ISSN: 1646-9895

risti

Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

DEZEMBRO 11 • DICIEMBRE 11

Nº 8

©AISTI 2011 http://www.aisti.eu

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Edição / Edición

Nº 8, 12/2011

Tiragem / Tirage: 500

Preço por número / Precio por número: 12,5€

Subscrição anual / Suscripción anual: 20€ (2 números)

ISSN: 1646-9895

Depósito legal:

Indexação / Indexación

DOAJ, EBSCO, GALE, IndexCopernicus, Index of Information Systems Journals, LatinIndex, ProQuest, Scielo e Ulrich's.

Propriedade e Publicação / Propriedad y Publicación

AISTI – Associação Ibérica de Sistemas e Tecnologias de Informação

Rua de Lagares 523 - Silvares, 4620-646 Lousada, Portugal

E-mail: risti@aisti.eu

Web: http://www.aisti.eu

Ficha Técnica

Director
Álvaro Rocha, Universidade Fernando Pessoa

Coordenador da Edição / Coordinador de la Edición
Luís Paulo Reis, Universidade do Minho

Conselho Editorial / Consejo Editorial
Carlos Ferrás Sexto, Universidad de Santiago de Compostela
Jose Antonio Calvo-Manzano Villalón, Universidad Politécnica de Madrid
Luís Borges Gouveia, Universidade Fernando Pessoa
Luís Paulo Reis, Universidade do Minho
Manuel Pérez Cota, Universidad de Vigo
Maria Manuela Cruz-Cunha, Instituto Politécnico do Cávado e do Ave
Ramiro Gonçalves, Universidade de Trás-os-Montes e Alto Douro

Secretariado Editorial
Avelino Victor, Instituto Superior da Maia e Instituto de Informática do Porto
Paulo Teixeira, Instituto Politécnico do Cávado e do Ave

Conselho Científico / Consejo Científico
Adolfo Lozano-Tello, Universidad de Extremadura
Alberto Fernández, Universidad Rey Juan Carlos
Aldemar Santos, Universidade Federal de Pernambuco
Alípio Jorge, Universidade do Porto, FCUP
Armando Sousa, Universidade do Porto, FEUP
Ana Paula Afonso, Instituto Politécnico do Porto, ISCAP
Antoni Lluís Mesquida Calafat, Universitat de les Illes Balears
António Godinho, Instituto Superior de Línguas e Administração-Gaia
Antonia Mas Pichaco, Universitat de les Illes Balears
Armando Mendes, Universidade dos Açores
Arturo J. Méndez, Universidad de Vigo

Carlos Costa, Universidade de Aveiro
Carlos Rabadão, Instituto Politécnico de Leiria
Carlos Vaz de Carvalho, Instituto Politécnico do Porto, ISEP
Carmen Galvez, Universidad de Granada
Daniel Castro Silva, Universidade de Coimbra, FCT
Didac Busquets, Imperial College London
Eduardo Destefanis, Universidad Tecnologica Nacional da Argentina
Eduardo Luís Cardoso, Universidade Católica Portuguesa - Porto
Feliz Gouveia, Universidade Fernando Pessoa
Fernando Bandeira, Universidade Fernando Pessoa
Fernando Diaz, Universidad de Valladolid
Francisco Restivo, Universidade do Porto, FEUP
García Pérez-Schofield Baltasar, Universidad de Vigo
Gonçalo Paiva Dias, Universidade de Aveiro
Ivan Garcia, Universidad Tecnologica de la Mixteca
Jaime S. Cardoso, Universidade do Porto, FEUP
Javier Garcia Tobio, CESGA-Centro de Supercomputacion de Galicia
Jesús Pardillo, Universidad de Alicante
Jezreel Mejia, Instituto Tecnologico de Orizaba
João Barroso, Universidade de Trás-os-Montes e Alto Douro
João Paulo Costa, Universidade de Coimbra, FEUC
João Sarmento, Universidade do Minho
João Tavares, Universidade do Porto, FEUP
Joana Maria Segui Pons, Universitat de les Illes Balears
Joaquim Cezar Filipe, Universidade de São Paulo
Joaquim Reis, Instituto Superior de Ciências do Trabalho e da Empresa
Jörg Thomaschewski, University of Applied Sciences OOW - Emden
José Adriano Pires, Instituto Politécnico de Bragança
José Paulo Lousado, Instituto Politécnico de Viseu
Jose M Molina, Universidad Carlos III de Madrid
Juan Carlos González Moreno, Universidad de Vigo
Juan José de Benito Martín, Universidad de Valladolid

Juan Manuel Fernández-Luna, Universidad de Granada
Juan-Manuel Lopez-Zafra, Universidad Complutense de Madrid
José Silvestre Silva, Universidade de Coimbra, CIUC
Luciano Boquete, Universidad de Alcalá
Luis de Campos, Universidad de Granada
Luis Fernandez-Sanz, Universidad de Alcalá
Luisa María Romero-Moreno, Universidad de Sevilla
Magdalena Arcilla Cobián, Universidade Nacional de Educación a Distancia
Marco Painho, Universidade Nova de Lisboa, ISEGI
Maria Clara Silveira, Instituto Politécnico da Guarda
Maria Helena Monteiro, Universidade Técnica de Lisboa, ISCSP
María J. Lado, Universidad de Vigo
Maria João Castro, Instituto Politécnico do Porto, ISCAP
Martin Llamas-Nistal, Universidad de Vigo
Mercedes Ruiz, Universidad de Cádiz
Miguel de Castro Neto, Universidade Nova de Lisboa, ISEGI
Miguel Mira da Silva, Universidade Técnica de Lisboa, IST
Mirna Ariadna Muñoz Mata, Instituto Tecnologico de Orizaba
Montserrat Sebastià, Universitat de Barcelona
Nuno Lau, Universidade de Aveiro
Nuno Ribeiro, Universidade Fernando Pessoa
Orlando Belo, Universidade do Minho
Paulo Pinto, Universidade Nova de Lisboa, FCT
Pedro Abreu, Universidade de Coimbra, FCT
Pedro Miguel Moreira, Instituto Politécnico de Viana do Castelo
Paulo Silva, Universidade Lusíada de Famalicão
Pedro Nogueira Ramos, Instituto Superior de Ciências do Trabalho e da Empresa
Pedro Pimenta, Universidade do Minho
Pedro Sanz Angulo, Universidad de Valladolid
Reinaldo Bianchi, Centro Universitario da FEI
Sergio Gálvez Rojas, Universidad de Málaga
Thomas Panagopoulos, Universidade do Algarve

Tomás San Feliu Gilabert, Universidad Politécnica de Madrid
Vitor Santos, Universidade de Trás-os-Montes e Alto Douro
Xose A. Vila, Universidad de Vigo

Editorial

O oitavo número da RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) é dedicado à temática dos Sistemas Inteligentes e de Apoio à Decisão. Integra contribuições originais e relevantes nestas temáticas, nas suas diferentes dimensões, com ênfase para os Agentes Inteligentes e Sistemas Multi-Agente; Ambientes Inteligentes; Aprendizagem Computacional e Mineração de Dados; Reconhecimento de Padrões; Análise Inteligente de Vídeo/Imagem; Representação do Conhecimento; Sistemas de Apoio à Decisão; e Aplicações de Sistemas Inteligentes e de Apoio à Decisão.

O conjunto de seis artigos que se publica neste número da RISTI resultou da avaliação cuidada realizada pelos membros da comissão científica dos trinta e sete trabalhos submetidos pelos autores a este número especial da revista. Cada trabalho foi sujeito à avaliação por três membros da comissão científica, resultando numa taxa de aceitação final de 16%. Dos trabalhos submetidos, 14 foram oriundos de Espanha, 8 do Brasil, 7 de Portugal, 3 de Cuba, 2 da República Dominicana, 1 do México, 1 da Argentina e 1 da Colômbia.

Os artigos publicados cobrem diversas vertentes das áreas dos Sistemas Inteligentes e de Apoio à Decisão. Mais especificamente abordam: o reconhecimento de acções humanas em vídeo, algoritmos de recomendação para comércio electrónico, desenvolvimento eficiente de sistemas multi-agente, classificação de textos usando a Wikipedia, análise de opiniões expressas em redes sociais e decisão na selecção de provedores de Tecnologias da Informação.

O primeiro artigo aborda o cada vez mais actual e importante problema do reconhecimento de acções humanas a partir de vídeos. É apresentado um novo algoritmo para a detecção de comportamentos humanos em tempo real. Inclui ainda resultados da sua aplicação a repositórios de dados públicos, contendo vídeos com acções humanas, gravados em diversas condições.

O segundo artigo apresenta um algoritmo de recomendação para aplicações de E-commerce, baseado em passeios aleatórios num grafo bipartido. Apresenta também a avaliação do algoritmo, realizada com dados provenientes de transacções comerciais reais, através da sua comparação com cinco outros algoritmos, usando uma base de dados comum.

O terceiro artigo aborda o problema do desenvolvimento eficiente de Sistemas Multi-Agente. Apresenta um modelo/arquitectura que funciona como base para uma construção mais eficiente de Sistemas Multi-Agente. O modelo baseia-se em cinco princípios fundamentais: estandardização, especialização, facilidade de implementação, reutilização e computação distribuída. Assenta no conceito de comunidades inteligentes especializadas com funcionalidades específicas.

O quarto artigo descreve um classificador que tem como base a Wikipedia para a criação do corpus que define uma dada categoria mediante técnicas de processamento de linguagem natural que analisam sintacticamente os textos a classificar.

O quinto artigo está centrado na análise de opiniões expressas em redes sociais. Para tal, utiliza técnicas de Análise Sentimental, para verificar se a informação existente no Facebook e no Twitter pode ser utilizada para estimar valores úteis para a comercialização, de bens ou serviços, com vista ao seu lançamento no mercado.

Finalmente, no último artigo, é tratada a problemática da decisão na selecção de provedores de Tecnologias da Informação. Através da aplicação de um protocolo de revisão sistemática de Engenharia de Software o artigo apresenta a selecção de dezasseis critérios de avaliação de provedores de Tecnologias de Informação.

Por último gostaria de agradecer, a todos os autores, por terem escolhido a Revista Ibérica de Sistemas e Tecnologias de Informação, e este número temático em particular, para submeterem o seu trabalho de investigação, contribuindo para a qualidade deste número e da revista. Finalmente, gostaria de agradecer o esforço e dedicação do conselho científico, que teve árduo trabalho de avaliação dos artigos, permitindo a sua avaliação cuidada.

Luís Paulo Reis

Universidade do Minho

Índice

Método espacio-temporal para el reconocimiento de acciones humanas en el espacio canónico 1

I. Gómez-Conde, D.N. Olivieri, X.A. Vila

Algoritmo de Recomendação Baseado em Passeios Aleatórios num Grafo Bipartido 15

Ricardo Gama , Nuno André, César Pereira, Luís Almeida, Pedro Pinto

SCODA para el Desarrollo de Sistemas Multiagente 25

Jesús A. Román, Dante I. Tapia, Juan M. Corchado

Clasificación de textos en lenguaje natural usando la Wikipedia 39

Jose María Quinteiro-González, Ernestina Martel-Jordán, Pablo Hernández-Morera, Juan A. Ligero-Fleitas, Aaron López-Rodriguez

Análise de opiniões expressas nas redes sociais 53

Diogo Teixeira, Isabel Azevedo

Análisis de Decisiones en la Selección de Proveedores de Tecnologías de la Información: Una Revisión Sistemática 67

Lucas Grossi, Jose A. Calvo-Manzano

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido / Recibido: 13/07/2011
Aceitação / Aceptación: 09/12/2011

Método espacio-temporal para el reconocimiento de acciones humanas en el espacio canónico

I. Gómez-Conde [1], D.N. Olivieri, X.A. Vila [1]

ivangconde@uvigo.es, olivieri@ei.uvigo.es, anton@uvigo.es

[1] Departamento de Informática, Universidad de Vigo. Edificio Politécnico, As Lagoas s/n 32004 Ourense (España)

Resumen: El reconocimiento de acciones humanas es un campo de investigación muy activo en visión artificial, donde los esfuerzos se centran actualmente en la detección de comportamientos humanos en vídeos en tiempo real. En este trabajo, se presenta un algoritmo espacio-temporal, que denominamos Motion Vector Flow Instance (MVFI) para la clasificación de acciones sobre vídeos, y se muestran los resultados de su aplicación a dos conjuntos de datos, "KTH" y "MILE", que contienen escenas de acciones humanas con diferentes condiciones de grabación (varios ángulos de cámara, iluminación, diferentes prendas de vestir, calidad de vídeo...) La plantilla MVFI codifica la información de la velocidad del movimiento de una persona, a partir del flujo óptico que se obtiene en cada fotograma de un vídeo. A continuación, mediante aprendizaje supervisado, se proyectan las imágenes MVFI en el espacio canónico y se buscan los límites de decisión para varias acciones con máquinas de soporte vector (SVM). En este artículo, mostramos que este método para detectar acciones humanas, es robusto y permite un reconocimiento en tiempo real.

Palabras clave: reconocimiento de acciones humanas; visión por computador; análisis componentes principales; plantillas espacio-temporales

Abstract: The recognition of human actions is a very active research field in computer vision, where efforts are presently focused on the detection of human behavior in real-time video. In this paper, we present a novel spatio-temporal algorithm, called the Motion Vector Flow Instance (MVFI), for classification of actions in videos. We show the results of applying this algorithm to two public datasets, "KTH" and "MILE" that contain scenes of human actions with different recording conditions (multiple camera angles, lighting, different clothes, and video quality). The MVFI spatio-temporal template encodes information about the speed and direction of human motion from the optical flow vectors obtained within each video frame. Then, by using supervised learning, MVFI images are projected into a canonical vector space and decision boundaries are determined for various actions by using a support vector machines (SVM) algorithm. Thus, in this paper, we demonstrate that our method is robust for detecting human actions across different datasets and provides real-time recognition.

Keywords: human action recognition; computer vision; principal component analysis; spatio-temporal templates

1. Introducción

La detección automática de acciones y gestos humanos sobre vídeos es un campo muy activo de investigación y presenta un interés creciente para empresas de diversos sectores. Por ejemplo, en el campo del cuidado de la salud, no existen sistemas automatizados que permitan monitorizar la actividad de una persona en todo momento. Existen algunos sistemas, como las pulseras de emergencia o la monitorización de los signos vitales de una persona, pero ninguno de ellos permite seguir la actividad de esa persona ni detectar automáticamente la acción que realiza. Los sistemas de videovigilancia actuales tampoco sirven al requerir de supervisión humana, puesto que no son capaces de detectar automáticamente la acción que realiza esa persona.

Clasificar acciones humanas es un problema complejo en visión artificial. Existen diferentes enfoques como son el seguimiento del cuerpo en 3D mediante múltiples cámaras, el uso de modelos para caracterizar las imágenes en 2D de un vídeo o el uso de imágenes que guardan el historial del movimiento de fotogramas anteriores. El artículo de Poppe (2010) ofrece una revisión reciente de este campo. Si nuestro objetivo es detectar tipos de movimiento, tales como actividades anormales o caídas, el seguimiento de la persona en 3D produce excesiva información y supone un elevado coste computacional. Una forma de reducir ese coste podría basarse en la reducción de la dimensionalidad de los datos del vídeo transformando una secuencia de imágenes en puntos en un espacio canónico, en el que se podrían distinguir formas de "andar" normales (Huang, 1999) de otras irregulares, como es el caso de las personas que padecen la enfermedad de Parkinson (Cho, 2010).

En este artículo se describe un algoritmo de visión artificial para la detección de actividades humanas usando una transformación en el autoespacio de las imágenes y se muestra su aplicación práctica para discriminar dos actividades muy comunes como son "andar" y "hacer footing". En la sección 2 se analiza el estado del arte en el que se encuentra la detección y reconocimiento de acciones humanas en vídeos. A continuación, se describen las bases de datos de vídeo utilizadas para realizar los análisis y se detalla la plantilla espacio-temporal desarrollada así como las matemáticas asociadas a la transformación canónica. En el punto siguiente se comenta la metodología que se utilizó para dividir los vídeos en conjuntos de entrenamiento y conjuntos de test y finalmente, se presentan los resultados y conclusiones que se han obtenido de este trabajo.

2. Estado del arte

El reconocimiento de acciones humanas sobre vídeo es un problema antiguo para los investigadores en visión artificial, y ha madurado lo suficiente para permitir la existencia de varios artículos recientes que resumen las técnicas empleadas hasta el momento. Un ejemplo, es el artículo de Forsyth et. al. (2006) que han descrito la cinemática del seguimiento del cuerpo humano, y centran su interés en entender su movimiento. Más recientemente un extenso trabajo de Poppe (2010) describe los principales enfoques existentes para etiquetar acciones humanas sobre secuencias de vídeo.

Para describir el movimiento de todo el cuerpo humano, se han desarrollado varios sistemas que permiten realizar un seguimiento en 2-dimensiones y trasladarlo a 3-dimensiones. El trabajo de Deutscher (2005), por ejemplo, describe el uso de modelos con 26 ángulos de movilidad y usa filtros de partículas en cada nodo para reconstruir el movimiento del cuerpo humano al completo. Bobick et. al. (2001) se centran en el reconocimiento de actividades a través de la representación mediante plantillas de movimiento. Ellos introducen los conceptos de instancias de historial de movimiento (Motion History Instance MHI) y de instancias de energía de movimiento (Motion Energy Instance MEI) que permiten capturar información a través de muchas secuencias de imágenes. Existen otros métodos modernos como el que describen Meeds et. al. (2008) que usa modelos gráficos bayesianos para inferir modelos de actividades humanas y animales sin necesidad de conocimiento previo del tipo de movimiento. Ikizler y Forsyth (2008) han mostrado que los movimientos atómicos de las diferentes partes del cuerpo en 3-dimensiones, cuando son combinados con modelos ocultos de Markov (HMMs), pueden ser usados para inferir composiciones de movimientos, tales como la combinación de fonemas para construir palabras.

El objetivo de estos métodos es proporcionar una descripción completa del movimiento humano, pero son computacionalmente muy costosos para aplicaciones en tiempo real. Las técnicas basadas en información espacio-temporal son más adecuadas y eficientes en esos casos. La información espacio-temporal, dirección y sentido del movimiento, puede ser codificada en plantillas que representan el movimiento y permiten capturar características del mismo a través de diferentes fotogramas. Cuando se combinan con algoritmos para reducir la dimensionalidad de los datos, como análisis de componentes principales (PCA), estos métodos son capaces de clasificar diferentes tipos de movimientos. La idea tiene sus orígenes en el trabajo de Etemad y Chellappa (1997) para el reconocimiento de caras, quienes utilizaron análisis lineal discriminante (LDA), también llamado análisis canónico, para optimizar la separación de diferentes caras. La idea fundamental de este método es que, dado que las características esenciales de una cara son las mismas, es la covarianza la que juega un papel importante en la discriminación. Esta idea, aplicada a las transformaciones en los autoespacios para el movimiento humano, se utilizó por primera vez para diferenciar personas basándose en su forma de "andar". Huang et. al. (1999) usaron PCA para clasificar formas de "andar" de diferentes personas. Ellos introducen la idea de una transformación en el autoespacio canónico, que consiste primero en aplicar PCA seguido de LDA para reducir la varianza entre los elementos de una clase mientras se maximiza la varianza entre los elementos de diferentes clases. Das et. al. (2006) usan PCA en dos etapas, donde la primera consiste en reducir la dimensionalidad y la segunda en comparar las curvas en el nuevo espacio. En una interesante aplicación de reconocimiento de formas de "andar", Cho et. al. (2009) han mostrado el uso de PCA para analizar el grado de enfermedad de Parkinson en pacientes de avanzada edad basándose en su forma de "andar".

3. Metodología

En nuestro algoritmo mediante aprendizaje supervisado, se entrena el sistema con varias secuencias de vídeo de diferentes acciones y diferentes actores. Estas secuencias son transformadas mediante las plantillas MVFI (Olivieri et. al., 2011) y posteriormente

se mapean en diferentes puntos en el autoespacio. A continuación, se realizan una serie de transformaciones canónicas (PCA y LDA) para reducir drásticamente la dimensión de cada secuencia de imágenes a un conjunto de puntos en un espacio multidimensional mucho más pequeño.

Una vez que disponemos en el espacio canónico de la trayectoria de puntos de cada una de las acciones, se realiza un entrenamiento utilizando máquinas de soporte vector (SVM). Cada uno de los puntos es etiquetado en función de su categoría ("andar o "hacer footing"). El reconocimiento de una nueva secuencia de una acción consiste en formar la plantilla MVFI correspondiente y realizar la transformación al espacio canónico creado en el entrenamiento, para posteriormente determinar mediante la SVM si pertenece a una categoría u otra. En la figura 1 se puede ver el flujo de trabajo de todo el proceso.

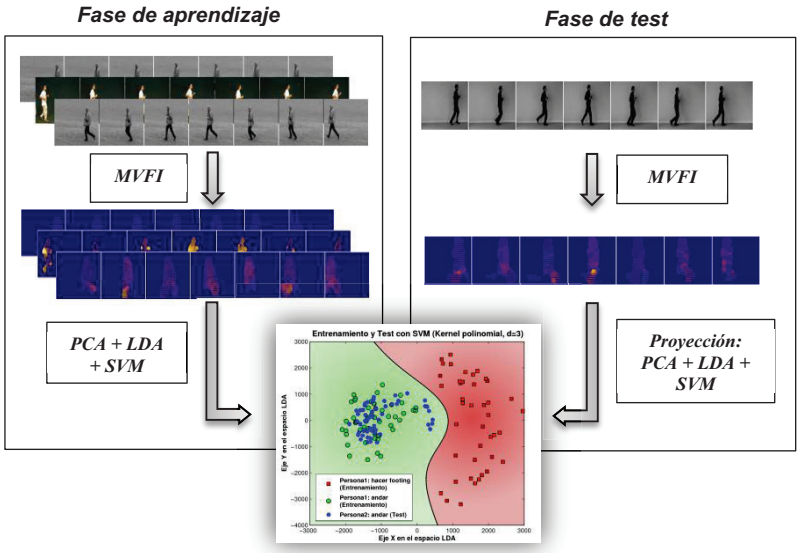


Figura 1 – Flujo de trabajo para el entrenamiento y la clasificación de nuevas secuencias de vídeo

3.1. Conjuntos de datos: KTH y MILE

Hemos realizado un extenso entrenamiento y validación para evaluar el rendimiento del método propuesto. En la figura 2 se pueden observar los dos tipos de acciones utilizadas para realizar el entrenamiento: "andar" y "hacer footing". Se han utilizado dos conjuntos de datos: KTH-Human, una base de datos de vídeo pública y MILE-Human, creada por nosotros para realizar las pruebas.

Figura 2 – Ejemplos de las acciones "andar" y "hacer footing" en las bases de datos de vídeo de KTH y MILE.

KTH Human (Vapnik, 1995): este conjunto de vídeos contiene seis tipos de acciones: andar, hacer footing, correr, boxear, agitar las manos y aplaudir. En este trabajo sólo se utilizaron las acciones de "andar" y "hacer footing". Tomamos los vídeos de 11 personas en tres tipos de escenarios diferentes: exteriores con ropa uniforme, exteriores con ropas diferentes e interiores. Todas las secuencias utilizadas fueron grabadas sobre fondos homogéneos con una cámara estática a una velocidad de 25 fotogramas por segundo. La duración media de cada secuencia es de 4 segundos y tienen una resolución de 160x120 píxels.

MILE Human (MILE, 2011): se trata de una base de datos de vídeo creada para realizar las pruebas con este nuevo algoritmo; contiene varios tipos de acciones: andar, hacer footing, sentarse, acostarse, caerse… En este artículo sólo se han utilizado dos acciones: "andar" y "hacer footing". Se grabaron un total de 5 secuencias por cada acción, con 12 personas. Todas las actividades fueron grabadas utilizando la misma distancia focal a una velocidad de 25 fotogramas por segundo. La duración de cada uno de los vídeos es de 5 segundos de media y tiene una resolución de 572x512 píxels.

Se dividió el conjunto de vídeos en un conjunto de entrenamiento y un conjunto de test con diferente número de personas. Las imágenes han sido procesadas a bajo nivel para generar la plantilla de movimiento MVFI, utilizando programas escritos en Python con llamadas a funciones de la librería de visión artificial OpenCV2.1 (Bradski, 2008). Estos algoritmos procesan los vídeos y producen las secuencias de imágenes del movimiento: extraen la plantilla de movimiento de los fotogramas individuales, reducen el tamaño de las imágenes a 128x128 usando un algoritmo adaptativo de redimensionado, y finalmente las convierten a escala de grises para realizar la transformación al espacio canónico.

3.2. Plantilla espacio-temporal MVFI

La Instancia del Vector de Flujo de Movimiento (Motion Vector Flow Instance, MVFI) es un nuevo método espacio-temporal, desarrollado por nosotros (Olivieri et. al., 2011), que usa el flujo óptico para guardar la información de la magnitud de la velocidad y dirección del movimiento en cada fotograma. Muchas de las actividades que se realizan diariamente, se caracterizan por aumentos bruscos de velocidad en la dirección horizontal y/o vertical. Nuestra plantilla de movimiento codifica la información de la velocidad y dirección del movimiento para ser sensible a movimientos bruscos del cuerpo humano.

En primer lugar, caracterizamos el vídeo de entrada $V\ (w,\ h,\ N_f)$ mediante su resolución (ancho w, alto h) y el número de fotogramas N_f para inicializar el vídeo de

salida *O(w, h)* que contendrá la plantilla MVFI de ese movimiento. En cada instante de tiempo *t* el fotograma correspondiente (matriz con los valores en escala de grises de cada píxel) se divide en una cuadrícula que contiene el campo de flujo óptico, donde cada uno de los puntos de esa cuadrícula se representan por *(x, y)*. En cada uno de esos puntos evaluamos el vector de movimiento (figura 3a) a través de un algoritmo piramidal muy conocido (Farnebäck, 2003), y creamos un rectángulo por cada vector (figura 3b). En cada instante *t* se inicializa una lista de objetos *s* para contener los vectores de flujo *f* en dicho instante. Por cada punto de la cuadrícula *(x, y)* en el flujo óptico, se almacenan en *s* los componentes del vector *(fx, fy)* junto con la magnitud $||f||$. Posteriormente, ordenamos los vectores de *s* por la magnitud $||f||$, para colocar los vectores de mayor peso al principio de la lista. Finalmente, se escribe el vídeo de salida, codificando cada vector de flujo óptico en *s*, desde el más pequeño al más grande, para asegurar que los vectores de flujo más grandes no sean sobrescritos por otros más pequeños. Cada uno de los rectángulos leídos de la lista, se codifican en función de la magnitud $||f||$ como se muestra en la figura 3b.

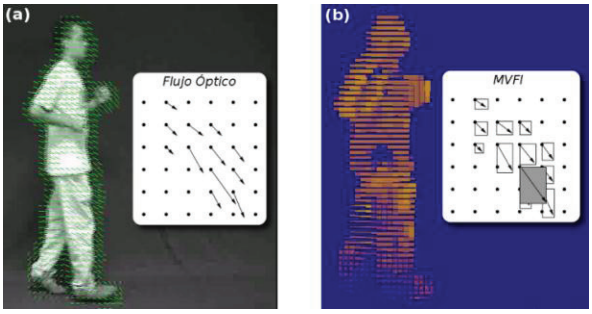


Figura 3 – Vectores de flujo óptico y MVFI en un fotograma de un vídeo de una persona haciendo footing.

Nuestro método MVFI difiere del historial de flujo de movimiento (MHI) propuesto por Bobick y Davis (2001) en dos aspectos fundamentales. En primer lugar en que no se mantiene un historial de los *x* fotogramas anteriores, sino que usa una fotografía instantánea de los vectores de flujo en un instante *t* que permite codificar los cambios bruscos de velocidad más eficientemente. Y en segundo lugar, en que el tamaño de los vectores se codifica en el tamaño de los rectángulos, que son almacenadas en capas de tal forma que las velocidades mayores se almacenan en la cima. En la figura 4 se puede observar el resultado de aplicar la plantilla MVFI sobre un vídeo de una persona haciendo footing. En nuestra página Web (MILE, 2011) se puede observar el resultado de aplicar la plantilla MVFI a diferentes tipos de acciones humanas.

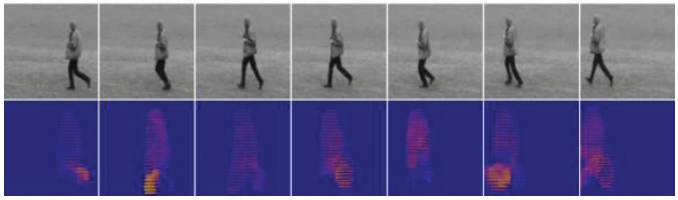


Figura 4 – Secuencia de imágenes de "hacer footing" junto con el resultado de la plantilla espacio-temporal MVFI.

3.3. Transformación canónica

Las transformaciones canónicas, basadas en PCA y LDA, se describen en la literatura científica y en recientes libros de texto (Fukunaga, 1990 y Gonzalez, 2003). Aquí resumiremos los detalles matemáticos de esta técnica, aplicados a nuestro problema particular, utilizando la notación estándar. En primer lugar, definimos una secuencia de movimiento *s*, que tiene N_f imágenes (o fotogramas) $x_{i=1 \ldots Nf}$. A continuación, definimos un tipo de acción humana (clase) *c*, la cual puede estar formada por N_s secuencias de movimiento. Luego formamos un vector de entrenamiento con el conjunto de las secuencias de imágenes, pertenecientes a diferentes clases, en total tenemos N_c clases. Obtenemos así un vector *X*, donde cada elemento $x_{k,\,i,\,j}$, se corresponde con la imagen de la clase *k* perteneciente a la secuencia *i*, y al frame *j*. El número total de imágenes en *X* es N_T , con $N_T = N_c N_s N_f$. Por lo tanto, el conjunto de entrenamiento viene dado por el vector:

$$X = [x_{1,1,1} \ldots x_{1,Ns,Nf}, \ldots, x_{Nc,1,1} \ldots x_{Nc,Ns,Nf}]$$

donde cada $x_{k,i,j}$ es la matriz de píxels del fotograma *j* de la secuencia *i* de la clase *k*. El espacio PCA canónico se construye a partir de los vectores ortogonales que poseen la mayor varianza entre todas las imágenes en *X*. Por lo tanto, para construir el espacio PCA, debemos encontrar la media del vector *X*, dada por m_x, y la covarianza C_x, que representa la desviación de los píxels de la media:

$$m_x = \frac{1}{N_T} \sum_{k=1}^{N_c} \sum_{i=1}^{N_s} \sum_{j=1}^{N_f} x_{k,i,j}$$

$$C_x = \frac{1}{N_T} \sum_{k=1}^{N_c} \sum_{i=1}^{N_s} \sum_{j=1}^{N_f} (x_{k,i,j} - m_x)(x_{k,i,j} - m_x)^T = \frac{1}{N} \bar{X} \bar{X}^T$$

A partir de los vectores propios o autovectores u_i y los valores propios o autovalores λ_i de C_x se determinan las direcciones ortogonales con la mayor varianza:

$$C_x u_i = \lambda_i u_i$$

asumiendo que C_x puede ser diagonalizada. Teniendo en cuenta que $\bar{X}\bar{X}^T$ es una matriz muy grande de *n x n* (*n* es el número total de píxels de $\bar{X}$), existe una simplificación práctica y bien conocida (ver *Descomposición en valores singulares en Fukunaga*, 1990) que consiste en considerar la matriz relacionada $\tilde{C}_x = \bar{X}^T \bar{X}$, la cual es solamente N_T *x* N_T. Esta transformación permite calcular el nuevo conjunto de autovectores y autovalores $(\tilde{u}_i, \tilde{\lambda}_i)$. Se puede demostrar que este nuevo espacio de $\tilde{C}_x$ está relacionado con el espacio original de C_x a través de una simple transformación en *X*.

Utilizando la teoría estándar de PCA, este nuevo espacio es reducido a solamente los $z \leq N_t - 1$ autovalores de mayor peso $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_z|$, lo cual se justifica demostrando que $|\lambda_j| \approx 0$ para $j > z$. El conjunto parcial de autovectores forman un nuevo espacio y = $[y_{1,1,1} \ldots y_{Nc,Ns,Nf}]$ en el que se representan las proyecciones de las imágenes originales:

$$y_{k,i,j} = [u_1 \ldots u_z]^T x_{k,i,j} = E x_{k,i,j}$$

Esta técnica proporciona un método directo para proyectar las imágenes originales de las secuencias dentro de un nuevo espacio multidimensional, donde cada punto representa una imagen y una trayectoria representa una secuencia de imágenes. Para clasificar diferentes tipos de acciones humanas a partir de la transformación de secuencias de vídeo en puntos $y_{k,i,j}$ en el nuevo espacio, aplicamos el conocido criterio de Fisher (Theodoridis, 2010) que usa la matriz de covarianza para maximizar la varianza *entre diferentes clases* (S_b), mientras que al mismo tiempo minimiza la varianza entre los elementos de *la misma clase* (S_w).

$$S_w = \frac{1}{N_T} \sum_{k=1}^{N_c} \sum_{i=1}^{N_s} \sum_{j=1}^{N_f} (y_{k,i,j} - m_{x,k})(y_{k,i,j} - m_{x,k})^T$$

$$S_b = \frac{1}{N_T} \sum_{k=1}^{N_c} N_k (m_{x,k} - m_x)(m_{x,k} - m_x)^T$$

donde $m_{x,k} = \frac{1}{N_T} \sum_{i=1}^{N_s} \sum_{j=1}^{N_f} x_{k,i,j}$. Obtenemos así, una nueva base ortogonal $[v_1, \dots, v_{c-1}]$, que toma los puntos $y_{k,i,j}$ del espacio PCA y los transforma en puntos del espacio LDA:

$$z_{k,i,j} = [v_1, \dots, v_{c-1}] y_{k,i,j} = V y_{k,i,j} = VE x_{k,i,j}$$

Hemos comprobado experimentalmente que, para obtener una buena clasificación, era suficiente la transformación en un nuevo espacio menor o igual a 10 dimensiones. En la figura 5 se muestra los excelentes resultados que se obtienen con las transformaciones en PCA y LDA de dos secuencias de acciones de "andar" y "hacer footing" representadas con únicamente 3 dimensiones, los tres autovectores de mayor peso.

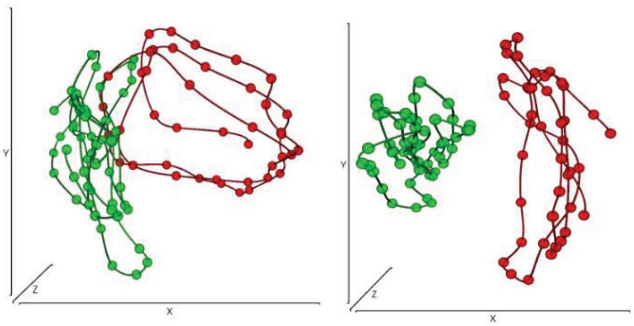


Figura 5 – Ejemplos de las transformaciones en PCA y LDA de dos secuencias de vídeo de las acciones "andar" y "hacer footing" representadas con los tres autovectores de mayor peso.

3.4. Entrenamiento

Para realizar el entrenamiento y la clasificación se dividió la base de datos de vídeo en la siguiente estructura: *acción : persona: secuencia*, donde *acción* representa las diferentes clases de movimiento, *personas*, son los diferentes sujetos que realizan la acción y *secuencias* es el conjunto de secuencias de vídeos procesados por acción y persona, y que utilizan la plantilla de movimiento espacio-temporal (MVFI). Todo ello constituye tres conjuntos para KTH (A_{KTH}, $P_{KTH,}$ S_{KTH}) y otros tres para MILE (A_{MILE},

$P_{MILE,}$ S_{MILE}) con un número total de (N_A=2 acciones, N_P =11 personas y N_S =4 secuencias por persona y acción) para KTH y (N_A =2 acciones, N_P =8 personas y N_S =4 secuencias por persona y acción) para MILE.

El entrenamiento se construyó encontrando todas las posibles combinaciones, como un producto cartesiano *n-ario*, a partir de los diferentes conjuntos definidos. Por ejemplo, un entrenamiento con todas las combinaciones de 2 clases con 3 personas y 1 secuencia por persona se representará como: T = [2][3][1].

El conjunto de test se obtiene considerando el mismo conjunto de acciones, A, pero usando el complemento de las secuencias de entrenamiento. De esta forma, el conjunto complementario, o conjunto de test, ($\tilde{T}$), es calculado por:

$$\tilde{T} = A\tilde{P}S \cup A\tilde{P}\tilde{S} \cup AP\tilde{S}$$

donde $\tilde{P}$, y $\tilde{S}$ son los complementarios a P y S respectivamente.

Para determinar la clase a la que pertenece una secuencia de test desconocida, se utiliza un clasificador de máquinas de vectores de soporte vector (SVM). Desde el punto de vista matemático, un conjunto de entrenamiento es de la forma:

$$D = \{(point_i, class_i) | class_i \in \{-1,1\}\}_{i=1}^{n}$$

donde $class_i$ es -1 o 1, indicando la clase a la que pertenece ("andar" o "hacer footing") cada uno de los puntos $point_i$. Cada $point_i$ es un vector de reales de *Z-dimensiones*. Y nosotros queremos encontrar el hiperplano que maximiza la distancia entre los puntos que tienen clase $class_i = -1$ de los que tienen clase $class_i = 1$.

El algoritmo original del hiperplano óptimo propuesto por Vapnik (1963) era un clasificador lineal. Sin embargo, Boser, Guyon y Vapnik (1992) sugirieron una manera de crear clasificadores no lineales aplicando un kernel para maximizar la separación entre los hiperplanos. En este artículo, se han analizado los resultados con varios kernels (lineal, cuadrático, gaussiano de base radial, polinomial y perceptrón multicapa). Los mejores resultados fueron presentados con el kernel polinomial que se define como: $k(point_i, class_i) = (point_i \cdot class_i)^d$ donde d es la dimensión del espacio del kernel. Se realizaron pruebas con diferentes valores para d, obteniendo los mejores resultados con un valor de d=3. Por lo tanto, todos los resultados presentados en este artículo se obtuvieron utilizando un kernel polinomial con d=3.

En la figura 6 podemos ver el resultado de aplicar un SVM con un kernel polinomial (d=3) a un conjunto de datos de entrada. Vemos como todos los puntos de la nueva secuencia de test (aspas en la figura) son correctamente mapeados en la parte del espacio que el SVM clasifica como "andar".

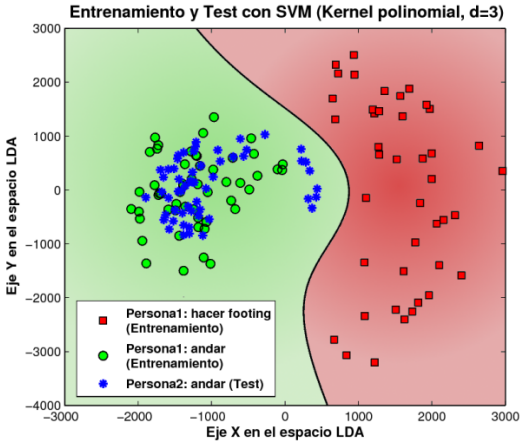


Figura 6 – Clasificicación de dos secuencias de vídeo ("andar" y "hacer footing") y test de un nuevo vídeo de una persona andando, mediante un SVM con kernel polinomial.

4. Experimentos y resultados

Hemos realizado un entrenamiento con todas las posibles combinaciones de vídeos de las bases de datos que hemos utilizado. Analizamos la influencia de diferentes ropas, fisionomías y condiciones de luz (escenas de interior o exterior) así como el número de personas necesarias en el entrenamiento. En esta sección, presentamos los resultados más relevantes.

La tabla 1 muestra un resumen estadístico de algunas de las combinaciones de secuencias de vídeo utilizadas en el proceso de entrenamiento. La primera columna, es la combinación de entrenamiento: $T = (N_{acciones}, N_{personas}, N_{secuencias})$.

Tabla 1 – Resumen de algunas estadísticas de los entrenamientos, mostrando el número medio de imágenes, el tiempo medio de cada entrenamiento y el tiempo total de ejecución.

T_{KTH}	Nº Combinaciones	Nº Medio Imágenes	Media Entrenam. (segundos)	Total Entrenam. (segundos)
2,2,1	55	118.2	5.87	341.85
2,6,1	373	392.4	42.8	16071
T_{MILE}				
2,2,1	28	187.5	11.2	321.6
2,6,1	70	367.8	39.61	2794.7

Como se puede apreciar, el número medio de fotogramas por secuencia varía en función de la base de datos de vídeo, aproximadamente 120 imágenes en KTH-Human y 187 en MILE-Human con $N_p = 2$. Al aumentar el número de personas en el entrenamiento (N_P =6) vemos como aumenta el número de imágenes en cada una de las combinaciones de entrenamiento a 390 en KTH-Human y 370 en MILE-Human

aproximadamente. En ambas bases de dato de vídeo los tiempos medios de ejecución por entrenamiento están entre 6 y 11 segundos con entrenamientos del tipo (2, 2, 1) y aproximadamente 40 segundos en el caso de entrenamientos con 6 personas.

En el estudio realizado para este artículo se han analizado los resultados de entrenamiento y clasificación de acciones humanas tanto de KTH-Human como MILE-Human en función de algunos parámetros. Por ejemplo, comparamos las tasas de acierto en el reconocimiento de nuevas secuencias de vídeo en función del número de personas N_p incluidas en el entrenamiento comprobando que con sólo 6 personas se logró una tasa de acierto media de 97.48%

Para analizar la influencia de cambios en las bases de datos, hemos realizado una validación cruzada con todas las posibles combinaciones de entrenamiento y test posibles entre las dos bases de datos. En la tabla 2 se pueden observar todas las matrices de confusión para las combinaciones de entrenamiento y test. Por ejemplo, el caso $Tr_{KTHMILE} - Tst_{MILE}$ hace referencia a los resultados obtenidos con un entrenamiento que incluye todas las combinaciones de vídeos de 6 personas y 1 secuencia por persona en KTH-Human y MILE-Human (T_{KTH} = (2,6, 1) y T_{MILE} = (2,6,1)). Posteriormente intentamos clasificar los vídeos que pertenecen al conjunto complementario del entrenamiento en la base de datos de vídeo de KTH-Human ($\bar{T}_{KTH}$), y obtuvimos un 99.44% de clasificaciones correctas para la acción de "andar" y un 93.75% para la acción de "hacer footing".

Tabla 2 – Matrices de confusión para los diferentes conjuntos de entrenamiento y test.

Tr_{KTH} - Tst_{KTH}	Andar	Footing
Andar	98.70	1.29
Footing	11.77	88.22

Tr_{MILE} - Tst_{MILE}	Andar	Footing
Andar	100	0
Footing	0	100

$Tr_{KTHMILE}$ - Tst_{MILE}	Andar	Footing
Andar	99.44	0.55
Footing	6.25	93.75

$Tr_{KTHMILE}$ - Tst_{KTH}	Andar	Footing
Andar	99.59	0.41
Footing	5.38	94.88

Tr_{KTH} - Tst_{MILE}	Andar	Footing
Andar	99.56	0.44
Footing	4.35	95.65

Tr_{MILE} - Tst_{KTH}	Andar	Footing
Andar	100	0
Footing	0	100

Se observó que los mejores resultados se obtuvieron entrenando y clasificando con nuestra propia base de datos de vídeo. Esto es debido a que todas las escenas fueron grabadas en el mismo lugar, las personas siempre se encontraban a la misma distancia de la cámara aproximadamente y no había cambio de ángulos. En KTH-Human sin embargo, se han utilizado vídeos con personas a diferentes distancias, escenas de interior y de exterior, lo cual justifica que los resultados no sean tan buenos.

La tabla 3 muestra los resultados obtenidos por diferentes estudios que han tratado de discriminar acciones y que han utilizado la base de datos de KTH para validación y test. Y más concretamente, que han comparado las acciones de "andar" y "hacer footing". Como podemos apreciar nuestros resultados son mejores que en trabajos previos.

Nuestro método es capaz de distinguir entre las acciones de "andar" y "hacer footing" con un porcentaje de acierto del 96.90%, incluyendo simplemente 6 personas en el entrenamiento. La mayoría de trabajos previos habían utilizado 16 personas para obtener buenos resultados.

Tabla 3 – Comparación de diferentes trabajos que han utilizando la base de datos pública de KTH para discriminar las acciones "andar" y "hacer footing".

Trabajo	**Tasas de acierto**	**N_P en el entrenamiento**
Jhuang, 2007	91.6 %	16 personas
Laptev, 2008	94 %	16 personas
Liu, 2008	94.5 %	16 personas
Cao, 2010	94.01 %	8 personas
Nuestro trabajo	96.90 %	6 personas

5. Conclusiones

Este artículo describe y presenta los resultados de la validación de una técnica denominada MVFI con diferentes bases de datos de vídeo. En concreto se ha utilizado una base de datos de vídeo propia (MILE-Human) y una base de datos pública muy utilizada en visión artificial (KTH-Human). Nuestro sistema es capaz de distinguir acciones humanas complejas como son "andar" y "hacer footing" sobre múltiples vídeos grabados en diferentes situaciones, cumpliendo la única característica de que la cámara sea estática. En futuros trabajos, se considerarán diferentes ángulos de cámara y distancias.

La técnica MVFI que hemos desarrollado, aunque es innovadora, está basada en métodos existentes, como el flujo óptico. Hemos demostrado que MVFI es capaz de detectar acciones humanas caracterizadas por cambios de velocidad. Por lo tanto, este trabajo sugiere la importancia de preservar la información de velocidad en cada secuencia de imágenes si deseamos utilizar técnicas de transformaciones canónicas para discriminar actividades humanas.

La ventaja de este sistema frente a otros existentes en la actualidad es la capacidad de clasificar acciones humanas en tiempo real, una vez que se ha realizado el entrenamiento, para ello se realizaron pruebas en un ordenador personal típico (Intel Core 2 Duo E8400 3.00GHz, 4 Gb RAM) obteniendo excelentes resultados. El proceso de coger una nueva secuencia de vídeo, transformarla con la plantilla de movimiento MVFI y proyectarla en el espacio canónico creado en el entrenamiento se puede realizar en tiempo real. Una de las aplicaciones en las que pensamos utilizar esta metodología es la segmentación automática de secuencias de vídeo, en función de los tipos de acciones realizadas.

Una futura línea de investigación en la que estamos trabajando, es la búsqueda y recuperación de vídeos basándose en contenido. El usuario podrá grabar un vídeo y nuestro algoritmo será capaz de obtener en una base de datos todos los vídeos con movimientos y acciones similares a los que esa persona ha realizado. El análisis se realizará con nuestra técnica de reconocimiento de acciones MVFI combinada con un algoritmo de seguimiento.

Referencias bibliográficas

Bobick, A. F., Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(3), 257-267.

Boser, B. E., Guyon, I. M., Vapnik, V.Ñ. (1992). A training algorithm for optimal margin classifiers. *In Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, 144-152.

Bradski, A. (2008). Learning OpenCV, Computer Vision with OpenCV Library. *O'Reilly Media, 1. edition*.

Cao, L., Liu, Z., Huang, T. (2010). Cross-dataset action detection. *In Computer Vision and Pattern Recognition, 2010. CVPR 2010.*

Cho, Ch. W., Chao, W. H., Lin, S. H., Chen, Y. Y. (2009) A vision-based analysis system for gait recognition in patients with parkinson's disease. *Expert Systems with Applications, 36*(3), 7033-7039.

Das, S. R., Wilson, R. C., Lazarewicz, M. T., Finkel, L. H. (2006) Two-stage PCA extracts spatiotemporal features for gait recognition. *Journal of Multimedia, 1*(5), 9-17.

Deutscher, J., Reid, I. (2005). Articulated body motion capture by stochastic search. *Int. J. Computer Vision, 61*(2), 185-205.

Etemad, K., Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *In Proceedings of the First International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA '97,* 127-142.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. *In Proceedings of the 13th Scandinavian Conf. on Image Analysis,* 363-370, 2003.

Forsyth, D. A., Arikan, O., Ikemoto, L., O'Brien, J., Ramanan, D. (2006) Computational studies of human motion: Part 1, tracking and motion synthesis. *Foundations and Trends in Computer Graphics and Vision, 1*(2/3):77-254.

Fukunaga, K. (1990). Introduction to statistical pattern recognition (2nd ed.). *Academic Press Professional, Inc.,* San Diego, CA, USA

Gonzalez, R. C., Woods, R. E., Eddins, S. L. (2003). Digital Image Processing Using MATLAB. *Prentice-Hall, Inc.*, Upper Saddle River, NJ, USA.

Huang, P. S., Harris, C. J., Nixon, M. S. (1999). Recognising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering, 13*(4), 359-366.

Ikizler, N., Forsyth, D. A. (2008) Searching for complex human activities with no visual examples. *Int. J. Computer Vision, 80*(3), 337-357.

Jhuang, H., Serre, T., Wolf, L., Poggio, T. (2007). A biologically inspired system for action recognition. *In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 1-8.

Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B. (2008). Learning realistic human actions from movies. *In 2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.

Liu, J., Shah, M. (2008). Learning human actions via information maximization. *In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 1-8.

Meeds, E. W., Ross, D.A., Zemel, R.S., Roweis S.T. (2008) Learning stick-figure models using nonparametric bayesian priors over trees. *In Proceedings of the EEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-8.

MILE (2011). Demos de MVFI en la Web de nuestro grupo de investigación. Disponible en: http://www.milegroup.net/demos

Olivieri D.N., Gómez Conde, I., Vila, X.A (2011). Eigenspace-based fall detection and activity recognition from motion templates and machine learning. *Expert Systems with Applications*, doi:10.1016/j.eswa.2011.11.109.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image & Vision Computing, 28*(6), 976-990.

Theodoridis, S., Pikrakis, A., Koutroumbas, K., Cavouras, D. (2010) Introduction to Pattern Recognition : A Matlab Approach. *Elsevier Inc.*

Vapnik, V. Ñ. (1995). The nature of statistical learning theory. *Springer-Verlag New York, Inc.*, New York, USA.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido / Recibido: 08/10/2011
Aceitação / Aceptación: 09/12/2011

Algoritmo de Recomendação Baseado em Passeios Aleatórios num Grafo Bipartido

Ricardo Gama [1], Nuno André, César Pereira, Luís Almeida, Pedro Pinto

[1]rgama@estgl.ipv.pt

Centro de Estudos em Educação, Tecnologias e Saúde, ESTGL, Instituto Politécnico de Viseu

Av. Coronel José Maria V. de Andrade, Campus Politécnico, 3504-510 VISEU, Portugal

Resumo: Neste trabalho apresentamos um novo algoritmo de recomendação para aplicações de *E-commerce*, baseado em passeios aleatórios num grafo bipartido. O desempenho do mesmo é testado com dados provenientes de transações comerciais reais e os resultados comparados com os de cinco outros algoritmos. Por fim, apresentamos possíveis generalizações do algoritmo, tornando-o muito versátil em aplicações práticas.

Palavras-chave: Algoritmos de Recomendação, Filtragem Colaborativa, Passeios Aleatórios.

Abstract: In this paper we present a new recommendation algorithm for E-commerce applications, based in random walks in a bipartite graph. Its performance is tested using real commercial transaction data and the results are compared with five other algorithms. Finally, we present possible generalizations for our algorithm making it very versatile for practical applications.

Keywords: Recommendation algorithms, Collaborative Filtering, Random Walks.

1. Introdução

Os algoritmos de recomendação representam, atualmente, um esforço acrescido, por parte das empresas de venda online, em fornecer aos seus clientes uma recomendação de produtos que vão ao encontro das características do consumidor, potenciando não só a venda dos seus produtos, mas também fornecendo um serviço de valor acrescentado ao consumidor que apenas deseja adquirir um produto.

Apesar do enorme sucesso de alguns algoritmos de recomendação (Zanker et al, 2007) grande parte apresenta alguns problemas, nomeadamente, na iniciação e esparsidade dos dados (Huang et al., 2004) contribuindo assim, para uma redução da qualidade das recomendações. Numa fase inicial, a informação que relaciona os consumidores com os produtos é escassa, o que dificulta a aprendizagem dos algoritmos e o bom

desempenho dos mesmos. Deste modo, é necessário criar métodos para reduzir ao máximo o tempo de aprendizagem dos algoritmos e dotá-los de mecanismos de previsão, baseados na pouca informação inicial disponível. Motivados pelo grande êxito de vários algoritmos aplicados a motores de busca na Internet, como os algoritmos *PageRank*, HITS e SALSA (Langville et al., 2006) surgiram, nos últimos anos, vários trabalhos que sugerem a utilização de modelos semelhantes em sistemas de recomendação (Huang et al., 2004,2007; Meyer et al., 2008; Pucci et al, 2007, Li et al., 2009).

2. Descrição do Algoritmo

Neste artigo, apresentamos um novo algoritmo que efetua recomendações baseadas em passeios aleatórios no grafo bipartido de consumidores/produtos.

À semelhança dos trabalhos apresentados por (Huang et al., 2004, 2007) no nosso modelo, interpretamos os dados das transações como um grafo bipartido, onde num dos grupos temos os consumidores e no outro os produtos possíveis de adquirir (2.1). Apesar de ter uma origem idêntica, o algoritmo apresentado neste artigo destaca-se dos trabalhos de Huang et al., pelo facto de assentar num modelo de passeios aleatórios diferente. Consequentemente, para além de ter um desempenho diferente, o algoritmo proposto utiliza apenas matrizes estocásticas (matrizes normalizadas por linhas) no seu funcionamento, evitando assim a necessidade de efetuar normalizações intermédias no decorrer do mesmo, que aumentam os tempos de execução e criam instabilidades numéricas desnecessárias.

De modo a preparar os dados para o inicio do algoritmo, a informação contida no grafo é armazenada numa matriz consumidores/produtos, A, cujas entradas a_{ij} indicam o número de vezes que o consumidor i comprou o produto j. Considerando que temos nc consumidores e np produtos, a matriz A vai ter dimensões $nc \times np$.

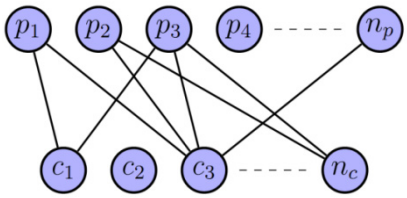


Figura 2.1: Grafo Consumidores/Produtos

Partindo deste modelo, pudemos criar as matrizes Nc e Np normalizando por linhas e por colunas a matriz A. Para contornar os problemas que surgem com clientes que não efetuaram compras, ou com produtos que nunca foram comprados, criámos dois vetores ac e ap de dimensões $nc \times 1$ e $np \times 1$ respetivamente, cujas entradas vão ser 1, caso o cliente (produto) tenha efetuado compras (sido comprado) e zero no caso contrário. Construímos as matrizes seguintes:

$$Mc = Nc + (1/np)\, ac \times ep^T, \qquad (1)$$

$$Mp = Np + (1/nc) ep \times ac^T, \qquad (2)$$

onde os vetores ec e ep de dimensões $nc \times 1$ e $np \times 1$ respetivamente, são vetores com todas as entradas iguais a um. Assim, contornámos o problema da inicialização, garantindo que todos os clientes possam ter recomendações. Um outro problema que pode surgir é a existência de grupos de clientes sem ligações entre eles. Para contornar esta situação, definimos a matriz:

$$Mm = (1/nc)\, ec \times ep^T, \qquad (3)$$

e criamos uma matriz Mc generalizada:

$$M_c^g = \beta_0 Mc + (1 - \beta_0) Mm. \qquad (4)$$

O nosso algoritmo consiste em simular nc passeios aleatórios, com recomeço, pelo grafo de consumidores/produtos, tendo como matrizes de transição, as matrizes M_c^g e Mp construídas anteriormente. Inicialmente, cada cliente inicia o seu passeio no vértice do grafo que lhe corresponde, avançando para o conjunto dos produtos, tendo em conta as probabilidades da matriz de transição M_c^g. Após este passo, cada cliente pode regressar para o grupo dos consumidores utilizando a matriz de transição Mp, com uma probabilidade α, ou retomar o seu ponto de partida, com uma probabilidade 1 - α. Este procedimento é repetido até se obter convergência ou até atingirmos um determinado número de iterações.

O resultado final do procedimento será uma matriz Cp, de dimensão $nc \times np$, que contém os vetores estacionários dos passeios de todos os clientes. Ou seja, obtemos uma matriz cujas entradas c_{ij} contêm a representatividade que o produto j tem para o cliente i.

Este procedimento pode ser resumido no seguinte algoritmo:

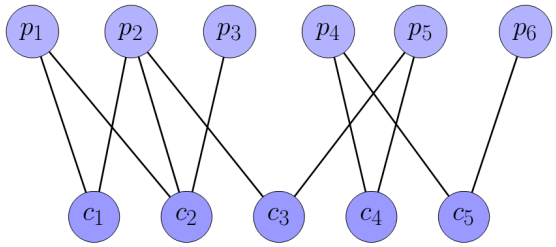
Entrada: M_c^g, M_p, α, erro máximo tolerado ϵ_m, número máximo de iterações k_m

C_0 =matriz identidade $n_c \times n_c$
$C^{(0)} = C_0$, $k = 1$
enquanto $k < k_m \wedge \epsilon > \epsilon_m$ **faça**
  $C_p^{(k)} = C^{(k-1)} \cdot M_c^g$
  $C^{(k)} = \alpha C_p^{(k)} \cdot M_p^T + (1 - \alpha) C^0$
  $k = k + 1$
  $\epsilon = |C_p^{(k)} - C_p^{(k-1)}|$
fim
retorna C_p

Figura 2.2: Algoritmo de Recomendação

A matriz Cp pode então ser utilizada para efetuar recomendações, escolhendo para cada cliente os produtos ainda não comprados com maior representatividade.

Vejamos um exemplo ilustrativo do algoritmo. Suponhamos que temos os seguintes registos de transações:



Produtos

Clientes	p_1	p_2	p_3	p_4	p_5	p_6
c_1	1	1	0	0	0	0
c_2	1	1	1	0	0	0
c_3	0	1	0	0	1	0
c_4	0	0	0	1	1	0
c_5	0	0	0	1	0	1

Figura 2.3: Exemplo de um grafo de transações e respectiva matriz

A partir desta informação podemos construir, tal como explicado anteriormente, as seguintes matrizes de transição:

$$M_c = \begin{pmatrix} 0.5000 & 0.5000 & 0 & 0 & 0 & 0 \\ 0.3333 & 0.3333 & 0.3333 & 0 & 0 & 0 \\ 0 & 0.5000 & 0 & 0 & 0.5000 & 0 \\ 0 & 0 & 0 & 0.5000 & 0.5000 & 0 \\ 0 & 0 & 0 & 0.5000 & 0 & 0.5000 \end{pmatrix} \qquad M_p = \begin{pmatrix} 0.5000 & 0.3333 & 0 & 0 & 0 & 0 \\ 0.5000 & 0.3333 & 1 & 0 & 0 & 0 \\ 0 & 0.3333 & 0 & 0 & 0.5000 & 0 \\ 0 & 0 & 0 & 0.5000 & 0.5000 & 0 \\ 0 & 0 & 0 & 0.5000 & 0 & 1 \end{pmatrix}$$

Inicializando $Cp^{(0)} = Mc$, aplicamos o algoritmo anterior até ser atingido o critério de paragem, obtendo a seguinte matriz:

$$C_p^{(\infty)} = \begin{pmatrix} 0.3057 & 0.3850 & 0.1112 & 0.0594 & 0.1168 & 0.0219 \\ 0.2843 & 0.3582 & 0.1732 & 0.0552 & 0.1087 & 0.0204 \\ 0.1532 & 0.3304 & 0.0738 & 0.1326 & 0.2609 & 0.0490 \\ 0.0723 & 0.1559 & 0.0348 & 0.3267 & 0.2896 & 0.1207 \\ 0.0424 & 0.0914 & 0.0204 & 0.3984 & 0.1698 & 0.2777 \end{pmatrix}.$$

Após a eliminação das compras previamente efetuadas, obtemos a matriz de recomendações:

$$\begin{pmatrix} - & - & 0.1112 & 0.0594 & 0.1168 & 0.0219 \\ - & - & - & 0.0552 & 0.1087 & 0.0204 \\ 0.1532 & - & 0.0738 & 0.1326 & - & 0.0490 \\ 0.0723 & 0.1559 & 0.0348 & - & - & 0.1207 \\ 0.0424 & 0.0914 & 0.0204 & - & 0.1698 & - \end{pmatrix}$$

A título de exemplo, podemos observar que se desejarmos efetuar duas recomendações ao cliente c_3, podemos ver que recomendaríamos os produtos $\{p_1; p_4\}$.

3. Avaliação Experimental e Resultados

Com o objetivo de avaliar o desempenho do algoritmo proposto neste trabalho, utilizámos dois conjuntos de dados cedidos em http://isl.ifit.uni- klu.ac.at, por (Zanker et al, 2007), *Buys Advisor Data* e *Buys All Data*, contendo informação sobre transações comerciais de uma loja online de cigarros cubanos.

Efetuámos a comparação dos resultados do nosso algoritmo com os resultados obtidos por cinco algoritmos de recomendação descritos na literatura:

- Recomendação aleatória, onde para cada cliente é escolhida, aleatoriamente, para recomendação uma lista de n produtos ainda não comprados pelo mesmo, (Aleat.);
- Recomendação dos produtos mais vendidos, onde para cada cliente é escolhida para recomendação, a lista dos *n* produtos mais comprados, que ainda não foram adquiridos pelo mesmo, (*Top n*);
- Algoritmos clássicos de filtragem colaborativa baseados em produtos e em clientes, usando o coeficiente de Pearson na criação das matrizes de semelhança [8, 2], (*CFUB* e *CFIB*);
- Algoritmo de recomendação proposto em (Huang et al., 2007), (L.A.ALg:).

A metodologia usada consistiu na utilização de dois métodos: dado n e todos menos n.

O método dado *n* consiste em utilizar n produtos selecionados aleatoriamente da lista de compras de um dado consumidor, que tenha pelo menos *n+1* compras, como dados de treino, considerando as restantes compras como conjunto de teste, *CT*.

No método todos menos *n*, por outro lado, usámos todos os produtos comprados pelos utilizadores que tenham pelo menos *n+2* compras, como dados de treino, exceto uma seleção aleatória de *n* deles, considerada como conjunto de teste.

Nos nossos testes, repetimos ambos os métodos 250 vezes para todos os utilizadores que verificavam as condições e medimos o desempenho dos vários algoritmos, calculando as medidas clássicas de precisão (P) e sensibilidade (S) (Huang et al., 2004; Zanker et al, 2007).

Após ter um conjunto de *nr* de artigos a recomendar para cada utilizador, *CR*, confrontámos o mesmo com o conjunto de teste e determinámos o conjunto de acertos do algoritmo de recomendação para esse cliente: $CA = CR \cap CT$, calculando posteriormente a precisão e a sensibilidade:

$$P = \frac{\#CA}{\#CR}\ ; \quad S = \frac{\#CA}{\#CT}\ .$$

Como o algoritmo apresentado neste trabalho depende do parâmetro x, usámos o mesmo, para valores de α de 0.01 a 0.99 por passos de 0.01, escolhendo no final o que apresentou valores mais elevados das medidas.

Os resultados obtidos estão sintetizados nas tabelas seguintes:

Buys Advisor Data

Dados n							
$n = 2$ & $n_r = 5$	α	proposto	Aleat.	topN	CFUB	CFIB	L.A.ALg.
S	0.82	16.08 ± 6	3.55 ± 4	15.50 ± 6	9.25 ± 3	9.02 ± 5	11.36 ± 4
P	0.93	7.16 ± 2	2.44 ± 2	8.24 ± 2	5.44 ± 2	4.37 ± 2	5.21 ± 2
$n = 3$ & $n_r = 5$							
S	0.95	14.71 ± 7	3.61 ± 5	13.62 ± 5	13.31 ± 5	6.64 ± 5	13.81 ± 7
P	0.95	8.59 ± 3	2.44 ± 2	9.79 ± 3	7.44 ± 3	5.24 ± 3	6.85 ± 3
Todos menos n							
$n = 1$ & $n_r = 5$	α	proposto	Aleat.	topN	CFUB	CFIB	L.A.ALg.
S	0.62	5.95 ± 3	1.26 ± 2	6.48 ± 3	3.70 ± 2	3.50 ± 2	4.82 ± 2
P	0.90	3.22 ± 1	0.66 ± 1	3.33 ± 1	2.22 ± 1	1.80 ± 1	2.23 ± 1
$n = 2$ & $n_r = 5$							
S	0.94	10.55 ± 4	2.27 ± 3	8.18 ± 4	7.25 ± 3	3.69 ± 3	8.70 ± 3
P	0.94	6.32 ± 3	1.47 ± 2	5.26 ± 2	4.15 ± 2	2.76 ± 2	4.97 ± 2

Tabela 1: Desempenho dos algoritmos em 250 repetições para o conjunto de dados *Buys Advisor Data*, contendo informação sobre as compras efetuadas por 37 clientes tendo à disposição 143 produtos, (média±desvio padrão).

Buys All Data

Dados n							
n=2 & nr=5	α	proposto	Aleat.	topN	CFUB	CFIB	L.A.ALg.
S	0.82	17.50 ± 2	3.64 ± 1	13.78 ± 2	14.68 ± 2	5.71 ± 1	16.5 ± 2
P	0.82	7.47 ± 1	1.64 ± 1	5.54 ± 1	6.32 ± 1	2.62 ± 1	6.92 ± 1
$n = 3$ & $n_r = 5$							
S	0.63	19.38 ± 3	3.70 ± 2	12.04 ± 3	14.72 ± 3	6.26 ± 2	16.52 ± 3
P	0.64	7.96 ± 1	1.69 ± 1	5.25 ± 1	6.27 ± 1	2.24 ± 1	6.82 ± 1
Todos menos n							
$n = 1$ & $n_r = 5$	α	proposto	Aleat.	topN	CFUB	CFIB	L.A.ALg.
S	0.81	6.89 ± 1	1.30 ± 1	5.41 ± 1	5.77 ± 1	2.59 ± 1	6.95 ± 1
P	0.47	3.80 ± 1	0.74 ± 1	2.81 ± 1	3.22 ± 1	1.42 ± 1	3.74 ± 1
n=2 & nr=5							
S	0.65	13.64 ± 2	2.65 ± 1	8.51 ± 2	11.82 ± 2	4.77 ± 1	11.56 ± 2
P	0.73	7.38 ± 1	1.44 ± 1	4.69 ± 1	6.31 ± 1	2.43 ± 1	6.21 ± 1

Tabela 2: Desempenho dos algoritmos em 250 repetições para o conjunto de dados *Buys All Data*, contendo informação sobre as compras efetuadas por 508 clientes tendo à disposição 143 produtos, (média±desvio padrão).

Os resultados apresentados nas tabelas anteriores, demonstram que, escolhendo o parâmetro α no intervalo 0.6 a 0.95, os resultados do algoritmo proposto neste trabalho superam os outros algoritmos, em alguns casos com melhorias significativas, evidenciando a vantagem da utilização deste método em sistemas de recomendação.

Observámos também que, os valores de desempenho em função do parâmetro α não variavam significativamente para $\alpha \in (0.6; 0.95)$, indicando que, em aplicações práticas se possa fixar um α escolhido deste intervalo.

4. Generalização e Aplicação em Contexto Real

Uma mais-valia do algoritmo proposto, face aos descritos em (Huang et al., 2004, 2007) é o facto de poder ser facilmente generalizado para adaptação a situações práticas. Concretamente, ao contrário dos trabalhos de Huang et al., o algoritmo deste artigo possibilita que se inclua, na determinação das recomendações, mais informação do que apenas a disponível nas transações comerciais, permitindo uma utilização mais alargada do mesmo, como se passa a descrever.

Seguindo uma generalização do algoritmo de *PageRank* inicialmente proposta em (Meyer et al., 2008), podemos alterar o passo iterador 4 do algoritmo da seguinte forma:

$$Cp^{(k)} = C^{(k-1)} \times (\beta_0 Mc + \beta_1 M1 + \beta_2 M2 + \cdots + \beta_m Mm) \qquad (5)$$

com $0 \leq \beta_i \leq 1$ para todo i e $\sum_{i=0}^{m} \beta_i = 1$.

As matrizes Mi serão matrizes $nc \times np$, normalizadas por linhas, construídas de forma a conterem informação útil para o sistema de recomendação. Em particular, podem conter informação relativa às preferências dos clientes, personalizando as recomendações ou serem construídas de modo a realçar determinados produtos em campanhas promocionais, melhorando assim a qualidade das recomendações. Os valores dos pesos β_i podem ficar acessíveis numa eventual aplicação, permitindo ao gestor da mesma adaptar o comportamento do algoritmo consoante os seus interesses.

O teste em contexto real do algoritmo foi realizado numa empresa de comércio online de t-shirts Marvelclick (http://dezpeme.com). Inicialmente esta empresa contava com uma base de dados em MySQL e uma página realizada em PHP. A empresa não dispunha de nenhum sistema de recomendação, embora fosse apresentado aos visitantes uma lista de produtos escolhidos aleatoriamente da base de dados. Para a implementação do algoritmo optamos pela linguagem Python, especialmente pela grande rapidez da sua biblioteca numérica Numpy e também pela simplicidade das suas bibliotecas de acesso a bases de dados, em particular a MySQLdb.

Antes de proceder a implementação do algoritmo, foram efetuadas simulações de desempenho de modo a testar exaustivamente a escalabilidade do algoritmo. Estes testes tiveram como objetivo a previsão do comportamento do mesmo em contexto real, onde se espera que as bases de dados tenham um crescimento contínuo, como é o caso das páginas web de comércio eletrónico. Visto que a implementação final do algoritmo iria utilizar uma base de dados em MySQL, optámos então, por realizar os testes também em MySQL usando a biblioteca MySQLdb, tendo assim a possibilidade de medir o desempenho do algoritmo, e também o desempenho do mesmo quando combinado com a base de dados. Assim, foi possível ter uma noção real do futuro desempenho do conjunto prevendo possíveis problemas.

Os testes foram realizados, simulando bases de dados com um determinado número de clientes e produtos onde o número de transações era escolhido aleatoriamente. Foram repetidas cinco vezes as simulações, para cada par de números clientes/produtos, e registado o tempo de consulta na base de dados – processamento do algoritmo – preenchimento da base de dados com as recomendações. Os testes foram realizados num computador com um processador Intel Core Duo CPU P8400 2.26GHz, com 4Gbyte de RAM e correndo num sistema operativo Ubuntu. Os tempos de execução registados foram compilados na figura seguinte:

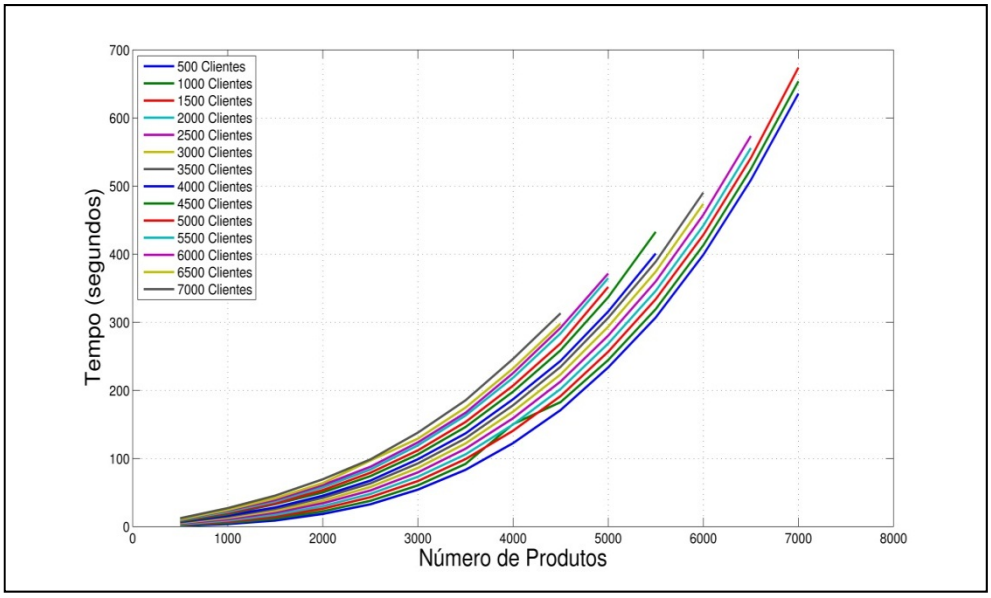


Figura 4.1: Gráfico dos Resultados dos Testes de Desempenho

Como se observa na figura anterior, podemos concluir que, numa implementação prática real, o algoritmo não poderá ser executado em tempo real, visto que o tempo de espera previsto seria da ordem das décimas de segundo, sendo um tempo de execução demasiado elevado para uma navegação rápida entre páginas. De facto, para empresas de dimensão pequena e média, como é a Marvelclick, a dinâmica de transações não é suficientemente rápida que exija uma atualização das recomendações em tempo real. Assim, é suficiente efectuar uma implementação que contemple uma atualização dos rankings de produtos guardados na base de dados com uma periodicidade da ordem das horas ou mesmo de dias. Desta forma, o algoritmo poderá correr em "back office" com uma frequência programável ou sempre que o administrador do *Website* o desejar, não decorrendo deste facto nenhuma desvantagem real do algoritmo proposto.

A implementação final consistiu num sistema que incluía o algoritmo generalizado proposto neste trabalho, assim como o interface que permitia a iteração do administrador com o mesmo. Neste interface foram definidas três modalidades de recomendações, no sentido de possibilitar um controlo da execução do algoritmo através de um interface intuitivo, permitindo adaptar a execução do mesmo a diversas

situações, tais como promoções ou alterações ao tipo de recomendação. A primeira modalidade é baseada nas compras efetuadas no website, a segunda tem em conta as preferências dos utilizadores e a última modalidade tem como base as escolhas do administrador. O controlo por parte do administrador advém do facto de que nem sempre a recomendação mais precisa é a economicamente mais viável. Por tal facto, foi dada a possibilidade ao administrador de definir pesos para todos os produtos que deseje, fazendo com que estes tenham maior probabilidade de recomendação. No interface, o administrador pode fazer pesquisas detalhadas de produtos, alterar o peso que os mesmos terão nas recomendações e influência que estes têm no algoritmo, alterando os valores dos pesos β_i, ou seja, escolhendo se as recomendações vão ser baseadas nas compras, nas preferências do administrador ou nas preferências dos clientes.

Com isto, foi possível obter um sistema que, para além de efetuar recomendações personalizadas potenciando as vendas, pode ter um papel fundamental em possíveis campanhas de marketing, assim como, no escoamento do stock residual.

5. Conclusões

Neste artigo foi proposto um algoritmo de recomendação para funcionar em aplicações de *E-commerce*. A qualidade das recomendações do mesmo foi testada, comparando o seu desempenho com outros algoritmos, usando uma base de dados comum. Os resultados dos testes realizados são muito promissores, indicando que o mesmo supera o desempenho dos algoritmos de recomendação clássicos.

Após provada a qualidade do algoritmo, foram desenvolvidas e testadas algumas funcionalidades mais vocacionadas para uma implementação em contexto empresarial de *E-commerce*. Devido à sua estrutura, o algoritmo pode ser facilmente generalizado para incluir vários tipos e níveis de informação, permitindo assim a sua utilização num leque alargado de cenários. Esta facilidade de generalização foi comprovada com uma implementação em ambiente real, no qual o algoritmo pode correr utilizando informação de transações anteriores, preferências dos clientes/utilizadores e preferências do gestor da empresa. Assim, neste trabalho foi apresentado um novo algoritmo de recomendação que, para além de registar muito bons resultados em testes controlados, é de fácil adaptação e implementação para funcionamento em lojas *on-line* reais.

Agradecimentos

Os autores agradecem os comentários e sugestões feitas pelo nosso colega Paulo Lousado e por dois revisores anónimos, que muito ajudaram a enriquecer a versão final do artigo.

Referências

Huang Z., Zeng D., Chen, H., (2005). A Link Analysis Approach to Recommendation under Sparse Data. Americas Conference on Information Systems.

Huang Z., Zeng D., Chen H., (2007). A Comparative Study of Recommendation Algorithms in E-commerce Applications, IEEE Intelligent Systems, 22(5), pp. 68-78.

Langville A., Meyer C., (2006). Google's PageRank and Beyond: The Science of Search Engine Rankings, Princeton University Press.

Li M. et al., (2009). Grocery Shopping Recommendations Based on Basket Sensitive Random Walk, KDD09, June 28 July 1, Paris, France.

Meyer C., Albright R., Govan A., (2008). Generalizing Google's PageRank to Rank National Football League Teams, Proceedings of SAS Global Forum 2008, San Antonio, TX, March 16-19, Article 151-2008.

Pucci A., Gori M., Maggini M., (2007). A Random-Walk Based Scoring Algorithm Applied to Recommender Engines , Lecture Notes in Computer Science - Springer.

Yildirim H., Krishnamoorthy M., (2008). A random walk method for alleviating the sparsity problem in collaborative filtering, RecSys 08: Proceedings of the 2008, ACM conference on Recommender systems, pp 131-138.

Zanker M., Jessenitschnig M., Jannach D., Gordea S., (2007). Comparing recommendation strategies in a commercial context, IEEE Intelligent Systems, Special issue on Recommender Systems, Vol. 22(3), pp. 69-73.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido / Recibido: 12/10/2011
Aceitação / Aceptación: 09/12/2011

SCODA para el Desarrollo de Sistemas Multiagente

Jesús A. Román[1], Dante I. Tapia [2], Juan M. Corchado [2]

zjarg@usal.es, dantetapia@usal.es, corchado@usal.es

[1] E. P. S. de Zamora (Universidad de Salamanca), 49022, Zamora, España.

[2] Universidad de Salamanca, 37008, Salamanca, España.

Resumen: El planteamiento de implementar sistemas multiagente de forma que puedan reutilizarse con distintos objetivos, dota a los desarrolladores de una metodología más eficaz en sus fines. Este artículo presenta SCODA, una arquitectura basada en cinco principios: estandaridad, especialización, facilidad de implementación, reutilización y computación distribuida. SCODA pretende facilitar el desarrollo de sistemas multiagente, basándose en el concepto de pequeños grupos de agentes denominados Comunidades Inteligentes Especializadas (CIE). Las CIE cuentan con funcionalidades específicas que permiten llevar a cabo la implementación de sistemas multiagente de forma escalable, dentro del marco de SCODA, pudiendo ser reutilizadas en diversos desarrollos.

Palabras clave: Agentes, Sistemas, Comunidades, Especialización, Reutilización.

Abstract: The approach to develop multi-agent systems that they can be reused for different purposes, provides to developers an effective methodology in their aims. In this paper is presented SCODA, an architecture based on five principles: standarization, specialization, facility of implementation, reuse and distributed computing. SCODA has intended to facilitate the development of multi-agent systems relying on the concept of small groups of agents, known as Intelligent Specialized Communities (CIE). The CIE have specific features that allow to carry out the scalable implementation of multi-agent systems, within the framework of SCODA and can be reused in several developments.

Key words: Agents, Systems, Communities, Specialization, Reuse.

1. Introducción

En general, las personas u organizaciones crean la capacidad necesaria para solventar los problemas que puedan producirse en un momento determinado, lo cuál es fácil pensar que existan otras entidades que dispongan ya de estas capacidades (Jensen, 1992). Este hecho, traducido a los sistemas computacionales, lo encontramos en la tecnología Orientada a Objetos (Elrad et al., 2001; Rashid et al., 2003; Gay et al., 2010; Langtangen, 2011), donde los objetos son encapsulados de forma independiente y

pueden ser reutilizados en diferentes desarrollos con finalidades muy distintas. El planteamiento de llevar a cabo este enfoque a los sistemas multiagente es, precisamente, uno de los objetivos de SCODA, en el cual se busca una arquitectura basada en *Comunidades Inteligentes Especializadas* (CIE) que puedan ejecutarse de forma independiente y tengan la capacidad de colaborar atómicamente entre ellas.

SCODA *(Specialized COmmunities for Distributed multiAgent systems)* es una arquitectura que se centra en el desarrollo de sistemas multiagente basándose para ello en cinco principios: estandaridad, especialización, facilidad de implementación, reutilización y computación distribuida. SCODA integra una o más Comunidades Inteligentes dotadas de especialización, que denominaremos **Comunidades Inteligentes Especializadas (CIE)**. Las Comunidades Inteligentes Especializadas tienen la capacidad de funcionar como un sistema multiagente independiente y especializado, donde los servicios que ofrece son ejecutados de forma distribuida, pudiendo colaborar con otras Comunidades Inteligentes Especializadas en la consecución de objetivos, que de forma individual no puedan alcanzar. Esta filosofía se basa en las *"Redes Empresariales"* (Sonquist y Koenig, 1975; Galaskiewicz, 1979; Borgatti y Foster, 2003; Viedma, 2004; Pöyhönen y Smedlund, 2004; Ibarra et al., 2005; Eraydin y Armatli-Köroglu, 2005; Cabus y Vanhaverbeke, 2006) a través de las cuales, un conjunto de grupos, instituciones u organizaciones interactúan entre ellos para obtener unos resultados favorables tanto a las unidades individuales como a la red en su conjunto.

SCODA puede ser implementada sobre cualquier plataforma para sistemas multiagente que soporte agentes BDI (Belief-Desire-Intention), los cuales tienen asociados una serie de estados mentales, como son creencias, deseos e intenciones (Bratman, 1988; Rao y Georgeff, 1995; Winikoff, 2009; Winikoff y Cranefield, 2010) mediante los que se implementa una capacidad de razonamiento que los hace idóneos para su integración en las Comunidades Inteligentes Especializadas. La implementación de SCODA se ha realizado sobre JADEX (Pokahr et al., 2003; 2007; Pokahr y Braubach, 2009), una extensión de la plataforma JADE (Bellifemine, 1999; Bellifemine et al., 2007) para agentes BDI, debido a que se considera como un motor de razonamiento y puede ser ejecutado de forma independiente, lo que implica una ventaja si se tiene en cuenta el enfoque distribuido que se pretende conseguir.

El presente artículo se estructura de la siguiente forma: En la sección 2 se describe la arquitectura SCODA, incidiendo en los aspectos más relevantes de la misma. La sección 3 describe la plataforma de agentes que integran SCODA. Finalmente, en la sección 4 se presentan las conclusiones obtenidas.

2. Descripción de SCODA

La mayor parte de los sistemas multiagente que se desarrollan tienen un diseño cerrado, es decir, ningún componente o agente externo tiene la capacidad de entrar a participar en el sistema, lo que implica que en la fase de diseño se ha de conocer el conjunto de los agentes que participan y las interacciones que se producen entre ellos (Zambonelli et al. 2000). En el desarrollo de sistemas abiertos la complejidad en cuanto a su diseño es mucho mayor, sobretodo en torno a las comunicaciones, ya que el diseñador no conoce a priori que tipos de componentes van a ser incluidos en el

sistema y por lo tanto la forma de comunicarse (Gonzalez-Palacios y Luck 2007). Es por ello que SCODA surge como alternativa intermedia entre estos dos enfoques, permitiendo la escalabilidad de Comunidades Inteligentes Especializadas, en tiempo de ejecución, para la consecución de objetivos complejos, que de forma individual no se alcanzarían.

La arquitectura SCODA se basa en cinco principios a través de los que se pretende una mayor eficiencia de los sistemas multiagentes desarrollados con la misma:

Estandaridad: A través de este principio se busca que a partir de SCODA, las diferentes Comunidades Inteligentes Especializadas que la conforman tengan una misma estructura que sea independiente de la finalidad que se persiga en el sistema multiagente que la implemente, es decir, la implementación de la estructura de una Comunidad Inteligente Especializada que tenga asociados determinados servicios ha de ser la misma que si los servicios fueran diferentes.

Especialización: Este principio se basa en el funcionamiento de las *"Redes Empresariales"* de forma que una Comunidad Inteligente Especializada persiga unas metas concretas dentro de un contexto determinado para lo cual, se especialice en la resolución de problemas determinados, y la cooperación entre Comunidades Inteligentes Especializadas de forma atómica, se resuma en la eficiencia que este tipo de cooperación genera en las organizaciones humanas.

Facilidad de Implementación: El diseño y la posterior implementación de un sistema multiagente es una labor costosa para los desarrolladores (Bellifemine et al. 2001), por lo que los sistemas construidos a partir de la cooperación de varias Comunidades Inteligentes Especializadas, dentro el marco de la arquitectura SCODA, han de ser fáciles de implementar. Esta finalidad se consigue ya que la estructura de las Comunidades Inteligentes Especializadas es estándar, siendo los servicios que ofrecen los que han de ser programados.

Reutilización: Debido a que dentro de SCODA cada Comunidad Inteligente Especializada es adoptada como un sistema multiagente independiente, el cual provee una serie de servicios especializados, es decir, orientados hacia unas metas concretas y comunes, la reutilización de estas Comunidades Inteligentes Especializadas ha de ser viable en cualquier desarrollo de sistemas multiagente que requiera los servicios que éstas puedan ofrecer, basado en la arquitectura SCODA. Esta visión se basa en el paradigma de la tecnología de orientación a objetos, los cuales son encapsulados y utilizados en diferentes desarrollos donde se requiera su funcionalidad.

Computación Distribuida: Este principio conlleva la necesidad de descentralizar, y por tanto distribuir la carga computacional de los agentes que conforman las Comunidades Inteligentes Especializadas de SCODA (Shen y Norrie, 1998; Camarinha-Matos y Afsarmanesh, 2007), de forma que los servicios requeridos no los presten directamente los agentes integrantes de las Comunidades Inteligentes Especializadas, sino que tendrán la capacidad de ejecutarse de forma distribuida para que la carga computacional asociada a los agentes disminuya y la estructura de la arquitectura no tenga variaciones.

Tal y como se muestra en la **Figura 1,** SCODA se define en seis módulos básicos: Aplicaciones Externas, Protocolo de Comunicaciones, Módulo de Control, Plataforma de Agentes, Comunidades Inteligentes Especializadas y Servicios de las Comunidades.

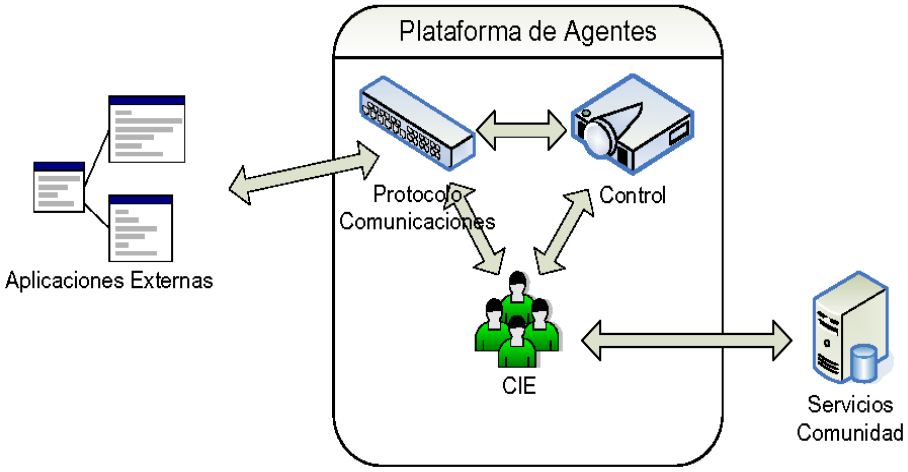


Figura 1 – Esquema estructural de SCODA

Las Aplicaciones Externas constituyen los programas y usuarios que hacen uso de SCODA solicitando los servicios que ofrece. El Protocolo de Comunicaciones es el encargado de atender las peticiones de las Aplicaciones Externas y solicitar una respuesta a las Comunidades Inteligentes Especializadas. El módulo de Control realiza un seguimiento de la coordinación y de las funcionalidades que la arquitectura provee atendiendo a una política de tolerancia a fallos en su funcionamiento. Las Comunidades Inteligentes Especializadas son el núcleo de SCODA, a través de las cuales se hacen efectivas las peticiones y respuestas de forma deliberada y optimizada. Los Servicios de la Comunidad se ejecutan de forma distribuida y es donde reside la capacidad de proceso de cada Comunidad Inteligente Especializada. Finalmente, la Plataforma de Agentes representa el entorno donde se ejecuta SCODA y está compuesta por los agentes que conforman la arquitectura y por la Comunidades Inteligentes Especializadas.

Los agentes que componen SCODA siguen el modelo deliberativo BDI y los servicios que ofrecen las Comunidades Inteligentes Especializadas son gestionados por este tipo de agentes. Concretamente uno de los agentes, el *PlannerAgent*, es el responsable de planificar que servicio es el óptimo y que parámetros son necesarios para que la solución demandada se ajuste a las necesidades del usuario o aplicación externa que realiza la petición. Otra característica aplicada a todos los agentes que componen SCODA, es que al ser agentes deliberativos BDI, éstos pueden hacer uso de mecanismos de razonamiento y técnicas de aprendizaje para realizar la gestión de funcionalidades y coordinación de las mismas, en función de las particularidades del contexto en que se ejecuten.

2.1. Aplicaciones Externas

Representan todos los programas que pueden utilizar las funcionalidades que provee el sistema multiagente implementados sobre SCODA. Estas aplicaciones son dinámicas y reaccionan de forma diferente de acuerdo a situaciones particulares como puede ser un gestor de agenda personal. Éstas pueden ser ejecutadas de forma local o remota, incluso desde dispositivos móviles con capacidad de proceso restringida, ya que las tareas que requieran una carga computacional alta, se realizarán de forma distribuida por los agentes que conforman SCODA a través de los servicios de las comunidades.

2.2. Protocolo de Comunicaciones

SCODA implementa un protocolo de comunicaciones basado en REST *(REpresentational State Transfer)* (Fielding, 2000; Griffin y Flanagan 2011), el cual permite a las aplicaciones externas comunicarse con los servicios ofrecidos por las Comunidades Inteligentes Especializadas a través de las mismas. El protocolo es completamente independiente de los lenguajes de programación utilizados, y está basado en peticiones sobre HTTP *(HyperText Transfer Protocol)* (RFC2616, 1999). Una petición HTTP *(HTTPRequest)* es enviada por las aplicaciones externas para especificar el servicio requerido, siguiendo el formato que se muestra a continuación:

Request = Simple-Request | Full-Request

Simple-Request = "GET" SP Request-URI CRLF

Full-Request = Request-Line

*(General-Header | Request-Header | Entity-Header)

CRLF

[Entity-Body]

En esta petición se informa de todos los parámetros necesarios para completar la tarea requerida. Todas las peticiones externas siguen las mismas pautas de comunicación y por lo tanto el mismo protocolo, mientras que las comunicaciones internas de la plataforma de agentes, siguen la especificación propia de la plataforma en la que se implemente SCODA. En el caso de JADEX se utiliza *FIPA Agent Comunication Language (ACL)* (FIPA, 2005). Finalmente la invocación de un servicio de una comunidad por parte de los agentes de la misma se realiza creando un nuevo hilo que se asocia a un socket, a través del cual se mantiene la comunicación abierta hasta que la tarea haya sido completada y el resultado se haya enviado a la Comunidad Inteligente Especializada que corresponde, o exista algún error en la ejecución de este servicio y por lo tanto también se informa del mismo. El formato de respuesta a la aplicación externa que solicita el servicio se realiza en forma de *HTTPResponse* como se muestra a continuación:

Response = Simple-Response | Full-Response

Simple-Response = [Entity-Body]

Full-Response = Status-Line

*(General-Header | Response-Header | Entity-Header)

CRLF

[Entity-Body]

Dentro del cuerpo de la *HTTPResponse*, y siguiendo la especificación REST se puede tener una representación de la respuesta en formato HTML, XML, etc (Muehlen et al., 2005).

La utilización de sockets en la invocación de servicios permite el procesamiento de múltiples solicitudes de forma simultánea. En el caso de procesamiento de servicios que requieran una alta carga computacional es la Comunidad Inteligente Especializada la responsable de realizar un balanceo de cargas invocando otro servicio que tenga la capacidad de satisfacer la petición. En la **Figura 2** se muestra un ejemplo de solicitud de un servicio a SCODA, a partir de la cual una aplicación externa realiza una petición a través de HTTP, siendo los agentes que conforman SCODA los responsables de seleccionar el servicio adecuado, llevar a cabo su ejecución y responder a la petición realizada.

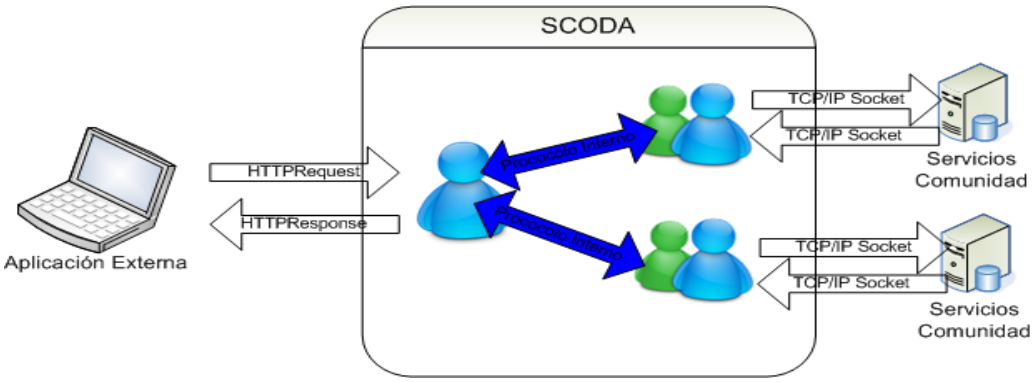


Figura 2 – Ejemplo de solicitud de un servicio a SCODA

2.3. Módulo de Control

SCODA implementa un mecanismo de control sobre los agentes integrantes de la misma, sobre los servicios, y sobre las comunicaciones realizadas. Así pues, a partir de este módulo, implementado por varios agentes, la arquitectura está preparada para asumir fallos en el funcionamiento de las aplicaciones y solventarlos de forma autónoma tomando las decisiones necesarias de forma deliberada, de forma que si alguno de los agentes activos en tiempo de ejecución tuviera un funcionamiento anómalo o dejara de funcionar, este módulo de control tiene la capacidad de reiniciarlos corrigiendo así su funcionamiento. También, en el caso de un mal funcionamiento de los servicios asociados a una Comunidad Inteligente Especializada, este módulo tiene la capacidad de buscar una réplica del servicio que pueda responder a una petición dada. Por otra parte, este módulo controla las comunicaciones realizadas entre las aplicaciones externas, los agentes integrantes de SCODA y los servicios asociados a las Comunidades Inteligentes Especializadas, de forma que si existiese algún tipo de comunicación anómala intenta su corrección reiniciando la misma, o en su caso reinicia todas las entidades implicadas en dicha comunicación.

2.4. Comunidades Inteligentes Especializadas (CIE)

Las Comunidades Inteligentes Especializadas son la esencia de SCODA. En ellas se encuentra la capacidad para distribuir y seleccionar el trabajo de forma inteligente a través de los agentes BDI que la componen. Constituidas por un agente controlador *(ComunityController)*, y un equipo de trabajo formado por un agente planificador y otro ejecutor *(PlannerAgent y ExecutorAgent)*, que son instanciados en tiempo de ejecución cuando son necesarios, y liberados al finalizar su trabajo. La propia arquitectura interna de la comunidad y la filosofía que se persigue en cuanto a la instanciación y liberación de agentes bajo demanda hace que SCODA sea una arquitectura distribuida en cuanto a sus servicios y eficiente en cuanto a la gestión de recursos internos de la plataforma que la ejecuta. Estas Comunidades Inteligentes Especializadas funcionan como sistemas autónomos, de forma que una implementación basa en SCODA puede estar compuesta por una o varias Comunidades Inteligentes Especializadas y hacer uso de los servicios que éstas proveen por separado, o utilizar varias Comunidades Inteligentes Especializadas y sus servicios de forma escalada y coordinada, y así tener la capacidad de resolver problemas mayores de forma colaborativa. A diferencia de otras organizaciones de agentes como las instituciones electrónicas (Arcos et al., 2005), las Comunidades Inteligentes Especializadas no constituyen un sistema multiagente abierto, sino pequeños sistemas multiagentes cerrados con capacidad de crecer de forma escalada y colaborativa en la consecución de objetivos más complejos, que de forma individual no puedan ser abordados, lo que implica una sencillez en el diseño de las mismas, ya que no es necesaria una fuerte normativa como en las instituciones electrónicas.

La especialización que estas entidades ofrecen de forma atómica es importante debido a que pueden ser reutilizadas en diversos desarrollos donde se requieran los servicios que provean.

2.5. Servicios de la Comunidad

Estos servicios representan las funcionalidades que ofrece la arquitectura. Los servicios de las comunidades son el grueso del procesamiento ofrecido por el sistema, los cuales son accedidos de forma ubicua y distribuida liberando a los agentes de SCODA de carga computacional y haciendo de la misma una arquitectura ligera. Los Servicios de la Comunidad se encuentran permanentemente activos para recibir peticiones de su Comunidad Inteligente Especializada correspondientes y están diseñados para su acceso de forma remota a través de sockets, debido a su facilidad de implementación y a su buen resultado en otros trabajos como son (Balaji et al. 2006; Douglas y Pai 2006; Sunwook et al. 2009), y están organizados por categorías en cuanto a su especialización, es decir los servicios relativos a una determinada funcionalidad están agrupados para su acceso por la comunidad especializada que ofrezca estos servicios, de esta forma se pretende buscar la eficiencia basada en la especialización, en forma de redes empresariales.

Los Servicios de la Comunidad, así como otro tipo de servicios distribuidos como los *Servicios Web y los Servicios Orientados a Arquitecturas* (Papazoglou et al. 2007) han de disponer de mecanismos de publicación y descubrimientos de servicios. También han de poseer un directorio actualizado de los mismos de forma que estén accesibles

cuando se requieran (Leymann et al. 2002; Papazoglou et al. 2007), es por lo que SCODA posee un directorio de servicios flexible desde el cuál pueden ser invocados de forma dinámica a través de las Comunidades Inteligentes Especializadas que los proveen. Sin embargo, la inserción, modificación o eliminación de los Servicios de la Comunidad se realiza de forma manual por razones de seguridad (López et al. 2006).

3 Plataforma de Agentes en SCODA

Este módulo es el núcleo de SCODA. Integra los agentes BDI que conforman las Comunidades Inteligentes Especializadas, y los agentes especiales de control y de gestión de las comunicaciones, cada uno de ellos con características especiales y comportamientos determinados. Estos agentes, a través de la inteligencia que poseen, actúan como administradores del sistema, de forma que sustentan el control de las comunicaciones, y administran el comportamiento del sistema.

Una ventaja que proporciona la arquitectura SCODA sobre los desarrollos que la implementan es que, las aplicaciones externas que hacen uso de los servicios que ésta provee pueden ser programadas en cualquier lenguaje de programación, con lo que aumenta la capacidad de utilización desde multitud de dispositivos. De esta forma, y como se muestra en la **Figura 3**, el acceso a los servicios ofrecidos por las Comunidades Inteligentes Especializadas puede realizarse desde dispositivos de sobre mesa, ordenadores portátiles, teléfonos móviles, PDAs, etc

A continuación se describen los diferentes agentes que integran la arquitectura SCODA y sus funcionalidades:

CommunicatorAgent: Este agente se responsabiliza de las comunicaciones externas con la plataforma. Recibe las peticiones de usuarios, aplicaciones o agentes externos y las transmite a las Comunidades Inteligentes Especializadas para su procesamiento, para ello mantiene un directorio de las Comunidades Inteligentes Especializadas y los servicios que estas ofrecen. También es el responsable de proporcionar las respuestas hacia el exterior. Este agente está constantemente en escucha a través de un *Server socket*. Se recibe una petición vía HTTP y es éste agente el que busca la comunidad adecuada y envía la petición a través del protocolo de comunicación interna de la plataforma. El hilo queda a la espera de obtener una respuesta para comunicarla al la entidad que realiza la petición. Paralelamente también se envía la respuesta al *QualityAgent,* el cual realiza las operaciones necesarias para mantener una estadística de funcionamiento, y en su caso, solucionar los problemas que se hayan podido plantear como la caída en el sistema de alguno de los agentes responsables de comunidad o incluso del propio *CommunicatorAgent*. Este agente se pertenece al Módulo de Control comunicándose con el *QualityAgent* para dar información a cerca de las comunicaciones realizadas en ambos sentidos, revisando así los mensajes recibidos, y además tiene la capacidad de inicializar al *QualityAgent* si se detecta un funcionamiento defectuoso del o inactividad en el mismo.

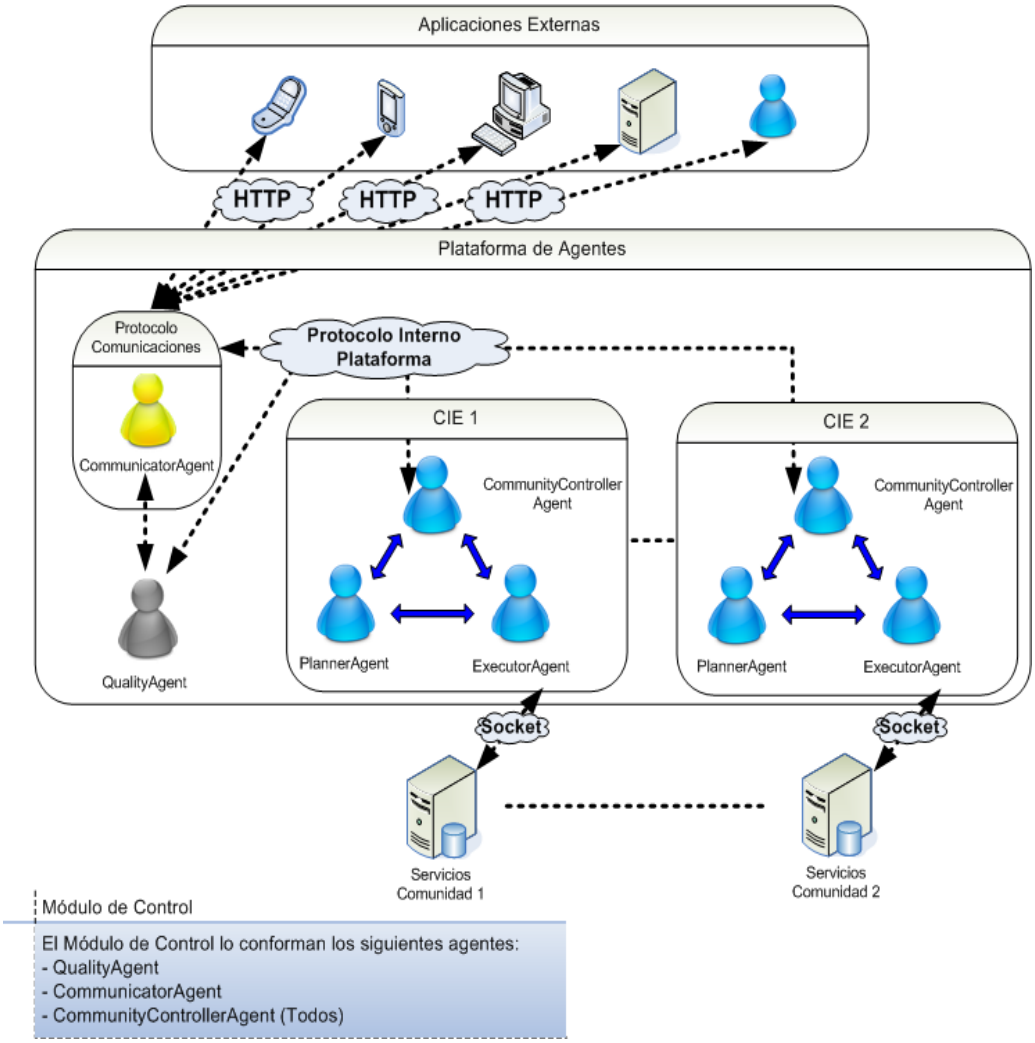


Figura 3 – Arquitectura SCODA

QualityAgent: Se responsabiliza del control del funcionamiento de los demás agentes, así como de las funciones que realizan. Se encuentra en constante comunicación con el *ComunicatorAgent* y con cada uno de los *ComunityControllers*. Tiene la capacidad de inicializar a los demás agentes que conforman SCODA, actuando sobre ellos en caso de un funcionamiento incorrecto. También realiza estadísticas de funcionamiento de las operaciones que se realizan, lo que implica que todos los movimientos operacionales son remitidos a éste agente en tiempo de ejecución.

CommunityControllerAgent: Este agente controla las Comunidades Inteligentes Especializadas. Una vez sido seleccionado por el *CommunicatorAgent* para resolver una petición externa, este agente crea un *"Equipo de Trabajo"* compuesto por un *PlannerAgent* y un *ExecutorAgent*, los cuáles son los responsables de ejecutar en sí

mima la petición. Este agente tiene la capacidad de crear y destruir al equipo de trabajo en tiempo de ejecución de forma que creará tantas instancias de los mismos como peticiones existan, y las irá eliminado una vez hayan finalizado su cometido, o cuando detecta un funcionamiento anómalo por parte de cualquiera de los agentes integrantes del *"Equipo de Trabajo"*. Este agente también pertenece al Módulo de Control, y se comunica con el *QualityAgent* para dar información a cerca de los servicios prestados y su resolución de forma que se pueda mantener un histórico de estadísticas de funcionamiento del sistema, y además tiene la capacidad de inicializar al *QualityAgent* si se detecta un funcionamiento anómalo o inactividad por parte de este agente.

PlannerAgent: Es uno de los componentes del *"Equipo de Trabajo"*, el cual es el responsable directo de la selección y ejecución del servicio adecuado que responda a una petición externa. Este agente es creado en tiempo de ejecución por el *CommunityController* cuando se ha de invocar un Servicio de la Comunidad y recibe la petición del propio *ComunityController* de su comunidad. Tiene la capacidad de seleccionar el servicio concreto de los que provee la comunidad y sus parámetros de forma que se optimice la ejecución del servicio, que posteriormente son comunicados al *ExecutorAgent*, el cual es el responsable de su invocación.

Se mantiene a la espera de un resultado por parte del *ExecutorAgent* lo cual confirma que el servicio requerido ha sido invocado y por lo tanto ha terminado el cometido del *"Equipo de Trabajo"*. Este agente tiene la capacidad de razonamiento que permite detectar el estado de los servicios en cada momento, para así seleccionar un servicio que no esté en ejecución y dotar de mayor agilidad a la Comunidad Inteligente Especializada.

ExecutorAgent: Este agente es el encargado de invocar al servicio pertinente para la resolución de una petición externa. La invocación se realiza a través de sockets, de forma que los servicios de cada Comunidad Inteligente Especializada pueden estar distribuidos de forma remota y por lo tanto la carga computacional repercute en la máquina donde se alberguen los servicios en sí. El *ExecutorAgent* se mantiene a la espera de una respuesta, y una vez recibida, la comunica de forma ascendente para advertir que su trabajo ya ha acabado.

4. Conclusiones

La filosofía seguida por SCODA permite implementar sistemas multiagente de una forma modular y escalada, dotando a los desarrolladores de la posibilidad de reutilizar las Comunidades Inteligentes Especializadas que la integran en diversos sistemas sin necesidad de volver a programar sus funcionalidades.

SCODA cuenta con una potente estructura de control, a partir de la cual se proporciona una completa capacidad en la gestión y recuperación de errores dentro del sistema y los servicios asociados al mismo, manteniendo en todo momento una estadística de funcionamiento, que permite adoptar decisiones y medidas en caso de un funcionamiento defectuoso del sistema y de los servicios asociados al mismo.

Otra de las ventajas que proporciona SCODA es que, junto a la metodología REST que adopta, permite la comunicación de forma estándar desde cualquier dispositivo y

aplicación independientemente del lenguaje de programación en la que esté desarrollada, o sistema operativo sobre la que se ejecute.

Para comprobar la validez del modelo teórico propuesto se ha llevado a cabo el desarrollo de un sistema multiagente, bajo el marco de SCODA, aplicado a la gestión logística de productos alimentarios. Este desarrollo se ha realizado en un entorno real empresarial, dedicado a la distribución de productos congelados. La metodología de trabajo de la empresa engloba tareas de gestión de inventarios, predicción de la demanda de productos y gestión de las rutas comerciales, lo cual hace idónea la filosofía modular que sigue SCODA permitiendo desplegar una Comunidad Inteligente Especializada por cada una de las tareas mencionadas.

Los resultados obtenidos, en cuanto al desarrollo del sistema, han permitido comprobar el correcto funcionamiento de las CIE que lo integran, tanto de forma individual, como de forma colaborativa, dando validez al modelo teórico propuesto. Sin embargo, en el caso de la tarea de predicción de la demanda de productos, los servicios desarrollados para mejorar los procesos seguidos por la empresa no han sido satisfactorios completamente, por lo que es necesario depurarlos con el fin de que el rendimiento del sistema alcance un mayor grado de optimización.

El desarrollo de este sistema se engloba dentro del trabajo de tesis doctoral *"Comunidades Inteligentes para la Construcción y Gestión de Arquitecturas Optimizadas de Sistemas Multiagente"*, que se está llevando a cabo por el autor principal. Una vez comprobada la validez del modelo teórico, se pretende llevar a cabo la creación de un repositorio de CIE de forma que se amplíe el campo de aplicación de SCODA en cuando al desarrollo de sistemas multiagente se refiere, lo que permitirá comprobar y optimizar el funcionamiento de SCODA en otro tipo de problemas. El desarrollo de estas CIE pretende ser abierto a la comunidad científica, de forma que se permita la aportación y recopilación de estas Comunidades Inteligentes Especializadas para su utilización.

Referencias bibliográficas

Arcos, J. L., Esteva, M., Noriega, P., Rodríguez, J. A., Sierra, C. (2005). Engineering open environments with electronic institutions. *Journal on Engineering Applications of Artificial Intelligence*. 18(2), 191-204.

Balaji, P., Bhagvat, S., Jin, H.-W., Panda, D.K. (2006). *Parallel and Distributed Processing Symposium. IPDPS.*

Bellifemine, F., Rimassa, G., Poggi, A. JADE - A FIPA-compliant Agent Framework. *In Proceedings of the 4th International Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agents, London, 1999.*

Bellifemine F., Poggi A., & Rimassa G. (2001). Developing multi agent systems with a fipa-compliant agent framework. *Software - Practice And Experience*, 31(2), 103–128.

Bellifemine, F., Caire, G., Greenwood, D. (2007). Developing Multi-Agent Systems with JADE. *John Wiley & Sons, NJ.*

Borgatti, S. Foster, P. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, *29*(6), 991-1013.

Bratman, M. E., Israel, D., Pollack, M. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence, 4, 349-355.*

Cabus, P. Vanhaverbeke, W. (2006). The territoriality of the network economy and urban networks: Evidence from flanders. *Entrepreneurship & Regional development*, *8*, 25-53.

Camarinha-Matos, L. M., Afsarmanesh, H. (2007). A Comprehensive Modeling Framework for Collaborative Networked Organizations. *Journal of Intelligent Manufacturing ,* 18 (5), 529-542.

Douglas, C., Pai, V.S. (2006). Seekable sockets: a mechanism to reduce copy overheads in TCP-based messaging. *Parallel and Distributed Processing Symposium, pp-6.*

Elrad, T., Filman, R. Bader, A. (2001), Theme Section on Aspect-Oriented Programming, *CACM,* 44(10), 29-32.

Eraydin, A. Armatli-Köroglu, B. (2005). Innovation, networking and the new industrial clusters: the characteristics of networks and local innovation capabilities in the Turkish industrial clusters, *Entrepreneurship and regional development, 17*, 237-266.

Fielding, R. (2000). *Architectural Styles and the Design of Network-based Software Architectures*, Doctoral Dissertation University of California, Irvine, CA.

FIPA. (2005). *Foundation for Intelligent Physical Agents*. Retrieved 7 14, 2006, from http://www.fipa.org

Galaskiewicz, J. (1979). *Exchange Networks and Community Politics*. Beverly Hills: Sage.

Gay, S., Vasconcelos, V., Ravara, A., Gesbert, N.,Caldeira, A. (2010). Modular session types for distributed object-oriented programming. *In Proceedings of the 37th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. 299-312.

Gonzalez-Palacios, J., Luck, M. (2007). Towards compliance of agents in open multi-agent systems. In *Software Engineering for Multi-Agent Systems V, Lecture Notes in Computer Science. Springer, 4408, 132-147.*

Griffin,K., Flanagan, C. (2011). Defining a call control interface for browser-based integrations using representational state transfer. *Computer Communications*. 4(2), 140-149.

Ibarra, H., Kilduff, M. Tsai, W. (2005). Zooming in and out: connecting individuals and collectivities at the frontiers of organizational network research. *Organization Science*, *16*(4), 359-371.

Jensen, M. C., Meckling, W. H. (1992). Knowledge, control and organizational structure. *Current Economics (Blackwell, Oxford). Lars Werin and Hans Wijkander, eds.* , 251-274.

Langtangen, P. (2011). Object-Oriented Programming. *Texts in Computational Science and Engineering*. 6, 437-496.

Leymann, F., Roller, D., Schmidt, M.-T. (2002). Web services and business process management. *IBM Systems Journal* 41(2), 198-211.

López, F., Luck, M., y d'Inverno, M. (2006). A normative framework for agent-based systems. *Computational and Mathematical Organization Theory,* 12,227–250.

Muehlen, M., Nickerson, J., Swenson, K. (2005). Developing Web Services Choreography Standards-The Case of REST vs. SOAP, *Decision Support Systems,* 40(1), 9-29.

Papazoglou, M. P., Traverso, P., Dustdar, S., Leymann, F. (2007). Service-Oriented Computing: State of the Art and Research Challenges. *Computer*. 40(11),38-45.

Pokahr, A., Braubach, L., Lamersdorf, W. (2003). Jadex: Implementing a BDI-Infrastructure for JADE *Agents. In EXP - in search of innovation (Special Issue on JADE),* 76-85.

Pokahr, A., Braubach, L., Walczak, A., Lamersdorf, W. (2007). Jadex - Engineering Goal-Oriented Agents. In Developing Multi-Agent Systems with JADE. *Eds., Wiley & Sons*, 254-258.

Pokahr, A., Braubach, L. (2009). From a Research to an Industrial-Strength Agent Platform: Jadex V2. *9. Int. Tagung Wirtschaftsinformatik*. 769-778.

Pöyhönen, A. Smedlund, A. (2004). Assessing intellectual capital creation in regional clusters. *Journal of Intellectual Capital*, *5*(3), 351-365.

Rao A. S. y Georgeff M. P. (1995) BDI Agents from Theory to Practice. *Proceedings of the First Int. Conference on Multi-Agents Systems. (ICMAS-95)*, 312-319.

Rashid, A., Moreira, A., Araujo, J. (2003). Modularization and composition of aspectual requirements. *In 2nd International Conference on Aspect-Oriented Software Development*, 11–20.

RFC 2616. (1999). Hypertext Transfer Protocol -- HTTP/1.1. *The Internet Society* 1999.

Shen, W., Norrie, D. H. (1998). An agent-based approach for distributed manufacturing and supply chain management. *In G. Jacucci (Ed.), Globalization of manufacturing in the digital communications era of the 21st century,* 579–590. Boston: Kluwer.

Sonquist, J. A., Koenig, T. (1975). Interlocking directorates in the top US corporations: a graph theory approach. *Insurgent Sociology*, *5*, 196-229.

Sunwook K., Chanho P., Seongwoon K., Yongwha C. (2009). The offloading of socket information for TCP/IP offload engine. *In 11th International Conference on Advanced Communication Technology, 1, 826-831.*

Viedma, J. (2004). Social capital benchmarking system: profiting from social capital when building network organizations. *Journal of Intellectual Capital*, *5*(3), 426-442.

Winikoff, M. (2009) Future Directions for Agent-Based Software Engineering. *In International Journal Agent-Oriented Software Engineering*. 3(4), 402-410.

Winikoff, M., Cranefield, S. (2010) On the testability of BDI agents. *8th European Workshop on Multi-Agent Systems (EUMAS2010).*

Zambonelli, F., Jennings, N. R., Wooldridge, M. (2000). Organisational abstractions for the analysis and design of multi-agent systems. *In 1st Int. Workshop on Agent-Oriented Software Engineering*, 127–141.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido / Recibido: 14/10/2011
Aceitação / Aceptación: 09/12/2011

Clasificación de textos en lenguaje natural usando la Wikipedia

Jose María Quinteiro-González [1,2], Ernestina Martel-Jordán [1,2], Pablo Hernández-Morera [1,2], Juan A. Ligero-Fleitas [1], Aaron López-Rodriguez [1]

{jquinteiro,emartel,pablo,jligero,alopez}@iuma.ulpgc.es

[1] IUMA Sistemas de Información y Comunicaciones. División Tecnología de la Información, Universidad de Las Palmas de Gran Canaria, Campus Universitario de Tafira, 35017 - Las Palmas de Gran Canaria, España.

[2] Departamento de Ingeniería Telemática, Universidad de Las Palmas de Gran Canaria, Campus Universitario de Tafira, 35017 - Las Palmas de Gran Canaria, España.

Resumen: La clasificación de textos, en entornos en los que el volumen de datos a clasificar es tan elevado que resulta muy costosa la realización de esta tarea por parte de humanos, requiere la utilización de clasificadores de textos en lenguaje natural automáticos. El clasificador propuesto en el presente estudio toma como base la *Wikipedia* para la creación del *corpus* que define una categoría mediante técnicas de *Procesado de Lenguaje Natural* (PLN) que analizan sintácticamente los textos a clasificar. El resultado final del sistema propuesto presenta un alto porcentaje de acierto, incluso cuando se compara con los resultados obtenidos con técnicas alternativas de *Aprendizaje Automático*.

Palabras clave: Categorización de textos; Wikipedia; tf-idf; Aprendizaje Automático; Procesado de Lenguaje Natural.

Abstract: Automatic Text Classifiers are needed in environments where the amount of data to handle is so high that human classification would be ineffective. In our study, the proposed classifier takes advantage of the Wikipedia to generate the corpus defining each category. The text is then analyzed syntactically using Natural Language Processing software. The proposed classifier is highly accurate and outperforms Machine Learning trained classifiers.

Keywords: Text Categorization; Wikipedia; tf-idf; Machine Learning; Natural Language Processing.

1. Introducción

Hoy en día la información se ha convertido en un recurso estratégico de primer orden para las organizaciones, que obtienen y almacenan grandes volúmenes de datos de diversas fuentes y de manera automática. En estos casos, los sistemas de clasificación

pueden ayudar a administrar, consultar y extraer información de grandes sistemas de documentos. La *categorización o clasificación de textos* (Sebastiani, 2005) consiste en asignar textos a una o varias categorías.

El proceso de clasificación de textos comienza con el indexado del documento, consistente en mapear el documento a una representación compacta de su contenido. Los métodos de indexación normalmente utilizados en la categorización de textos utilizan una representación del documento mediante el *Modelo del Espacio Vectorial* (Salton, Wong & Buckley, 1975), donde un documento en lenguaje natural se representa mediante vectores de términos.

Cada término del documento ocupa una posición en el vector de términos, originando un problema de dimensionalidad en documentos extensos. Por esta razón existe una fase previa que intenta reducir el tamaño del documento. Esta reducción hace el problema más manejable durante el aprendizaje, y repercute positivamente tanto en el tiempo de proceso como en el espacio ocupado. La reducción de la dimensionalidad del vector de términos de un documento normalmente se consigue mediante las siguientes acciones: tokenización, eliminación de *stopwords* y palabras nulas (que no aportan valor dentro del contexto del documento) y lematización.

Una vez reducida la dimensionalidad del documento, se calcula la relevancia de cada término del vector de términos. Para este fin, uno de los métodos más extendidos es la obtención del *tf-idf* (Salton & Buckley, 1988). El *tf-idf* comprende la frecuencia del término en el documento (*tf)* y la inversa de la frecuencia de documentos que poseen el término (*idf)*. El *tf* representa la importancia local que el término posee en el documento, es decir, cuanto más aparezca un término en un documento, más relevante será ese término para ese documento. El *idf* representa la importancia global de un término en relación inversa, es decir, cuantos más documentos incluyan el término, este término será menos relevante.

Para clasificar un documento se calcula la similitud entre el vector de términos característicos de la categoría y el vector de términos del documento. El ángulo que forman los dos vectores se usa como medida de divergencia y el coseno del ángulo se utiliza como valor de similitud: si el coseno vale 1, los vectores son idénticos; si vale 0 se trata de vectores ortogonales y no hay coincidencia entre ambos (el coseno del ángulo no puede tomar valores negativos por cuanto la medida de tf-idf siempre tiene valores mayores o iguales a cero).

El principal problema de modelar un documento mediante el *Modelo de Espacio Vectorial* es que los documentos se convierten en bolsas de palabras (*bag of words*). Esta aproximación tiene 3 inconvenientes cuando se aplica a la clasificación de textos (Wang & Domeniconi, 2008):

1. Si se separan palabras que reflejan un único concepto, se pierde el significado original.
2. Las palabras sinónimas se representan como diferentes dentro del *Modelo de Espacio Vectorial*, afectando a la frecuencia de aparición de las palabras. Dentro de este modelo, los términos *computador*, *ordenador* y *pc* se consideran términos sin relación.

3. Con palabras polisémicas se considera un único significado, cuando una misma palabra puede estar utilizándose con varios significados dentro del documento.

Existen diversas aproximaciones para enriquecer la bolsa de palabras, tratando así de solventar uno o varios de los problemas anteriormente citados. En este trabajo, se ha optado por usar la herramienta de PLN *FreeLing* (Atserias, Casas, Comelles, González, Padró & Padró, 2006) para extraer la información sintáctica de un texto, solventado en parte la separación de términos relacionados y el problema de la polisemia de las palabras.

En este artículo se presenta un proceso de clasificación de textos basado en la construcción de un vector de términos de cada categoría apoyándose en la Wikipedia.

A continuación, se indica el trabajo relacionado. En los apartados 3, 4 y 5 se presenta en detalle el proceso utilizado para la categorización de textos, los experimentos diseñados y los resultados obtenidos. Finalmente en los apartados 6 y 7 se presentan las conclusiones y el trabajo futuro.

2. Trabajo relacionado

El *Modelo de Espacio Vectorial*, siendo una aproximación válida a la clasificación de textos, presenta ciertos inconvenientes que se han intentado subsanar. Estos intentos han ido encaminados a enriquecer con conocimiento externo la bolsa de palabras, añadiéndole nuevos elementos.

En los últimos años, por el tamaño y notoriedad que ha alcanzado, ese conocimiento extra se ha buscado en la Wikipedia (Strube & Ponzetto, 2006), (Gabrilovich & Markovitch, 2006), (Chang, Ratinov, Roth & Srikumar, 2008), (Wang, Hu, Zeng & Chen, 2009). Pero el uso de la Wikipedia para la construcción de ontologías o de relaciones semánticas tiene una serie de obstáculos importantes. En general se considera que un artículo de la Wikipedia es un concepto *per se*, aunque esto no sea cierto siempre. Los artículos entre ellos tienen una estructura de enlaces entrantes y salientes, que permite construir una estructura de relaciones entre conceptos, ya sean estas relaciones jerárquicas o semánticas.

En (Strube & Ponzetto, 2006) se centran en medir la relación entre dos conceptos usando la Wikipedia. Aunque no se centra en la clasificación de texto, este trabajo si es interesante para resaltar que los artículos son descripciones de los conceptos a los que se refiere.

En (Gabrilovich & Markovitch, 2006) se relacionan los términos de los documentos a clasificar con conceptos de la Wikipedia. Para extraer los conceptos de la Wikipedia, se siguen una serie de pasos: se eliminan los artículos que no son conceptos, se eliminan las palabras raras (presentes en menos de tres artículos), *stop words*, y se realiza el stemming sobre el resto. Del texto a clasificar se toman sucesivamente diferentes elementos y con ellos se van buscando los conceptos relevantes. Posteriormente, la bolsa de palabras se ve enriquecida con estos nuevos conceptos y finalmente se decide la categoría a la que pertenece.

En (Gabrilovich & Markovitch, 2007) se introduce el *Explicit Semantic Analysis* (ESA). En este enfoque se emplean técnicas de clasificación de textos para representar explícitamente el significado de un texto en lenguaje natural en términos de un espacio multidimensional de conceptos Wikipedia. Básicamente elaboran un índice invertido de la Wikipedia, en el que cada palabra tiene asociada una lista de conceptos, también de la Wikipedia. Para clasificar un texto, se extraen sus términos y se construye un vector con pesos de conceptos relacionados. Se puede determinar la relación existente entre dos textos comparando los vectores de conceptos relacionados con sus términos.

En (Wang & Domeniconi, 2008) intentan resolver los inconvenientes que supone la representación de los documentos como meras *bolsas de palabras* mediante la inclusión de conocimiento procedente de la Wikipedia en un núcleo semántico que se utiliza para mejorar la representación de los documentos. Como principal aportación logran mantener como un único concepto los conceptos compuestos de más de un término, capturan la semántica de los sinónimos, y eliminan la ambigüedad de los términos polisémicos. Dado un texto a clasificar, primero elaboran una lista de *conceptos candidatos*, es decir, conceptos que están presentes en el texto y que se pueden mapear a conceptos de la Wikipedia. Con la lista de conceptos candidatos, se usa un tesauro elaborado con la Wikipedia para seleccionar conceptos relacionados semánticamente, aplicándose una medida de esta relación semántica según cada caso. Con una matriz de proximidad elaboran el *Modelo de Espacio Vectorial* extendido para el documento en cuestión.

En (Cui, Lu, Li & Chen, 2009) proponen un método para extraer conceptos de la Wikipedia sin proponer ningún clasificador. Simplemente se centran en encontrar los conceptos y separarlos de las instancias, cosa que no se hace en (Strube & Ponzetto, 2006) y (Gabrilovich & Markovitch, 2006). Pongamos por ejemplo que el concepto *empresa* tendría como instancia *Microsoft*. Lo interesante de este trabajo para nosotros es que usaron el Stanford POS Tagger, un software de PLN, para identificar los conceptos, analizando las frases que contienen elementos del tipo "*is a*","*type of*","*name of*", "*a kind of*" y "*one of*" para reconocer conceptos.

En este trabajo no se opta por la creación automática de ontologías o tesauros basados en la Wikipedia, ya que se desea dar al usuario la decisión sobre cómo construir sus categorías. Para ello, el trabajo está inspirado en (Chang, Ratinov, Roth & Srikumar, 2008) que usa las etiquetas como elemento para definir una categoría. Éstos, a su vez, se basan en (Gabrilovich & Markovitch, 2007) para incrementar el contenido semántico de un fragmento de texto. Básicamente, utilizan las etiquetas de las categorías como base para la clasificación, extrayendo, mediante ESA (*Explicit Semantic Analysis*), conceptos de la Wikipedia relacionados con dicho texto. Estos conceptos son utilizados posteriormente para determinar la pertenencia o no de un documento a una categoría, comparando vectores de conceptos.

Aún considerando que la descripción textual que se hace de una categoría es importante para la clasificación, en este artículo se opta por aprovechar la función sintáctica de las palabras dentro de un texto. Dada una o varias palabras, éstas se emplean para encontrar artículos de la Wikipedia donde dichas palabras son relevantes en función de su categoría gramatical (sustantivo, verbo y adjetivo).

3. Proceso de clasificación

Para el proceso de clasificación propuesto es necesario realizar el trabajo previo de preparación del contenido de la Wikipedia. Éste se encuentra disponible para su descarga vía Web. Debido a que no toda la información existente resulta útil para nuestro propósito, es necesario realizar un filtrado con objeto de identificar los artículos y desechar el resto de información.

Los artículos identificados se almacenan en una tabla, con objeto de utilizar las funcionalidades de indexación y búsqueda de texto completo de MySQL. MySQL mantiene un índice numérico para cada término (índice *Full-Text*) que constituye una variante del *tf-idf* (Oracle), y cuyo cálculo viene dado por la expresión:

$$W = \frac{\log(dtf)+1}{sumdtf} * \frac{U}{1+0.0115*U} * \log(\frac{N-nf}{nf}) \qquad (1)$$

donde *dtf* es el número de veces que un término aparece en el artículo, *sumdtf* es la suma de [log*(dft)*+1] para todos los términos del mismo artículo, *U* es el número de términos únicos en el artículo, *N* es el número total de artículos de la tabla y *nf* es el número de artículos que contienen el término.

A partir de este punto, el proceso de clasificación se puede dividir en dos etapas: una centrada en la creación de las categorías y otra dedicada a la clasificación de los textos en las categorías creadas. El proceso de clasificación global se muestra en la figura 1.

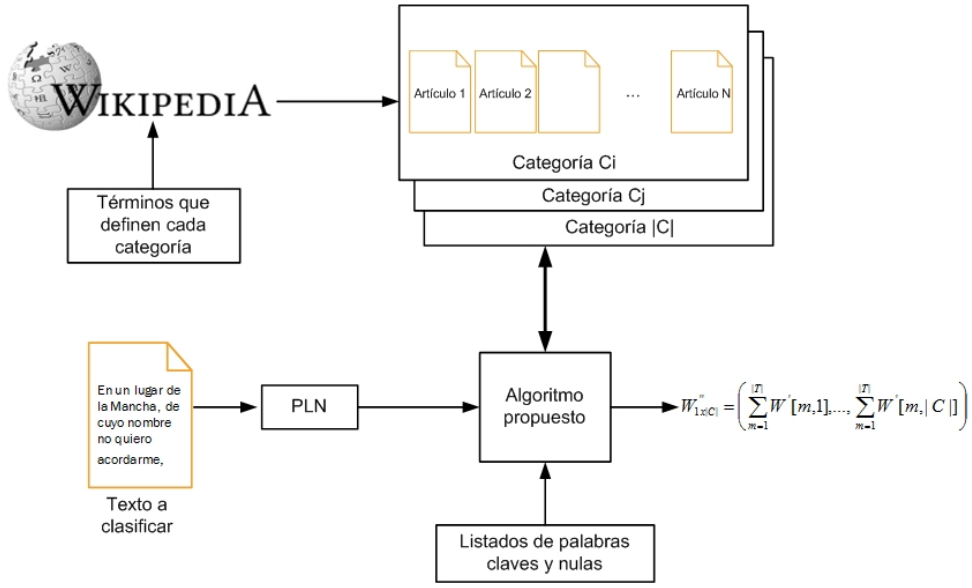


Figura 1 – Proceso global de clasificación

3.1. Creación de categorías

En este enfoque, el usuario aporta las categorías y algunos términos que considera relevantes para la definición de cada categoría. En general, lo importante no es que el usuario especifique un gran conjunto de palabras, sino un conjunto que permita desambiguar la categoría.

A partir de los términos de una categoría determinada, y apoyándose en la variante de *tf-idf* utilizada en (1), se extraen los *N* artículos de la Wikipedia en los que la relevancia de estos términos es elevada. La *relevancia* (*R*) de un término en un artículo se obtiene de la forma:

$$R = qf * w \tag{2}$$

donde *qf* representa el número de veces que el término se repite en la consulta, y *w* es el valor dado por la expresión (1).

A continuación se procesan los artículos extraídos mediante técnicas de PLN. Este procesado reduce la dimensionalidad en varias fases: se lematizan los términos encontrados reduciéndolos a una forma común, generalmente al masculino singular o al infinitivo. Posteriormente, se catalogan los elementos sintácticamente más relevantes: entidades o nombres propios, nombres comunes, verbos y adjetivos. Finalmente, se eliminan las palabras vacías y los signos de puntuación.

La capacidad de detección de nombres propios de la herramienta de PLN *FreeLing* se ha ampliado, considerando como un único nombre secuencias de caracteres como por ejemplo "*De la Rosa*", "*McLaren*", "*US Open*", "*ONU*".

Al final de este proceso, cada uno de los *N* artículos se reducen a un número términos con una representación homogénea. Los nombres propios se conservan como tales y el resto de términos relevantes comparten una representación similar en los distintos artículos. El conjunto de elementos identificados en cada uno de los *N* artículos conforman el conocimiento que el sistema tiene sobre esa categoría.

Cada categoría creada se almacena mediante una tabla, donde cada tupla almacena la información relativa a cada uno de los *N* artículos, esto es, la relevancia del artículo *n* en la categoría C_j, $R(n,C_j)$, y el conjunto de términos identificados para ese artículo. De este modo, el proceso de creación de categorías finaliza con tantas tablas como categorías se hayan creado.

3.2. Algoritmo de clasificación

La clasificación consiste en determinar a qué categoría, dentro de un conjunto predeterminado $\{C_1,\ldots, C_{|C|}\}$, pertenece un texto. Como paso previo al proceso de clasificación se debe procesar el texto con el fin de obtener los términos relevantes para el algoritmo propuesto. Este procesamiento transforma el texto en un vector de términos relevantes $\{t_1,\ldots,t_{|T|}\}$, mediante el mismo proceso y herramientas utilizadas con los artículos de la Wikipedia.

El peso de cada término t_i en una categoría C_j se obtiene mediante la siguiente expresión:

$$W(t_i, C_j) = \sum_{n=1}^{N} R(n, C_j) * P(t_i, n, C_j) \quad (3)$$

donde N es el número de artículos extraídos de la Wikipedia para crear cada una de las categorías, $R(n,C_j)$ es la relevancia del artículo n en la categoría C_j, y $P(t_i,n,C_j)$ representa el peso del término t_i dentro del artículo n, que viene dado por la expresión (1).

El valor obtenido $W(t_i,C_j)$ se modifica según el tipo de término t_i de que se trate, atendiendo a la siguiente clasificación: palabra especial, palabra nula, palabra normal o entidad. Las *palabras especiales* son aquellas específicas, que no son exclusivas, de cada categoría. Las *palabras nulas* son aquellas que el usuario considera que pueden introducir distorsión en los resultados finales. Las *entidades* se corresponden con nombres de personas y organizaciones, que tienen importancia para determinar la categoría a la que pertenece un texto. Las *palabras normales* son aquellas que no están en ninguno de los grupos anteriores. Así, el peso $W(t_i,C_j)$, se incrementa para las palabras especiales y las entidades en un factor 100 y 10 respectivamente, se anula para las palabras nulas, y se disminuye para las palabras normales en un factor 0,1, obteniendo el valor modificado $W'(t_i,C_j)$.

El cálculo anterior debe extenderse sobre todas las categorías existentes, y aplicarse a todos los términos relevantes del texto que se desea clasificar. De este modo, el resultado es la matriz:

$$W'_{|T|x|C|} = \begin{pmatrix} W'_{11} & \cdots & W'_{1|C1} \\ \vdots & \ddots & \vdots \\ W'_{|T|1} & \cdots & W'_{|T||C|} \end{pmatrix} \quad (4)$$

Finalmente la suma de los valores de cada columna de la matriz $W'_{|T|x|C|}$ representa la relevancia del texto en cada categoría:

$$W''_{1x|C|} = \left(\sum_{m=1}^{|T|} W'[m,1], \ldots, \sum_{m=1}^{|T|} W'[m,|C|] \right) \quad (5)$$

El sistema concluye que la categoría a la que pertenece el texto es la correspondiente al máximo valor de $W''_{1x|C|}$.

4. Resultados experimentales

Para realizar los experimentos se ha tomado como conjunto de prueba 206 titulares de noticias recogidas durante varios días consecutivos, de las categorías '*Fórmula 1*', '*Tenis*', '*Ciclismo*', '*Golf*' y '*Atletismo*' (ver tabla 1) del sitio de noticias deportivas www.marca.com/deporte/rss/index.html. En esta URL podemos encontrar las noticias y la categoría asociada.

Se ha considerado una categoría de control, '*Cultura*', para detectar comportamientos anómalos del clasificador.

Tabla 1 - Categorías y número de noticias por categoría

Categorías	# Noticias
Atletismo	40
Ciclismo	49
Motor	44
Golf	23
Tenis	50

En la tabla 2 se indican los términos indicados por el usuario para crear cada categoría, buscando en la Wikipedia los artículos en los que mayor peso tengan.

Las palabras que se han considerado clave para cada categoría son las que se indican en la tabla 2. Las palabras clave tienen un alto significado en sus categorías, pero no son exclusivas a éstas. Por ejemplo, *vuelta* puede aparecer en ciclismo o motor, pero se desea que se valore más en la categoría ciclismo.

Tabla 2 - Términos que definen cada categoría y palabras clave por categoría

Categoría	Términos	Palabras clave
Atletismo	Atletismo	Altura, longitud, metros, m, kilómetros, carrera, correr
Ciclismo	tour, ciclista	Tour, etapa, giro, crono, vuelta
Motor	F1, Fórmula 1	Pole, Fórmula 1
Golf	golf, golfista	Golf, PGA, green, golpes
Tenis	tenis, grand slam	Atp, tenis
Cultura	Cultura	

El listado de palabras nulas se indica en la tabla 3. Las palabras nulas son globales a todas las categorías. Estas palabras pueden introducir sesgos indeseables porque pueden repetirse más en ciertos artículos que en otros.

Tabla 3 - Lista de palabras nulas

Palabras nulas
clasificado, clasificación, líder, liderato, triunfo, primero, segundo, final, victoria, plaza, vuelta, tiempo, equipo, vencer, podio

4.1. Resultados de las pruebas

La evaluación del rendimiento de las pruebas se realizará mediante el *índice de Cohen*, corrección del porcentaje de aciertos que no computa aquellos que hayan podido ser fruto del azar (Japkowicz & Shah, 2011).

En todas las pruebas realizadas se han empleando 750 artículos de la Wikipedia para crear cada categoría. En las pruebas se ha aplicado el algoritmo propuesto, evaluando el impacto de la incorporación de las siguientes funcionalidades:

- La lista de palabras clave
- La lista de palabras nulas
- La modificación del peso del término en función de su tipo: entidad, palabra clave, palabra nula, otros

Experimento 1

Los resultados se obtuvieron aplicando el algoritmo propuesto a las noticias que forman nuestro grupo de prueba, junto con las funcionalidades indicadas.

La matriz de confusión obtenida es la de la Tabla 4. El valor del índice de Cohen es 0.9078, esto es, el clasificador puede alcanzar un 90,78% de acierto sin considerar aquellos que son fruto del azar. La categoría *cultura* no obtiene ningún resultado.

Tabla 4 - Resultados del experimento 1

		Clasificación del experimento 1							**Efectividad por categoría**		
		Atletismo	Ciclismo	Motor	Golf	Tenis	Cultura	Total	Precisión	Cobertura	F1-score
Categoría real	Atletismo	36	0	0	3	1	0	40	0,947	0,9	0,923
	Ciclismo	1	44	0	1	3	0	49	0,936	0,898	0,917
	Motor	0	3	40	1	0	0	44	1	0,909	0,952
	Golf	1	0	0	21	1	0	23	0,84	0,913	0,875
	Tenis	0	0	0	0	50	0	50	0,909	1	0,952
	Cultura	0	0	0	0	0	0	0	N/A	N/A	N/A
	Total	38	47	40	25	55	0	206			

Experimento 2

En esta prueba no se usa la lista de palabras clave ni la de palabras nulas. La matriz de confusión se muestra en la Tabla 5. El valor del índice de Cohen es 0.8522.

El resultado se degrada un 6,12%. Son 9 titulares más los que se clasifican mal frente al experimento 1. La categoría '*cultura*' no obtiene ningún resultado.

Tabla 5 - Resultados del experimento 2

		Clasificación del experimento 2							Efectividad por categoría		
		Atletismo	Ciclismo	Motor	Golf	Tenis	Cultura	Total	Precisión	Cobertura	F1-score
Categoría real	Atletismo	34	2	0	2	2	0	40	0,919	0,85	0,883
	Ciclismo	2	40	3	1	3	0	49	0,889	0,816	0,851
	Motor	0	2	40	1	1	0	44	0,93	0,909	0,92
	Golf	1	1	0	19	2	0	23	0,792	0,826	0,809
	Tenis	0	0	0	1	49	0	50	0,86	0,98	0,916
	Cultura	0	0	0	0	0	0	0	N/A	N/A	N/A
	Total	37	45	43	24	57	0	206			

Experimento 3

No se utilizan las listas de palabras claves y nulas, ni la modificación del peso del término en función de su tipo, de manera que todas las palabras serán consideradas iguales. La matriz de confusión que se obtiene se indica en la Tabla 6.

Tabla 6 - Resultados del experimento 3

		Clasificación del experimento 3							Efectividad por categoría		
		Atletismo	Ciclismo	Motor	Golf	Tenis	Cultura	Total	Precisión	Cobertura	F1-score
Categoría real	Atletismo	31	4	0	1	4	0	40	0,861	0,775	0,816
	Ciclismo	1	38	4	2	4	0	49	0,731	0,776	0,752
	Motor	2	3	34	4	1	0	44	0,895	0,773	0,829
	Golf	1	3	0	14	5	0	23	0,636	0,609	0,622
	Tenis	1	4	0	1	44	0	50	0,759	0,88	0,815
	Cultura	0	0	0	0	0	0	0	N/A	N/A	N/A
	Total	36	52	38	22	58	0	206			

El valor del índice de Cohen es 0.7222. Los resultados empeoran drásticamente: un 20,45% frente al experimento 1 y un 15,25% frente al experimento 2. Sigue sin haber titulares clasificados como pertenecientes a '*cultura*'.

Experimento 4

Se desea realizar la comparación del algoritmo propuesto con técnicas de *Aprendizaje Automático*. Como herramienta para esta tarea se ha usado el software Weka. El conjunto de entrenamiento se reduce a una propiedad por noticia estableciendo de esta forma las mismas condiciones que en el experimento 1, dado que es el más sencillo de los modelos y a la vez el que genera mejores resultados. Para la construcción del modelo se emplearon los algoritmos Bayes Multinomial y SVM (Máquinas de Vectores Soporte). El modelo fue validado mediante validación cruzada de las 206 noticias,

obteniéndose los resultados indicados en la tabla 7, donde no se ha realizado corrección del azar.

Tabla 7 - Resultados de las pruebas realizadas con validación cruzada

Prueba	% de acierto
Algoritmo propuesto (Experimento 1)	93,10%
Bayes Multinomial	91,26%
SVM	86,41%

Experimento 5

En los experimentos anteriores se han utilizado 750 artículos de la Wikipedia para crear las diferentes categorías. En esta prueba se evalúa la influencia del número de artículos en el resultado final. Para medir el efecto, se han mantenido las mismas condiciones que las utilizadas en el experimento 1, y se ha ido variando el número de artículos por categoría.

Los resultados se muestran en la figura 2. En el eje de abscisas se representa el número de artículos que se emplean. En el eje de ordenadas se representan dos magnitudes: número de aciertos y número de titulares asignados a '*cultura*'.

Los resultados muestran que 10 artículos son claramente insuficientes: 6 titulares son asignados a la categoría '*cultura*'. Al incrementar el número de artículos, va incrementándose progresivamente el número de aciertos, disminuyendo (hasta hacerse cero) el número de titulares asignados a '*cultura*'. Se observa un máximo sobre los 750 artículos. Incluir más artículos reduce el número de aciertos, pues se incrementa el vocabulario con palabras que no son realmente importantes en el corpus, y que se extraen de artículos cada vez menos relevantes.

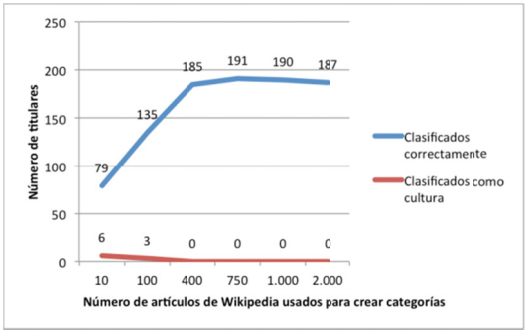


Figura 2 - Influencia del número de artículos en los resultados

5. Análisis de resultados

Del experimento 5 se observa un compromiso para escoger la cantidad de artículos para crear el corpus de las categorías. Un número pequeño de artículos no producen el vocabulario necesario para clasificar correctamente, y un número excesivo de artículos

introducen *ruido*, pues el vocabulario obtenido deja de ser plenamente representativo de la categoría.

Del resultado de los experimentos 1, 2 y 3, se observa que los mejores resultados se obtienen considerando las palabras clave y nulas, así como la modificación del peso del término en función de su tipo. En términos relativos, la modificación del peso del término en función de su tipo tiene un mayor impacto en los resultados, que la utilización de las listas de palabras especiales y nulas.

Los resultados obtenidos con las técnicas de *Aprendizaje Automático* son inferiores a los obtenidos con el algoritmo propuesto (91,26% de la prueba con Bayes Multinomial frente al 93,10% del algoritmo propuesto).

6. Conclusiones

Se ha creado un sistema de clasificación automático de textos que puede cumplir con las expectativas de efectividad que un usuario pudiera esperar en su uso. En el experimento 1 se comprueba que, dadas unas condiciones concretas, se alcanzan unas medidas del 93,10% de acierto en la clasificación, en una situación de partida desfavorable: se toman noticias de deportes, todas ellas compartiendo vocabulario.

El presente clasificador, muestra una forma más de salvar la visión que deja el *Modelo de Espacio Vectorial* de los documentos como meras *bolsas de palabras*, al valorar la función sintáctica que tiene una palabra dentro del texto a la hora de computar su peso.

Hemos de concluir, a la vista de los experimentos, que el rendimiento del clasificador está en relación con la configuración: dependiendo del número de artículos de la Wikipedia que se tomen o de los listados de palabras que se incluyan. El clasificador propuesto debe ser configurado según cada necesidad o la cercanía conceptual de los elementos a clasificar. El número de artículos seleccionados influye directamente en dos aspectos: el espacio que ocupan las categorías en las unidades de almacenamiento y el tiempo que se tarda en computar un texto en cada categoría.

Por otro lado, las listas de palabras clave y nulas tienen un impacto del 6,12% en el índice de aciertos entre el experimento 1 y 2, lo que indica que la capacidad de influencia de estas listas es limitada, aunque su aportación es deseable por cuanto incrementan el porcentaje de acierto. La mayor parte de los aciertos recae en el *Modelo de Espacio Vectorial* enriquecido con el análisis sintáctico propuesto.

Los resultados existentes en la literatura científica en su mayoría se basan en diferentes corpus en inglés, siendo más escasos los disponibles en español. En (Venegas, 2007) se utiliza una muestra de 222 artículos en español con unos resultados cercanos en fase de prueba al 76,74% con SVM y de 68,18% con Naive Bayesiano, aunque los resultados no son directamente comparables pues tanto el corpus de datos como las categorías elegidas (Química Industrial, Ingeniería en Construcción, Trabajo Social y Psicología) son diferentes.

Igual que dos humanos pueden no estar de acuerdo al clasificar ciertos titulares, el clasificador automático, en determinados momentos, también hace *interpretaciones*, aunque estas sean de base matemática. Pensemos que el titular "*Jaime Alguersuari y Sergio García, amigos y solidarios en un torneo de golf benéfico*" estaba originalmente

clasificado como *motor*. Pero la palabra *golf* junto con el nombre *Sergio García*, tienen mayor peso que *Jaime Alguersuari*, por lo que el titular se clasifica en la categoría *golf*, contabilizándose como un error.

7. Líneas futuras

En esta propuesta, el corpus es la fuente de la que emana el conocimiento que se emplea en la clasificación. Por lo tanto, la principal línea de actuación sería trabajar en el estudio del corpus de Wikipedia, introduciendo la autoconfiguración del clasificador mediante desambiguación supervisada.

Hay que estudiar especialmente aquellas palabras compartidas por dos o más categorías. Dentro de estas palabras habrá que encontrar la manera de diferenciar cuándo esa palabra es realmente determinante en el corpus.

Por último, sería interesante incluir reconocimiento de patrones sintácticos que ayuden a reconocer expresiones lingüísticas recurrentes (Cruz, Troyano, Enriquez & Ortega, 2008). En la propuesta actual, únicamente se incluyen para el reconocimiento de entidades.

Agradecimientos

Este trabajo ha contado con la financiación del Fondo Europeo de Desarrollo Regional (FEDER) y el Ministerio de Industria, Turismo y Comercio (MITYC) a través del Plan Avanza I+D (TSI-020302-2008-115).

Referencias bibliográficas

Atserias, J., Casas, B., Comelles, E., González, M., Padró, L. & Padró, M. (2006). FreeLing 1.3: Syntactic and Semantic Services in an Open-source NLP Library. In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006)*, 48-55.

Chang, M.-W., Ratinov, L., Roth, D. & Srikumar, V. (2008). Importance of Semantic Representation: Dataless Classification. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, 830-835.

Cruz, F. L., Troyano, J. A., Enriquez, F. & Ortega, J. (2008). Experiments in Sentiment Classification of Movie Reviews in Spanish. In *Sociedad Española de Procesamiento del lenguaje Natural*, 41, 73-80.

Cui, G., Lu, Q., Li, W. & Chen, Y. (2009). Mining Concepts from Wikipedia for Ontology Construction. In *Proceedings of the 2009 IEEE/WIC/ACM international Joint Conference on Web Intelligence and Intelligent Agent Technology* - Vol. 03. Web Intelligence & Intelligent Agent. IEEE Computer Society, Washington, DC, 287-290.

Gabrilovich, E. & Markovitch, S. (2006). Overcoming the Brittleness Bottleneck using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge. In

Proceedings of the 21st National Conference on Artificial Intelligence (AAAI), Boston, 1301–1306.

Gabrilovich, E. & Markovitch, S. (2007). Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, *IJCAI'07*, Hyderabad, India, 1606–1611.

Japkowicz, J. & Shah, M. (2011). Evaluating Learning Algorithms. A Classification Perspective. NY, USA: Editor Cambridge University Press. ISBN: 978-0-521-19600-0.

Oracle Corporation. MySQL Internals Algorithms - MySQL Forge Wiki. *MySQL Forge*. [En línea].

Quinlan, R. (1993). C4.5: Programs for Machine Learning. San Mateo, California: Editor Morgan Kauffman. ISBN: 978-1-55860-238-0.

Salton, G., Wong, A. & Buckley, C. (1975). A Vector Space Model for Automatic Indexing. *Communications of ACM, 8*(11), 613-620.

Salton, G. & Buckley, C. (1988). Term Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, 24(5), 513-523.

Sebastiani, F. (2005). Text Categorization. In Alessandro Zanasi (Ed.), *Text Mining and its Applications*. Southampton, UK: Editora WIT Press, 109-129.

Schapire, R. E., Freund, Y. & Schapire, R. (1996). Experiments with a new boosting algorithm. *International Conference on Machine Learning*, 148–156.

Strube, M. & Ponzetto, S. (2006). WikiRelate! Computing Semantic Relatedness Using Wikipedia. *Association for the Advancement of Artificial Intelligence*, 1419–1424.

Venegas, R. (2007). Clasificación de textos académicos en función de su contenido léxico-semántico. *Revista signos*, 40(63), 239-271. ISSN 0718-0934.

Wang, P. & Domeniconi, C. (2008). Building Semantic Kernels for Text Classification using Wikipedia. In Proceedings of the *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, USA, 713-721.

Wang, P., Hu, J., Zeng, H.-J. & Chen, Z. (2009). Using Wikipedia Knowledge to Improve Text Classification. *Knowledge and Information Systems, 19*, 265-281.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido / Recibido: 17/10/2011
Aceitação / Aceptación: 09/12/2011

Análise de opiniões expressas nas redes sociais

Diogo Teixeira, Isabel Azevedo

1070370@isep.ipp.pt, ifp@isep.ipp.pt

GILT, ISEP-IPP, Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal

Resumo: As redes sociais são cada vez mais utilizadas no nosso dia-a-dia. O recente aumento de popularidade deste tipo de serviço veio trazer novas funcionalidades e aplicações. Os utilizadores contribuem com as suas opiniões e conhecimentos, formando um repositório de informação de grandes proporções. Esta informação é cada vez mais utilizada por empresas, que vêem nas redes sociais uma forma de promover os seus produtos junto do público ou analisar de que forma os mesmos são considerados. O estudo apresentado neste artigo aplicou técnicas de Análise Sentimental para verificar se a informação existente em duas redes sociais (Facebook e Twitter) pode ser utilizada para estimar valores que podem vir a ser obtidos na comercialização de bens ou serviços a serem lançados no mercado.

Palavras-chave: Redes Sociais; Processamento de Linguagem Natural; Análise Sentimental.

Abstract: The social networks have been increasingly used. Their popularity brought new features and applications. In social networks users contribute with their opinions and knowledge, forming a huge information repository. The use of this information by companies, which consider social networks as a way of promoting their products, has been rising. This study, through the use of Sentimental Analysis, sustain the conclusion that the information obtained from social networks (Facebook and Twitter) can be used to determine values that can be obtained in the commercialization of goods or services to be launched in the market.

Keywords: Social Networks; Natural Language Processing; Sentimental Analysis.

1. Introdução

A quantidade de informação disponível tem aumentado, sendo cada vez mais comum a pesquisa de notícias em sítios Web, por exemplo, em vez de se proceder à consulta das mesmas nos meios de comunicação tradicionais. Desde muito cedo a Internet possibilitou a exposição e a troca de opiniões. Posteriormente, com o surgimento de

fóruns e blogues, as pessoas passaram a ter um espaço de partilha de opiniões mais personalizado.

Mais recentemente surgiram as redes sociais que adicionaram ao conceito inicial de conexão com os amigos, a facilidade de partilha de informação com os mesmos ou com todos os restantes utilizadores do serviço. As redes sociais mais utilizadas atualmente, Facebook e Twitter (Alexa, 2011), têm uma forte componente de partilha e troca de informação sobre os mais diversos assuntos, desde notícias a opiniões sobre produtos.

Com o aumento da importância das redes sociais, surgiu o interesse por parte de diversas empresas, que viram nas mesmas uma forma de promover os seus produtos e obter opiniões do público sobre os mesmos, algo que anteriormente exigia a realização de sondagens, com os custos inerentes às mesmas. Após a obtenção dos dados é realizado o seu tratamento de modo a obter as opiniões expressas nas mensagens e a sua polaridade, através de várias técnicas de processamento de linguagem natural e análise sentimental.

Este artigo descreve um trabalho de investigação que utiliza a informação disponível nas redes sociais para verificar como determinados produtos são percebidos pelos utilizadores e analisa a relação com os valores de comercialização obtidos na sua primeira semana no mercado. Foram escolhidos filmes de cinema lançados no mercado norte-americano, devido ao grande número de mensagens que geram, e pelo facto de grande parte delas indicarem opiniões dos utilizadores sobre os filmes. Duas redes sociais foram consideradas no estudo, o Facebook e o Twitter, pela abrangência e popularidade das mesmas.

O resto deste artigo está organizado conforme se explica em seguida. Na segunda secção descrevem-se as duas redes sociais estudadas, sendo exploradas as suas características. Na terceira secção é introduzido o conceito de análise sentimental. Na quarta secção é apresentado o problema, sendo posteriormente indicada a abordagem seguida e feita uma breve explicação da solução desenvolvida. Na quinta secção são apresentados e analisados os resultados obtidos. Na sexta secção apresentam-se as conclusões gerais e o trabalho futuro.

2. Redes Sociais

Uma rede social pode ser descrita como um conjunto de relações e intercâmbios entre entidades (indivíduos, grupos ou organizações) que partilham interesses, geralmente através de plataformas disponíveis na Internet.

A ideia básica de um SNS (*Social Networking Site*) é permitir às pessoas ter um espaço próprio, onde podem colocar dados pessoais que as caracterizam e relacionarem-se com outros utilizadores, mesmo que não tenham qualquer tipo de relação anterior com os mesmos. O propósito principal não é conhecer pessoas estranhas, mas sim permitir a conexão e troca de informação com pessoas que já fazem parte da sua rede social. O primeiro SNS, denominado SixDegrees (http://www.sixdegrees.com), surgiu em 1997.

Ao longo dos anos as redes sociais foram evoluindo, trazendo novas funcionalidades e diferentes formas de interacção com os amigos. A troca de informação entre utilizadores tornou-se cada vez mais relevante, sendo possível através de mensagens privadas ou públicas, ou até troca de mensagens em tempo real (*instanting*

messaging). Tornou-se possível partilhar todo o tipo de dados, seja texto, fotos, vídeo e até música.

A informação disponibilizada em redes sociais é cada vez mais usada com outros propósitos, seja prever a opinião dos utilizadores sobre factos ou produtos (Ku, Ke, & Chen, 2009), ou identificação de comunidades e interesses (Java, Song, Finin, & Tseng, 2007). Assim, as sondagens de opinião pública, normalmente feitas com recurso a chamadas telefónicas e entrevistas, podem ser complementadas ou substituídas por uma minuciosa análise das mensagens colocadas pelos utilizadores nas redes sociais, que estão disponíveis sem custos, recorrendo às suas APIs (*Application Programming Interface*).

Um exemplo de aplicação foi o estudo feito recorrendo às mensagens do Twitter para determinar a opinião política do público e a opinião pública em determinados assuntos (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). Os dados eram compostos por 1 bilião de mensagens relacionadas com as eleições presidenciais de 2008 dos Estados Unidos da América e com a opinião do público sobre os primeiros meses de governação, recolhidas entre 2008 e 2009 recorrendo à API do Twitter. Essas mensagens eram posteriormente analisadas para determinar se a opinião exposta era positiva ou negativa.

Para saber se os dados recolhidos nas mensagens do Twitter eram o reflexo da opinião pública, foi feita uma comparação com as duas principais sondagens realizadas no país que fazem a medição dos níveis de confiança do público, de periodicidade mensal. Foram também usados os dados de outra sondagem, com periodicidade diária. Para obter resultados sobre a opinião política do público em geral foram usadas duas sondagens, uma realizada em média a cada 3 dias, e outra que pretendia saber se o público iria votar em John McCain ou Barack Obama. Após a análise das mensagens recolhidas, conseguiram-se obter resultados equivalentes aos das sondagens na opinião pública para determinados assuntos.

Dentro das centenas de redes sociais em funcionamento actualmente, destacam-se algumas como o Facebook (http://www.facebook.com) e o Twitter (http://www.twitter.com), pela facilidade de partilha de informação para toda a gente, ou apenas para a rede pessoal do indivíduo.

2.1. Facebook

O Facebook começou em 2004 como uma rede social apenas para os alunos da Universidade de Harvard. Devido ao seu sucesso, ainda no mesmo ano expandiu-se para outras universidades, e para escolas do ensino secundário no ano seguinte. Finalmente, em 2006, o Facebook tornou-se uma rede social aberta a todos.

De momento o Facebook é a maior rede social do mundo, estimando-se que tenha mais de 600 milhões de utilizadores[1]. É também a segunda página Web mais visitada (Alexa, 2011).

[1] Dados de Outubro de 2011, obtidos de http://www.facebook.com/press/info.php?statistics.

Nesta rede social a interacção entre os vários utilizadores é feita através de páginas criadas pelos mesmos. Uma página desta rede social pode ser o perfil de um utilizador, uma página de fãs, uma página específica de um serviço, ou um evento, entre outros.

A extracção de informação para esta rede social é feita através da Graph API, sendo possível aceder a toda a informação inserida pelos utilizadores no Facebook, desde que se tenham as devidas permissões. O mecanismo de permissões utilizado pelo Facebook é o padrão aberto OAuth 2.0 (Hammer-Lahav, Recordon, & Hardt, 2011).

2.2. Twitter

O Twitter surgiu em 2006 como uma rede social que permite aos seus utilizadores partilharem informação, colocando mensagens (*tweets*) que tenham no máximo 140 caracteres. Ao contrário de outras redes, como o Facebook, o Twitter foca-se nas mensagens transmitidas entre utilizadores, sendo o perfil algo secundário. Estima-se que tenha actualmente 200 milhões de utilizadores activos, sendo trocados mais de 200 milhões de *tweets* diariamente (Twitter, 2011). Ocupa a nona posição entre as páginas Web mais visitadas (Alexa, 2011).

No Twitter os utilizadores podem seguir outros utilizadores, de modo a receber todas as mensagens que os mesmos escrevam. Os seguidores são denominados de *Followers*, sendo o ato de seguir outro utilizador denominado de *Following*. Devido à menor relevância dos perfis, na sua globalidade os tweets (mensagens dos utilizadores) são públicos, ficando acessíveis a qualquer pessoa através de pesquisas simples.

A Tabela 1 resume as diferenças entre a Streaming API, a REST API e a Search API (estas duas do tipo REST), mencionando as vantagens e desvantagens de cada uma.

Tabela 1 – Comparação entre as APIs do Twitter

API	Vantagens	Desvantagens
REST API	Implementação simples, não necessita de autenticação.	Não permite efectuar pesquisas, retornando apenas os últimos dados inseridos.
Search API	Implementação simples, permite obtenção de informação mais antiga, não necessita de autenticação.	Algum atraso no retorno dos Tweets, limitações de tráfego, não faz um pré-processamento dos Tweets.
Streaming API	Pesquisa feita em tempo real, sem limitações de tráfego, faz pré-processamento dos Tweets.	Não permite a obtenção de Tweets mais antigos, necessita de autenticação.

3. Análise Sentimental

A análise sentimental é bastante utilizada na determinação da opinião global que o público em geral tem sobre um determinado assunto ou produto, sem realização de sondagens ou entrevistas. Note-se que as mesmas, além de terem um custo relativamente elevado, são mais demoradas, permitem obter apenas a opinião sobre o que é perguntado e dependem da disponibilidade do público.

No entanto, muitas vezes para obter bons resultados na análise sentimental torna-se necessário implementar técnicas anteriores à análise propriamente dita que facilitam a

posterior procura de polaridade num texto. Algumas dessas técnicas vão ser descritas em seguida.

Caso o texto seja escrito de forma informal, poderá ser necessário um pré-processamento da frase, de modo a corrigir erros ortográficos e de pontuação que posteriormente iriam dificultar a procura de informação relevante.

O processamento de linguagem natural pode envolver diversas etapas, desde a divisão do texto em termos mais simples (*tokenizer*) até aos mais complexos como a análise sintática (*phrase chunking)* ou a identificação da classe gramatical das palavras (*part-of-speech tagging*). Alguns destes processos, como o *tokenizer*, são essenciais para o processamento correto do texto, existindo também outros mais específicos que, em certos contextos, podem ser úteis, com identificação de entidades, datas, nomes, entre outros elementos.

As técnicas de análise sentimental podem ser agrupadas em duas categorias: técnicas simbólicas e de aprendizagem máquina (Boiy, Hens, Deschacht, & Moens, 2007). As técnicas do primeiro tipo são mais simples, caracterizando-se pela avaliação das palavras recorrendo a recursos léxicos, para determinar se o seu sentido é positivo ou negativo, podendo, no entanto, estabelecer relações entre os diversos componentes e assuntos da frase.

Por sua vez, as técnicas de aprendizagem máquina, que constituem um ramo da Inteligência Artificial, procedem ao reconhecimento de comportamentos contidos em dados através de modelos treinados. Estes modelos são criados a partir de dados já processados e procedem à identificação de padrões idênticos nos dados a avaliar.

No trabalho de (Pang, Lee, & Vaithyanathan, 2002) foram comparadas três algoritmos distintos de aprendizagem máquina: Máxima Entropia, Naïve Bayes e Máquinas de Vetores de Suporte, na análise de polaridade relativa a vários filmes. Os resultados obtidos foram positivos, mas persistiram algumas dificuldades na determinação da relação entre os tópicos das mensagens e as opiniões relacionadas.

4. Problema e Solução Desenvolvida

Neste trabalho de investigação foram analisados 10 filmes (Tabela 2) e as respetivas mensagens recolhidas nos 7 dias anteriores às suas estreias.

Tabela 2 - Filmes Analisados

Filme	Data de Estreia	Nº Mensagens Analisadas - Facebook	Nº Mensagens Analisadas - Twitter	Resultados de Bilheteira (USD)
Thor	6 Maio 2011	3.168	101.472	$60M
Piratas das Caraíbas	20 Maio 2011	7.603	122.319	$90,1M
The Hangover II	27 Maio 2011	1.382	292.398	$86,5M
Kung Fu Panda	27 Maio 2011	4.873	43.919	$48M
X-Men: 1st Class	3 Junho 2011	2.475	127.989	$56M
Beginners	3 Junho 2011	95	1.477	$141.340

Filme	Data de Estreia	Nº Mensagens Analisadas - Facebook	Nº Mensagens Analisadas - Twitter	Resultados de Bilheteira (USD)
Submarine	3 Junho 2011	23	840	$41.832
50/50	30 Setembro 2011	627	19.893	$8,6M
What's Your Number	30 Setembro 2011	169	9.594	$5,4M
Dream House	30 Setembro 2011	232	4.754	$8,1M

Assim, a amostra extraída das duas redes sociais estudadas consiste em mensagens sobre filmes escritas pelos utilizadores na semana que antecedeu a sua estreia. Os valores de bilheteira foram obtidos através do serviço IMDB (The Internet Movie Database, disponível em http://www.imdb.com).

A solução desenvolvida foi implementada obedecendo a várias fases de desenvolvimento. Um esquema das funcionalidades da aplicação e das ferramentas utilizadas pode ser visto na Figura 1.

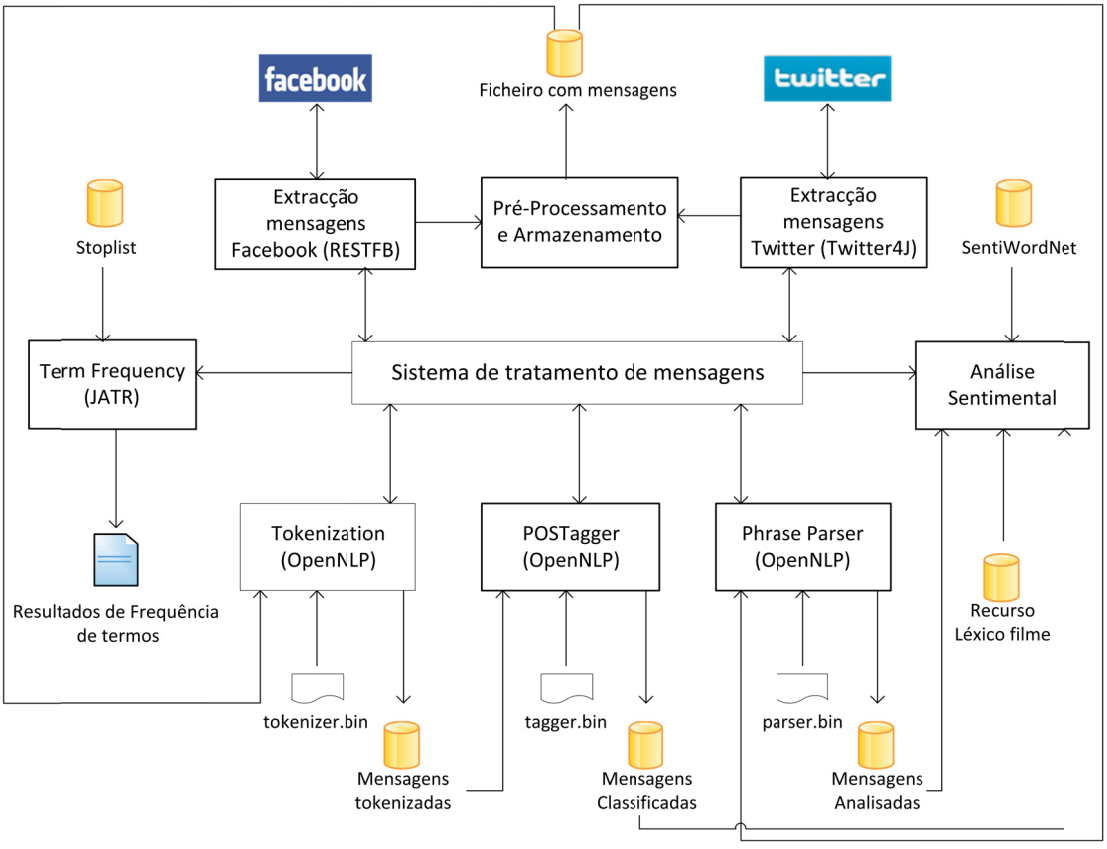


Figura 1 - Métodos e Ferramentas utilizados

A implementação da aplicação foi dividida em quatro fases distintas:

- 1ª Fase: Extracção, pré-processamento e armazenamento de mensagens;
- 2ª Fase: Processamento de linguagem natural;
- 3ª Fase: Análise sintática;
- 4ª Fase: Análise sentimental.

4.1. Extracção, pré-processamento e armazenamento de mensagens

A primeira fase inicia-se com a extracção de comentários com recurso às APIs das duas redes sociais e aplicações-cliente em linguagem java destas APIs: o RestFB[2] para o Facebook e o Twitter4J[3] para o Twitter. Dentro das APIs disponíveis, foi utilizada a Graph API do Facebook, e para as pesquisas de mensagens relevantes no Twitter, a Search API. Esta última foi escolhida em detrimento da mais recente Streaming API, visto permitir a obtenção de comentários mais antigos sem a necessidade de autenticação. O retorno dos resultados das duas redes sociais é feito em JSON (JavaScript Object Notation), um formato de fácil leitura e manipulação.

No Facebook, devido à existência de páginas de vários tipos, é feita antes da extracção uma pesquisa pelas páginas que contêm a informação pretendida, sendo avaliadas pelo seu nome e assunto a que se referem. Deste modo não é feita a extracção de mensagens de páginas que, apesar de partilharem o seu nome com o de um filme, não se referem ao mesmo.

Note-se que no Facebook, devido ao uso do padrão aberto de autenticação OAuth 2.0, todos os pedidos de informação à API têm de ser validados com um Access Token, sendo necessário o registo da aplicação no Facebook Developers[4] para a obtenção do mesmo. Por sua vez, no Twitter, devido à não existência de páginas pessoais, a extracção é feita de forma directa, com palavras-chave referentes ao filme e que são frequentes nas mensagens.

Após a obtenção de mensagens é realizado o seu pré-processamento, que altera a escrita informal, as abreviaturas e a ênfase dada às palavras através da repetição de caracteres (por exemplo, “goooooood”, em vez de “good”), bastante comuns nas redes sociais, mas que originam uma avaliação errada pelas técnicas habituais. O problema das abreviaturas foi resolvido com uma pesquisa pelas mais comuns, sendo feita em seguida a sua correção. Já para questão da ênfase das palavras pela repetição de um ou mais caracteres foram utilizadas expressões regulares para a identificação e correção das palavras com esse tipo de grafia. Assim, “goooooood”, por exemplo, passa a “good” com este pré-processamento das mensagens. As mensagens são armazenadas em ficheiros de texto logo após o pré-processamento.

2 http://restfb.com/

3 http://twitter4j.org

4 http://developers.facebook.com/

4.2. Processamento de linguagem natural

Na segunda fase é realizado o processamento de linguagem natural, que tem o objectivo de facilitar a posterior análise sentimental das mensagens. Para a implementação das técnicas de processamento de linguagem natural foi utilizada a ferramenta OpenNLP (Baldridge, 2005).

Como explicado na secção anterior, existem diversas técnicas de processamento de linguagem natural, umas mais específicas e outras mais genéricas. Neste projecto foram implementados dois processos considerados relevantes, *tokenization* e o *POS Tagger* (*Part of Speech Tagger*). O processo de *tokenization* realiza a divisão do texto em *tokens*, de modo a facilitar o tratamento posterior de cada token de forma independente. Para a realização desta operação é necessário importar os modelos treinados da ferramenta OpenNLP, que contêm as regras necessárias para a correcta execução dos diversos métodos de processamento.

Também é determinada a classe gramatical de cada componente das frases analisadas (*POS Tagger*), sendo necessária a importação do modelo treinado do OpenNLP para uma correta avaliação. Desta forma é possível identificar elementos importantes na frase, como os substantivos e adjectivos, para posteriormente estabelecer a correta polaridade da frase. Segue-se um exemplo[5] do resultado deste processo aplicado à uma mensagem específica:

```
Greatest->NNP
movie->NN
of->IN
all->DT
time->NN
!->.
```

4.3. Análise sintática

Na terceira fase é realizada a Análise Sintática das mensagens, com a utilização do componente Phrase Parser da ferramenta OpenNLP. O Phrase Parser faz a análise das mensagens, identificando os vários assuntos presentes na frase e quais os elementos (sujeitos, adjectivos, entre outros) que se relacionam com esse mesmo assunto, fazendo o seu agrupamento. Com os resultados deste processo é possível analisar apenas a parte da frase que interessa, atribuindo-lhe a polaridade correta na fase posterior de análise sentimental e ignorando o restante. Em seguida pode-se visualizar uma frase após a sua análise sintática:

[5] As diferentes classes gramaticais existentes podem ser consultadas em http://bulba.sdsu.edu/jeanette/thesis/PennTags.html.

```
(TOP (NP (NP (NNP Greatest) (NN movie)) (PP (IN of) (NP (DT all) (NN time))) (. !)))
```

4.4. Análise sentimental

Por fim, na quarta e última fase é realizada a análise sentimental das mensagens extraídas que tem como resultado a obtenção da sua polaridade. Para a sua realização, todos os processos descritos anteriormente nesta secção de desenvolvimento são necessários, visto procederem ao tratamento das mensagens que serão agora analisadas de modo a ser obtida a sua polaridade.

Neste método é analisada uma mensagem de cada vez, sendo dada maior relevância aos substantivos, que expõem os assuntos da frase e aos adjectivos, que os caracterizam como positivos ou negativos. Durante a análise das mensagens são tidos em conta os diversos assuntos da mesma, sendo verificado se estão relacionados com o filme pretendido. Caso estejam, os termos são avaliados, tendo em conta a sua classe gramatical. Estas duas operações dependem de uma base de dados e de um recursos léxico, que vão ser descritos em seguida.

A base de dados consiste em vários ficheiros de texto que contêm termos relacionados com cada filme, permitindo identificar nas mensagens os termos que estão associados ao filme e cuja polaridade será avaliada. Estes termos foram obtidos através do serviço IMDB, que permitiu recolher todo o tipo de informação sobre os mesmos, como o nome dos actores e personagens e os diversos assuntos abordados no filme, que poderiam dar origem a discussões nas redes sociais. Note-se que é importante saber se o que está ser comentado tem relação com o filme em análise ou não.

São também inseridos nesta base de dados termos comuns relacionados com o mundo do cinema, como "*trailer*" ou "*script*". Por sua vez, os valores de polaridade dos adjectivos são obtidos através do recurso léxico SentiWordNet (Baccianella, Esuli, & Sebastiani, 2010), uma variante da base de dados de palavras da língua inglesa Wordnet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). No exemplo abaixo é possível visualizar um extracto do recurso léxico SentiWordNet:

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
n	00340463	0	0	make#2	the act of mixing cards haphazardly
n	05845140	0.5	0	make#1	a recognizable kind; "what make of car is that?"
v	00012267	0	0	make#19	act in a certain way so as to acquire; "make friends"

A estrutura de cada termo no SentiWordNet é composta por cinco elementos principais, sendo eles a classe gramatical (POS), o ID do termo, o valor da polaridade positiva e negativa, e por fim o termo, seguido de uma pequena descrição e/ou de um exemplo numa frase. É de realçar que a mesma palavra pode ter várias classes gramaticais e significados com diferenças nos valores de polaridade.

Note-se que o processo de análise implementado não é infalível, visto ser possível a existência de frases relacionadas com os filmes com termos que não se encontram na base de dados, não sendo a sua polaridade avaliada. No entanto, foi feito um esforço para incluir o maior número de termos possível, de modo a esta limitação ser atenuada. São também avaliados os *emoticons* utilizados.

5. Resultados obtidos

Nesta secção apresenta-se uma análise dos dados obtidos através do processo de Análise Sentimental, sendo utilizada a correlação de Spearman para determinar a existência de uma relação entre alguns parâmetros e o valor de bilheteira para cada filme na sua semana de estreia. Trata-se de uma medida de correlação não paramétrica que mede o grau de relação entre duas variáveis.

A Tabela 3 apresenta o valor de correlação entre o número de mensagens positivas, negativas e neutras, divididas por rede social, e o valor de bilheteira.

Tabela 3 - Correlação entre o número de mensagens e o valor de bilheteira

Polaridade da mensagem	Rede Social	Valor de correlação (r_s)
Positiva	Facebook	0,88
Negativa	Facebook	0,89
Neutra	Facebook	0,89
Positiva	Twitter	0,93
Negativa	Twitter	0,95
Neutra	Twitter	0,98

Os valores encontrados permitem concluir que a correlação entre o número de mensagens positivas, negativas e neutras da rede social Twitter é bastante forte, atingindo valores entre 0,93 para o número de mensagens positivas, e 0,98 para o número de mensagens neutras. Para a rede social Facebook os valores são um pouco mais baixos, entre 0,88 para o número de mensagens positivas e 0,89 para o número de mensagens neutras.

O valor de correlação encontrado entre o número total de mensagens do Facebook e o valor de bilheteira é 0,89, e 0,93 para o Twitter. Assim, pode-se considerar que, na amostra estudada, quanto mais um filme for comentado, maiores são os seus resultados de bilheteira.

Também foi analisada a utilização de determinados termos com polaridade, considerada apenas a frequência de utilização dos termos, e sem análise sentimental das mensagens envolvidas. A Tabela 4 apresenta os valores de correlação entre a frequência de utilização de dois termos específicos com polaridade ("*good*" e "*bad*"), divididos por rede social, e os valores de bilheteira.

Tabela 4 - Correlação entre a frequência de utilização de determinados termos e o valor de bilheteira

Frequência de utilização de termo com polaridade	Rede Social	Valor de correlação (r_s)
Positiva	Facebook	0,34
Negativa	Facebook	0,25
Positiva	Twitter	0,75
Negativa	Twitter	0,71

Verificou-se assim que a correlação dos termos mais utilizados com polaridades positivas e negativas obtém valores mais fracos para a rede social Facebook (0,25 e 0,34, respetivamente) que para o Twitter (0,75 e 0,71, respetivamente). Estes valores mais baixos podem ser justificados pela sua limitação a uma contagem da frequência dos termos, não sendo realizada uma análise sentimental.

No entanto, apesar de não serem tão altos como os obtidos com a consideração da polaridade das mensagens, os valores para a rede social Twitter são significativos, sendo possível concluir que as mensagens do Twitter provavelmente refletem melhor a opinião do público que as do Facebook, possivelmente por serem mais curtas, o que incentiva os utilizadores a expressarem as suas opiniões em poucas palavras.

Na Figura 2 pode-se visualizar quais os 15 termos com polaridade positiva ou negativa mais utilizados nas mensagens do Twitter para o filme Kung Fu Panda II. Note-se que apenas pelos comentários terem determinados termos não é possível concluir se está a ser discutido o filme, bem como se o termo tem outro que o precede e lhe altera a polaridade (como em "*not good*", por exemplo).

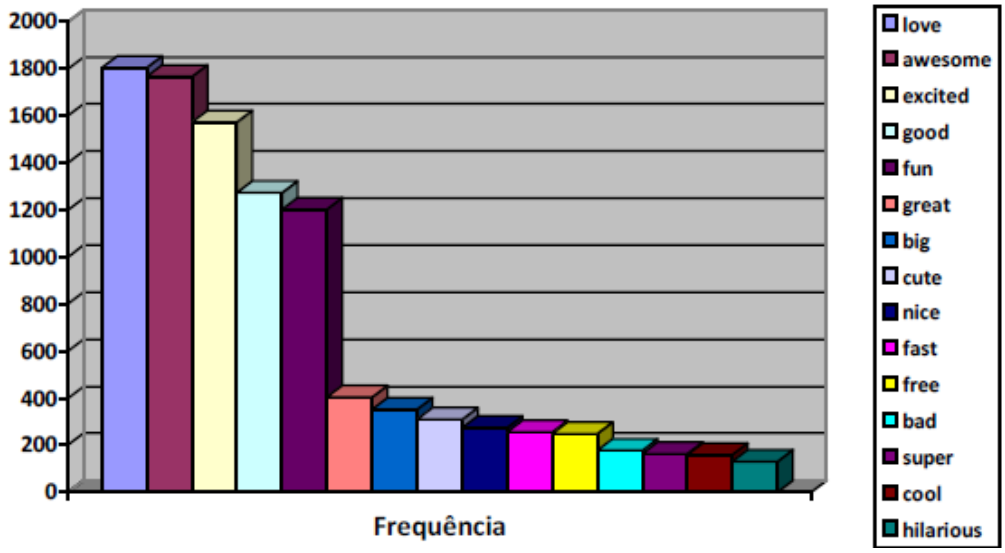


Figura 2 – Os termos com polaridade mais utilizados para o filme Kung Fu Panda II

Todos os valores de correlação obtidos têm significado ao nível de 5% se superior a $r_{s(0,05;10)}$=0,564 (valor crítico tabelado). Assim, pode-se afirmar que foram encontrados valores de correlação entre pares de variáveis que não são apenas casuais.

6. Conclusões

Neste artigo descreveu-se uma abordagem de análise sentimental que incidiu em mensagens disponibilizadas em redes sociais, tendo sido encontradas correlações significativas entre as percentagens de mensagens positivas e negativas e os valores de bilheteira conseguidos na semana de estreia de determinados filmes, entre outras.

Uma das vantagens da utilização da informação disponibilizada pelos utilizadores nas redes sociais é o grande volume de mensagens que se obtém, potenciando uma maior segurança nas conclusões obtidas através da análise das mesmas. A título comparativo, a amostra utilizada no estudo de (Pang, Lee, & Vaithyanathan, 2002) era composta por 2.053 críticas extraídas do serviço IMDB, escritas por 144 utilizadores. No trabalho mencionado em (Wu & Huberman, 2010), que pretendeu analisar as diferenças nas opiniões dos utilizadores ao longo do tempo, são analisados 407.557 comentários sobre 1.275 diferentes filmes. No trabalho desenvolvido e documentado neste artigo foram analisadas 745.302 mensagens sobre 10 filmes.

Os seguintes aspetos serão considerados no futuro próximo:

- A expansão da análise a outras línguas, nomeadamente a língua portuguesa, de modo a poderem ser analisados outros mercados. O recurso léxico subjetivo desenvolvido para a língua portuguesa descrito em (Rodrigues, 2009) poderá servir de base para este trabalho;
- A construção de um recurso léxico mais orientado ao tema "filmes", com base no recurso SentiWordNet que foi utilizado neste trabalho. A expansão a outras áreas é também uma possibilidade em análise;
- Dentro do mesmo tema será explorada a possível relação entre os comentários dos utilizadores e os valores de bilheteira nas semanas seguintes à estreia dos filmes, quando um grande número de utilizadores já terá visualizado os mesmos.

Pode-se concluir, com base nos valores encontrados na avaliação da amostra, que é possível a recolha, a análise e também a utilização dos comentários feitos nas redes sociais para se percecionar como determinados produtos estão a ser considerados pelos utilizadores e apoiar decisões.

Referências

Alexa. (2011). *Alexa Top 500 Global Sites*. Retrieved October 2011, from http://www.alexa.com/topsites.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, *Proceedings of the Seventh Conference*

on International Language Resources and Evaluation (pp. 2200–2204): European Language Resources Association.

Baldridge, J. (2005). *The OpenNLP Project*, from http://opennlp.sourceforge.net/.

Boiy, E., Hens, P., Deschacht, K., & Moens, M.-F. (2007). *Automatic sentiment analysis of on-line text.* Paper presented at the The 11th International Conference on Electronic Publishing, Vienna, Austria.

Hammer-Lahav, E., Recordon, D., & Hardt, D. (2011). *The OAuth 2.0 Authorization Protocol. IETF draft (work in progress)*, from http://tools.ietf.org/html/draft-ietf-oauth-v2-19.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities, *Procedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*. San Jose, California , USA.

Ku, L., Ke, K., & Chen, H. (2009). *Opinion Analysis on CAW2.0 Datasets*. Paper presented at the Content Analysis in Web 2.0 (CAW 2.0) Workshop, 21st April 2009, Madrid, Spain.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Introduction to WordNet: An On-line Lexical Database. *Journal of Lexicography, 3*(4), 235-244.

O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. (2010). *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.* Paper presented at the The International AAAI Conference on Weblogs and Social Media, Washington DC.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment Classification using Machine Learning Techniques.* Paper presented at the The Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia.

Rodrigues, D. H. (2009). *Construção automática de um Dicionário Emocional para o Português.* Master thesis, Universidade da Beira Interior, Covilhã.

Twitter. (2011). *Your world, more connected.* Obtido de Twitter Blog: http://blog.twitter.com/2011/08/your-world-more-connected.html.

Wu, F., & Huberman, B. A. (2010). Opinion Formation Under Costly Expression. *ACM Transactions on Intelligent Systems and Technology 1*(1), 1-13.

risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido / Recibido: 17/10/2011
Aceitação / Aceptación: 09/12/2011

Análisis de Decisiones en la Selección de Proveedores de Tecnologías de la Información: Una Revisión Sistemática[1]

Lucas Grossi[1], Jose A. Calvo-Manzano[1]

lgcgrossi@gmail.com, joseantonio.calvomanzano@upm.es

[1] Facultad de Informática, Universidad Politécnica de Madrid, 28660, Boadilla del Monte, España.

Resumen: En este artículo se aplica un protocolo de revisión sistemática de Ingeniería de Software para la decisión en la selección de proveedores de Tecnologías de la Información (TI). El objetivo es la búsqueda de documentos relacionados con la toma de decisión en la selección de proveedores de TI. Además, la revisión sistemática se centra en identificar las iniciativas y los informes de la toma de decisión en la selección de proveedores de TI. Los resultados obtenidos demuestran que existe una necesidad de más estudios y definen 16 criterios de evaluación para utilizarse en la selección de proveedores de TI.

Palabras clave: Selección de Proveedores; Análisis de Decisiones; Outsourcing; Revisión Sistemática.

Abstract: This paper applies a Systematic Review protocol for Software Engineering to IT outsourcing supplier selection decision. The objective is to search for papers related to IT outsourcing supplier selection decision. Furthermore, the systematic review is focused in order to identify initiatives and reports of IT outsourcing supplier selection decision proposals. Results obtained show that there is a need for more in-depth studies.

Keywords: Supplier Selection, Decision Analysis, Outsourcing, Systematic Review.

1. Introducción

Las actividades de negocios de hoy dependen en gran medida del apoyo de las Tecnologías de la Información (TI), por lo que su mantenimiento es un proceso crítico

[1] Este trabajo ha sido patrocinado por la Fundación everis y la Universidad Politécnica de Madrid a través de la Cátedra de Mejora de Procesos de Software para España y América Latina.

y debe realizarse con la mayor calidad posible. El problema de la selección de proveedores de TI requiere conocimientos específicos de la organización y de la técnica para lograr la deseada calidad de TI, y puede ser una tarea desalentadora para los gerentes que no están seguros acerca de sus múltiples implicaciones.

En los últimos años, el outsourcing de TI ha cobrado una gran importancia en el mercado y continúa creciendo cada año. Outsourcing como concepto ganó la aceptación común en la década de los 80 y, hoy en día, todavía se utiliza para describir "una relación contractual con un proveedor de servicios especializados para el trabajo tradicionalmente hecho en casa" (Corbet & Associates, 2002).

Analistas de tendencias, como Morgan & Chambers (Lee, 2009) e IDC (Gellings, 2007), predicen cifras de crecimiento anual de aproximadamente el 10%. IDC estima que los clientes en EE.UU. aumentarán su gasto en outsourcing a partir de 2011 cuando la economía reviva y espera que el gasto de estos servicios llegue a un total estimado de $57 mil millones en 2014 (Dialani, 2010). Aarkstore (Aarkstore, 2010) prevé que el gasto anual en outsourcing superará los $ 700 mil millones a nivel mundial en el año 2013. Huai (Huai, 2007) demostró que el mercado de outsourcing de SI (sistemas de información) creció rápidamente durante los últimos 20 años con estimaciones de que los ingresos de outsourcing crecerán un 6,2%.

Aunque la popularidad del outsourcing de TI creció en las últimas dos décadas, muchos acuerdos de subcontratación no han continuado. Un estudio de (Dorothy & Dwayne, 2006) encontró que aproximadamente un tercio de las empresas estudiadas habían cancelado los contratos de outsourcing. También, de acuerdo con un estudio de (CMMI, 2010), del 20 al 25 por ciento de los proyectos de adquisición de TI de gran tamaño fracasan dentro de dos años y el 50 por ciento fracasan dentro de cinco años. Los factores que contribuyen al fracaso del proyecto son: la mala gestión, la mala definición de requisitos, la falta de evaluaciones comprensivas, la selección de proveedores y la contratación inadecuada, los insuficientes procedimientos tecnológicos de selección, y los cambios de requisitos no controlados. La mayoría de los fracasos de proyectos se podrían evitar si el adquirente aprendiera a: comprender los problemas de decisión, hacer un mejor análisis de decisiones y usar el buen juicio (CMMI, 2010). (Gill, Grewal, & Sareen, 2008) también indican que la mala decisión en la selección del proveedor de TI es una de las razones fundamentales que causa el alto fracaso del outsourcing de TI. Por lo tanto, el proceso científico de toma de decisión en la selección de proveedores de outsourcing de TI es muy importante para aumentar la tasa de éxito del outsourcing.

El trabajo presentado en este documento muestra los resultados de una revisión sistemática sobre la decisión en la selección de proveedores de TI en las áreas de tecnologías de la información y sistemas de información.

2. Revisión Sistemática sobre Análisis de Decisiones en la Selección de Proveedores de TI

Antes del siglo XX no había métodos para integrar los resultados de la investigación. En 1904, Pearson calculó la media de los resultados de la correlación entre la inoculación de la fiebre tifoidea y la mortalidad. Entonces la revisión sistemática

comenzó a formalizarse y a finales de los años 80, la revisión sistemática logra legitimidad como campo de investigación (Aydin & Bakker, 2008; Basili, Craze, Jino, Mendonca, & Shull, 2007).

La búsqueda de temas relacionados con la toma de decisión en la selección de proveedores de outsourcing de TI se basó en el trabajo de (Biolchini, Mian, Natali, & Travassos, 2005), que propone un protocolo de revisión sistemática sobre las directrices propuestas por (El-Emam et al., 2001), y en la forma de extraer información de documentos de ingeniería de software, desarrollado por (Basili, et al., 2007) y otras revisiones sistemáticas de índole similar desarrolladas por (Garcia, Piattini, & Pino, 2008), (Calvo Manzano, Cuevas, Gasca, & San Feliu, 2009) y (Calvo-Manzano, Cuevas, Grossi, Mejia, & San Feliu, 2009). La revisión sistemática desarrollada durante este trabajo siguió el protocolo establecido por los autores indicados anteriormente. En las siguientes secciones se presenta la revisión sistemática realizada sobre la toma de decisión en la selección de proveedores de TI.

2.1. Formulación de la pregunta

Esta sección tiene por objetivo definir la sintaxis de la pregunta de investigación (el contexto/problema en el que se aplica la revisión y la pregunta que el estudio debe responder) y su especificidad semántica (alcance de la pregunta) descrito por los elementos restantes de esta sección - intervención, control, efecto, medida de resultado, población y aplicación (Basili, et al., 2007). A continuación, se describe cada uno de ellos para la toma de decisión en la selección de proveedores de TI.

- Problema: el proceso de análisis de decisiones es muy importante en cualquier selección de proveedores de TI. El éxito del proyecto se relaciona con una buena decisión en la selección de proveedores. Por ello, se necesita desarrollar un modelo de toma de decisión en la selección de proveedores de TI.
- Pregunta: ¿cuáles son las iniciativas o propuestas desarrolladas para la toma de decisión en la selección de proveedores de TI?
- Intervención: el desarrollo actual de la toma de decisión en la selección de proveedores de TI.
- Control: no hay datos iniciales para esta revisión sistemática.
- Efecto: iniciativas y propuestas relacionadas con la toma de decisión en la selección de proveedores de TI.
- Medida de resultados: número de propuestas identificadas.
- Población: publicaciones relacionadas con la toma de decisión en la selección de proveedores de TI.
- Aplicación: cualquier empresa que tiene que tomar decisiones en la selección de proveedores de TI. Los investigadores que trabajan con la toma de decisión en la selección de proveedores de TI.

2.2. Selección de Fuentes

El objetivo de esta sección es seleccionar las fuentes donde se llevará a cabo la búsqueda de estudios primarios (Basili, et al., 2007). Para llevar a cabo la selección, se propone a abordar las siguientes cuestiones (secciones).

2.2.1. Definición de los Criterios de Selección de Fuentes y Lenguajes de Estudio

- Criterios de selección de fuentes: mecanismos de búsqueda con palabras clave y sitios sugeridos por expertos, artículos recomendados por otros expertos y artículos disponibles en los sitios Web.
- Lenguaje de estudio: inglés.

2.2.2. Identificación de Fuentes

- Método de búsqueda de fuentes: la identificación de las fuentes se ha basado en el criterio de expertos en nuestra área de investigación. Estas fuentes incluyen revistas como: *European Journal of Operational Research, Information and Software Technology, Software: Practice and Experience, Software Process: Improvement and Practice, IEEE Software, Software Technology and Engineering Practice,* talleres de investigación en computación e informes técnicos del Software Engineering Institute – SEI.
- Cadena de búsqueda: fueron extraídas palabras clave del conjunto de palabras definido en la "pregunta", combinando estas palabras clave con los operadores lógicos "AND" y "OR". Se han obtenido cuatro cadenas de búsqueda. Además, al analizar las referencias de los documentos iniciales encontrados, se han definido tres cadenas más de búsqueda. Estas cadenas de búsqueda se han adaptado para cada buscador de las fuentes. Las cadenas de búsqueda definidas han sido: 1) "decision analysis AND IT Acquisition; 2) "decision analysis" AND outsourcing; 3) "decision analysis AND resolution"; 4) "formal evaluation process" AND "information technology"; 5) decision AND (analysis OR resolution) AND "information technology"; 6) (supplier OR provider) AND selection AND outsourcing; 7) "decision analysis" AND (supplier OR provider) AND selection.
- Lista de fuentes: estas fuentes se han seleccionado teniendo en cuenta el método de búsqueda de fuentes definido. La lista de fuentes definidas ha sido: 1) Science@Direct; 2) Springer Link; 3) Computer Database – GALE; 4) ISI Web of Knowledge; 5) IEEE Computer Science Digital Library; 6) ACM Digital Library.

2.3. Selección de Estudios

En esta revisión sistemática se ha utilizado un procedimiento iterativo e incremental para la selección de estudios: a) iterativo, para agrupar todas las actividades que podrían repetirse durante el procedimiento, y b) incremental, ya que los estudios se abordan y se registran uno por uno, hasta la obtención de los resultados de la revisión sistemática (Garcia, et al., 2008). Se ha utilizado este procedimiento iterativo e incremental debido a su funcionalidad en otras revisiones sistemáticas. Esta sección describe el proceso y los criterios para la selección de los estudios y la evaluación.

2.3.1. Definición de los Criterios de Inclusión y Criterios de Exclusión de los Estudios, y Proceso de Selección

Este elemento define como se han seleccionado los estudios (Basili, et al., 2007).

- Definición de los criterios de inclusión (CI) y de exclusión (CE) de los estudios: se presentan los criterios por los que los estudios serán evaluados para decidir si deben ser seleccionados o no en el contexto de la revisión sistemática (Basili, et al., 2007). Los criterios fueron definidos por los autores de este artículo teniendo en cuenta la propuesta de (El-Emam, et al., 2001) (véase la Tabla 1).
- Definición de los tipos de estudios: se tendrán en cuenta todos los estudios relacionados con la selección de proveedores de TI. Sin embargo, el mayor interés se centrará en estudios que muestren resultados en la toma de decisión en la selección de proveedores de TI, es decir, que definan un modelo o forma de estructurar el proceso de toma de decisiones en la selección de proveedores de TI.
- Procedimiento para la selección de estudios: teniendo en cuenta los criterios de selección, el título ha sido el primer criterio principal, pero, en algunos casos, no ha proporcionado suficiente información, por lo tanto la lectura del resumen de cada uno de ellos ha sido necesaria y, en otros casos ha sido necesaria una revisión del texto completo.

Tabla 1. Criterios de inclusión (CI) y exclusión (CE) de estudios

#	Criterios de inclusión (CI) y exclusión (CE)
CI1	Incluir artículos cuyo título se relaciona con la toma de decisión en la selección de proveedores de TI
CI2	Incluir artículos que contienen palabras clave que coinciden con los definidos en la cadena de búsqueda
CI3	Incluir artículos cuyo resumen está relacionado con el tema tratado
CI4	Incluir artículos después de la lectura parcial o completa
CE1	Excluir todos los artículos duplicados (aquellos seleccionados en varios motores de búsqueda)

2.3.2. Ejecución de la Selección

El proceso de selección de los estudios primarios y los resultados de su evaluación son presentados en esta sección (Basili, et al., 2007).

- Evaluación de la calidad de los estudios: aplicando los CIs y CE definidos anteriormente, fue posible encontrar los estudios primarios (véase Tabla 2). Esta tabla contiene los resultados obtenidos con las cadenas de búsqueda en cada motor de búsqueda, la cantidad de trabajos encontrados, los que no se repitieron, y los estudios relevantes y primarios encontrados. Aplicando los criterios de inclusión CI1, CI2 y CI3, sólo 72 de 1727 artículos encontrados fueron considerados relevantes. Aplicando el criterio CE1 para la exclusión de los estudios duplicados, 62 estudios fueron obtenidos (columna No Repetido). A partir de éstos, y aplicando el criterio CI4, sólo 52 estudios fueron seleccionados como primarios.

Los estudios encontrados aquí son la base para la revisión sistemática y para el análisis que será presentado en las próximas secciones.

Tabla 2. Resultados derivados de cada fuente de búsqueda

Fuente	Encontrado	Relevante	No Repetido	Primario (%)
ACM Portal	656	16	14	9 (17,3)
IEEE Explore	29	4	2	2 (3,8)
Computer Database Gale	86	4	3	3 (5,8)
ISI Web of Knowledge	484	31	31	31 (59,6)
Science@Direct	245	11	8	3 (5,8)
Springer Link	227	6	4	4 (7,7)
TOTAL	1727	72	62	52 (100)

2.4. Extracción de la Información

Una vez que se han seleccionado los estudios primarios, se puede iniciar el proceso de extraer la información útil. Este proceso puede comenzar justo después de la definición de los criterios de inclusión y de exclusión de la información.

Tabla 3. Criterios de inclusión (CI inf) y de exclusión (CE inf) de la información

#	Criterios
CI1 inf	Recopilar información sobre la tendencia de la organización relacionada con la toma de decisión en la selección de proveedores de TI
CI2 inf	Recopilar información sobre el área de trabajo (TI, SI, etc.) de los estudios
CI3 inf	Recopilar información sobre los modelos aplicados por los estudios para la toma decisión en la selección de proveedores de TI
CI4 inf	Identificar metodologías, métodos y cualquier procedimiento propuestos en los estudios para la toma de decisión en la selección de proveedores de TI
CE1 inf	Excluir las informaciones que no están relacionadas con los criterios de inclusión definidos arriba

Para analizar la información de los estudios primarios, éstos fueron agrupados por carpetas relacionadas con la fuente de búsqueda donde fueron encontrados. Entonces, los estudios primarios de cada una de estas carpetas fueron copiados y pegados en una carpeta final para eliminar los repetidos. La información relevante de cada estudio primario fue destacada y luego extraída en un archivo separado, conteniendo el nombre del estudio primario y toda la información útil relacionada.

La ejecución de la extracción fue de la siguiente forma:

- Extracción de resultados objetivos: en busca de cualquier información relevante relacionada con la toma de decisión en la selección de proveedores de TI, los estudios seleccionados fueron leídos al completo. La información encontrada fue seleccionada y organizada en tablas que contienen: estudio consecutivo (número de documento secuencial), metodología de estudio (los comentarios de las principales ideas relacionadas con la metodología) y el

resultado del estudio (datos e información de las conclusiones presentadas en cada estudio).

- Extracción de resultados subjetivos: para completar la información encontrada en la extracción objetiva, se añadieron dos datos más a la tabla: los datos sobre los autores (nombre completo e información de contacto disponible en los estudios) y notas adicionales (un campo específico para almacenar información general relacionada con el tema tratado en el estudio).

Siguiendo este procedimiento, la información obtenida fue organizada siguiendo los campos: número consecutivo, título del artículo, revista/conferencia, autores, tópico, tipo de estudio, fecha, país, metas (analizar, a fin de, con respecto a, desde el punto de vista de), método (nombre, tipo, valores posibles, procedimiento de recolección de datos), implementación (compañía, categoría, tamaño), resultados del estudio y otros temas. Este formato fue adaptado teniendo en cuenta otras revisiones sistemáticas, como la de (Garcia, et al., 2008) y la propuesta de extraer información del trabajo de (Basili, et al., 2007).

3. Resumen de Resultados

A continuación, hay que analizar la información extraída de los estudios primarios.

3.1. Propuestas de los Estudios

La mayoría de los estudios analizados están relacionados con el área de TI o SI, que son exactamente los temas que esta revisión sistemática se propuso encontrar. Estos estudios mencionan algunas ideas generales relacionadas con el proceso de la toma de decisiones utilizando modelos de criterios múltiples. La mayoría de las propuestas (71%) están relacionadas con el diseño de un modelo específico para la toma de decisiones, mostrando cómo seleccionar la mejor alternativa según los criterios definidos. Algunas de las propuestas (23%) analizaron el proceso de toma de decisiones, un 4% se centró en el uso de las primitivas de decisión en el proceso de toma de decisiones y un estudio (el 2%) presentó una lista de artículos relacionados con el asunto de toma de decisiones.

3.2. Clasificación de los Estudios

Los estudios primarios fueron seleccionados siguiendo las actividades definidas en la revisión sistemática y pudieron ser clasificados en áreas de interés. La clasificación muestra 5 tópicos diferentes relacionados con: Tecnologías de la Información, Sistemas de Información, Salud, Logística e Ingeniería Civil.

El 59% de los estudios estaban relacionados con Tecnologías de la Información y el 27% con Sistemas de Información. Esta información muestra que la revisión sistemática logró uno de sus propósitos, ya que la mayoría de los estudios están relacionados con el área que estábamos buscando (Tecnologías de la Información y Sistemas de Información).

La mayoría de los estudios (37%) utilizaron una mezcla de modelos (dos o tres) en el proceso de toma de decisiones. El Analytic Hierarchy Process (AHP) (Saaty, 1980) se utilizó en un 19% de los estudios y PROMETHEE (Brans, Mareschall, & Vincke, 1986)

en un 10%. Los modelos utilizados para hacer un análisis del proceso de decisión contribuyeron con un 8% (Analytical Study) y un 12% (Survey Methodology). Fuzzy Analytical Network Process (Mikhailov & Singh, 2003) con (10%) y Decision Primitives con (4%) completan la clasificación de los estudios por modelo. Por lo tanto, es posible ver que la mayoría de los estudios utilizaron AHP o PROMETHEE como modelo de análisis de decisión, pero ninguno de ellos los utilizó juntos, tratando de captar las ventajas de cada modelo y evitar los inconvenientes.

4. Análisis de los Resultados

(Dickson, 1966) llevó a cabo una encuesta enviada por correo a cerca de 300 organizaciones comerciales, principalmente empresas de manufactura. A los gerentes de compras de estas empresas, se les pidió identificar los factores que eran importantes para la selección de proveedores. Sus hallazgos fueron divididos en dos categorías: las prácticas de selección de proveedores de las empresas y las prácticas de selección de proveedores de los individuos. Calidad, precio y entrega fueron identificados como los factores más críticos en el proceso de selección de proveedor. De acuerdo con sus resultados, el porcentaje de sistemas utilizando cada factor fue: Calidad (96,6%), Precio (93,9%), Entrega (93,9%), Servicio (81,8%), Capacidad Técnica (63,6%), Fuerza Financiera (51,5%), Localización Geográfica (42,4%), Reputación (42,4%), Acuerdos de Reciprocidad (15,1%) y Otros Factores (12,1%.)

Para estudiar las prácticas de selección de proveedores por los encargados de la toma de decisiones individuales, (Dickson, 1966) seleccionó cuatro escenarios únicos de compra que abarcan productos y situaciones muy diferentes, y una lista de 23 factores descritos en la literatura como importantes para el problema de la toma de decisión en la selección de proveedores. Cada entrevistado (gerente de compras) fue preguntado para considerar cada situación y evaluar la importancia de cada uno de los factores enumerados en una escala de 0 a 4 (4 representando importancia extrema y 0 indicando poca o ninguna importancia). Los factores definidos fueron (presentados en orden y con la media entre paréntesis):

- Importancia extrema: calidad (3,5).
- Importancia considerable: entrega (3,4), historia de desempeño (3,0), garantías y políticas de reclamaciones (2,9), instalaciones de producción y capacidad (2,8), precio (2,7), capacidad técnica (2,5), y posición financiera (2,5).
- Importancia media: cumplimiento de los procedimientos (2,5), sistema de comunicación (2,4), reputación y posición en la industria (2,4), deseo de negocios (2,3), administración y organización (2,2), control de operación (2,2), servicio de reparación (2,2), actitud (2,1), impresión (2,0), capacidad de embalaje (2,0), registro de relación laboral (2,0), localización geográfica (1,9), cantidad de trabajos anteriores (1,6) y asistencia a la formación (1,5).
- Pequeña importancia: acuerdos de reciprocidad (0,6)

Teniendo en cuenta la extensa naturaleza del estudio de (Dickson, 1966) y el enfoque estructurado adoptado, era apropiado extender sus resultados para incluir la investigación obtenida sobre el problema de toma de decisión en la selección de

proveedores publicada hoy en día, de 1992 a 2010, donde 52 artículos, como se mostró en la Tabla 2 de la sección 2.3.2, fueron revisados y analizados. Es importante tener en cuenta que esta revisión se basa totalmente en la literatura académica, mientras que el estudio de (Dickson, 1966) se basó en una encuesta a los agentes de compras.

Tabla 4. Distribución de los Criterios de Evaluación por Artículos

#	Pr	autor	Capacidad de Respuesta	Recursos Humanos	Precio	Soporte	Flexibilidad	Compatibilidad	Riesgo	Calidad	Tiempo	Entrega	Posición Financiera	Reputación	Experiencia	Participación de Mercado	Administración y Organización	Capacidad Técnica	Aparición TOTAL de los Criterios por Autor
1		Alssaoui, N.; Haouari, M.; Hassini, E.																	*
2		Akarte, M. M.; Rangaraj, N.; Ravi, B.; Surendra, N. V.																	*
3	7	Albani, A.; Muessigmann, N.			x					x	x								3
4	7	Albani, A.; Müssigmann, N.			x					x	x								3
5		Almeida, A. T.																	0
6	4	Araz, C.; Ozfirat, P. M.; Ozkarahan, I.		x			x	x		x		x	x						6
7		Aydin, M. N.; Bakker, M. E.																	0
8	6	Bacon, C. J.					x						x				x	x	4
9		Bansal, V.; Pandey, V.																	0
10	7	Batis, D.; Chamodrakas, I.; Martakos, D.			x					x		x							3
11	4	Betencourt, P. R. B.; Borenstein, D.		x	x	x	x		x			x							6
12		Boonstra, A.																	0
13		Burton, B. K.; Ross, S. C.; Tyran, C. K.																	0
14		Chaudhury, A.; Nam, K.; Rajagopalan, S.; Rao, H. R.																	0
15		Chen, L. Y.; Chen, Y. H.; Wang, T. C.																	0
16	3	Chen, L. Y.; Wang, T. C.			x		x			x		x	x	x	x	x		x	9
17	7	Cheng, C. H.; Kun-Cheng, H.; Wang, J. W.			x		x					x							3
18	7	Chuang, Y. T.; Liou, J. J. H.			x			x	x										3
19		Chun-ping, H.; Pei, Z.; Zhen-xiang, Z.																	0
20	4	Creese, R. C.; Golmohammadi, D.; Kolassa, J.; Valian, H.			x		x	x		x		x						x	6
21		Cunningham, S. W.; van der Lei, T. E.																	0
22		Dağdeviren, M.																	0
23		Daim, T.; Hanna, R.																	0
24	8	Davari, S.; Turksen, I. B.; Zarandi, M. H. F.			x							x							2
25		Diamond, R.																	0
26	4	Dibbern, J.; Goles, T.; Hirschheim, R.; Jayatilaka, B.	x		x			x		x		x						x	6
27		Dorothy, Leidner; Dwayne, Whitten																	0
28	1	Du, Y.; Li, C.; Sun, Y.		x	x		x	x	x	x		x	x	x	x	x	x		12
29	5	Dulmin, R.; Mininno, V.					x			x	x	x						x	5
30	8	Ferreira, A. M. D.; Laurindo, F. J. B.			x				x										2
31		Fu-ji, C.; Jian, Z.; Ping, C.																	0
32	6	Gill, S.; Grewal, C. S.; Sareen, K. K.;	x		x		x		x										4
33		Hashemi, M.; Rafiei, H.; Razmi, J.																	*
34	5	Huang, S. H.; Keskar, H.	x		x		x					x						x	5
35		Huang, W.; Li, S.; Murat, A.																	0
36	6	Hwang, H. S.								x	x	x				x			4
37	5	Isik, E.; Kara, S. S.; Onut, S.			x					x	x	x		x					5
38	2	Jharkharia, S.; Shankar, R.		x	x		x	x		x		x		x	x	x	x		10
39		Kabak, O.; Kahraman, C.; Kaya, I.; Orhan, E.																	0
40		Krishnan, H. A.; Park, D.																	0
41	3	Lee, A. H. I.			x	x	x		x	x		x	x	x	x				9
42	5	Lin, C. L.; Lin, Y. T.; Tzeng, G. H.; Yu, H. C.	x		x	x				x		x							5
43		Ngwenyama, O.; Osei-Bryson, K. M.																	0
44		Osmanagic-Bedenik, N.; Varga, M.																	0
45		Prompoon, N.; Vantakavikran, P.																	0
46		Ravindran, A. R.; Wadhwa, V.																	0
47	6	Saen, R. F.			x					x	x			x					4
48		Shakir, M.; Viehland, D.																	0
49	5	Shuqin, C.; Yong, C.			x	x	x			x								x	5
50		Sucky, E.																	0
51		Udo, G. G.																	0
52	4	Wang, J. J.; Yang, D. L.	x		x		x		x	x							x		6
		Prioridad TOTAL del Artículo (menor es mejor)	**24**	**11**		**17**		**22**	**33**		**36**		**17**	**20**	**9**	**12**	**13**	**32**	
		Aparición TOTAL de los Criterios en los Artículos	**5**	**4**	**21**	**4**	**14**	**6**	**7**	**17**	**6**	**16**	**5**	**6**	**4**	**4**	**4**	**7**	

* mencionan la lista de Dickson.

La Tabla 4 clasifica a todos los estudios y los criterios de evaluación encontrados en cada uno de ellos. La primera columna es solamente un número de identificación del estudio, y la segunda columna muestra la prioridad del estudio (Pr) (donde los artículos con más criterios definidos fueron considerados los más prioritarios). Artículos con mismo número de criterios definidos tienen la misma prioridad, por eso los números repetidos en la segunda columna de la Tabla4 (por ejemplo, artículos 16 y 41, con prioridad 2), y artículos sin criterios definidos no tienen ninguna prioridad, por eso algunos artículos no tienen un número asignado (por ejemplo, artículos 1 y 2). La tercera columna lista los autores de los artículos encontrados en la revisión sistemática

y las demás columnas contienen los nombres de los criterios definidos en este artículo, con excepción de la última columna (*aparición total de los criterios por autor*) que será explicada más adelante.

Una "x" en la Tabla 4, debajo de las columnas de los criterios, confirma el reconocimiento de un criterio en la investigación de selección de proveedores publicada en el estudio. Los nombres de los criterios se basaron en la lista de (Dickson, 1966) y en los nombres encontrados en los documentos analizados y, cuando dos criterios tienen una misma definición pero nombres diferentes, los autores de este trabajo decidieron poner el mejor nombre (según su criterio). Es importante mencionar que sólo fueron seleccionados los criterios que aparecieron en cuatro o más trabajos, es decir, otros criterios que fueron encontrados en el análisis de los documentos, pero que no aparecen en tantos estudios, no han sido incluidos.

En la Tabla 4 algunos cálculos y análisis fueron realizados para obtener la Tabla 5 con el ranking de prioridades de los criterios. El primer paso fue el cálculo de la *aparición total de los criterios en los artículos* (última fila de la Tabla 4) para definir el criterio con mayor prioridad, donde el criterio que apareció en más artículos fue el más prioritario (por ejemplo, el "precio" con 21). El problema aquí, es que a veces dos criterios tienen la misma *aparición total de los criterios en los artículos*, por ejemplo, la "capacidad técnica" y el "riesgo", con 7 como número total. Para solucionar esto, se han añadido dos valores más a la Tabla 4: la *aparición total de los criterios por autor* (última columna) y la *prioridad total del artículo (penúltima fila).*

La *aparición total de los criterios por autor* sólo suma el número de criterios de evaluación que un artículo definió y pone este valor en la última columna de la Tabla 4. Por ejemplo, el artículo número 3 de la Tabla 4 definió 3 criterios y este número fue colocado en la columna *aparición total de los criterios por autor*. La *prioridad total del artículo* suma el número de prioridad de los artículos (Pr, columna 2 de la Tabla 4).

Volviendo al ejemplo de los criterios de "capacidad técnica" y "riesgo", donde antes se indicó que tienen la misma *aparición total de criterios en los artículos*, y es imposible definir cuál tiene la mayor prioridad. En la Tabla 4, el criterio de "riesgo" apareció en los artículos 11, 18, 28, 30, 32, 41 y 52, con los respectivos números de prioridad 4, 7, 1, 8, 6, 3 y 4, y la "capacidad técnica" apareció en los artículos 8, 16, 20, 26, 29, 34 y 49, con los respectivos números de prioridad 6, 3, 4, 4, 5, 5 y 5. Por lo tanto, la prioridad total del artículo fue 33 para el criterio "riesgo" y 32 para el criterio "capacidad técnica". Como el número de prioridad menor es mejor, también la menor suma es mejor, por lo tanto una mayor prioridad fue dada al criterio de "capacidad técnica". Aquí es importante recordar que estos cálculos para la *prioridad total del artículo*, se utilizaron sólo para resolver el problema cuando dos criterios tienen una *aparición total de los criterios en los artículos* igual para la definición de las prioridades de los criterios de evaluación en la Tabla 5.

Por lo tanto, dos consideraciones deben ser hechas: 1) Los criterios precio, flexibilidad, calidad y entrega no tienen ningún cálculo de *prioridad total del artículo* debido a que no tienen el mismo número de *aparición total de los criterios en los artículos* con los demás criterios y 2) No es posible utilizar la *prioridad total del artículo* como primera forma de definir la prioridad de los criterios de evaluación.

Tabla 5. El ranking actual y el anterior de los criterios de evaluación

Ranking Actual	Ranking Anterior	Factor	Ranking Actual	Ranking Anterior	Factor
1	6	Precio	16	15	Soporte (servicio de reparación)
2	1	Calidad	No	3	Desempeño
3	2	Entrega	No	4	Garantías y Políticas de Reclamación
4	Nuevo	Flexibilidad	No	5	Instalaciones de Producción y Capacidad
5	7	Capacidad Técnica	No	9	Cumplimiento de los Procedimientos
6	Nuevo	Riesgo	No	10	Comunicación
7	11	Reputación y Posición en la Industria	No	12	Deseo de Negocios
8	Nuevo	Compatibilidad	No	14	Control de Operación
9	Nuevo	Tiempo	No	16	Actitud
10	8	Posición Financiera	No	17	Impresión
11	Nuevo	Capacidad de Respuesta	No	18	Capacidad de Embalaje
12	21	Experiencia (cantidad de negocios anteriores)	No	19	Registro de Relación Laboral
13	Nuevo	Recursos Humanos	No	20	Localización Geográfica
14	Nuevo	Participación de Mercado	No	22	Asistencia a la Formación
15	13	Administración y Organización	No	23	Acuerdos de Reciprocidad

La Tabla 5 proporciona la clasificación actual así como la anterior de los diferentes criterios de evaluación. La clasificación actual (columna “Ranking Actual”) se ha basado en los cálculos explicados anteriormente para definir el criterio con más prioridad. Primeramente, se ha utilizado la *aparición total de los criterios en los artículos*, como se ha explicado anteriormente, para definir el criterio con mayor prioridad. En los casos donde los criterios tenían una *aparición total de los criterios en los artículos* igual, se ha utilizado la *prioridad total del artículo* para el desempate. La columna "Ranking Anterior" se refiere a la clasificación que los criterios obtuvieron en el estudio de (Dickson, 1966). La Tabla 5 también incluye los nuevos criterios que no estaban presentes en la lista de 23 criterios.

5. Conclusiones

Esta revisión sistemática relacionada con la toma de decisión en la selección de proveedores de TI se ha llevado a cabo utilizando el protocolo propuesto por (Biolchini, et al., 2005). Este protocolo tiene cinco características principales relacionadas con el proceso de revisión sistemática, realizándose un análisis final estadístico de los estudios primarios para lograr resultados y también para comprobar la calidad de los trabajos seleccionados. La revisión sistemática lleva más tiempo y también más esfuerzo que la revisión de literatura normal, pero los resultados alcanzados son más consistentes, ya que se basan en un prototipo bien definido que detalla todos los métodos para la inclusión y exclusión de los estudios y de la información pertinente.

El gran porcentaje (86%) de los estudios analizados se relacionó con TI y SI, que son los principales temas que esta revisión sistemática se propuso analizar. En cuanto a los modelos utilizados y las propuestas definidas por los estudios, el modelo más utilizado fue el Analytic Hierarchy Process (AHP) con un 19%, seguido por el modelo PROMETHEE con un 10%. El 37% de los estudios utilizaron una mezcla de dos o tres modelos de criterios múltiples. Estos resultados indican una falta de estudios en la toma de decisión de selección de proveedores de TI, ya que sólo 52 estudios primarios fueron encontrados y no todos ellos están específicamente relacionados con la definición de un modelo de decisión en la selección de proveedores de TI. La mayoría de los de los estudios (71%), que están relacionados con el diseño de un modelo específico para la toma de decisiones, no desarrolló un modelo completo para ser utilizado en el área de TI, lo que es una motivación para el desarrollo de un trabajo futuro en relación con esta área.

Este estudio indica un cambio significativo de los diferentes criterios de evaluación y en su importancia relativa en comparación con la investigación realizada por (Dickson, 1966). Comenzando con el estudio de (Dickson, 1966) y utilizando los resultados de la revisión sistemática desarrollada por los autores de este artículo, se han definido 16 criterios para la selección de un proveedor de TI en base a la lista de 23 criterios definidos por (Dickson, 1966). Estos nuevos criterios de evaluación previamente definidos serán utilizados como parte de un modelo de decisión en la selección de proveedores de TI que será desarrollado por los autores de este artículo.

Referencias bibliográficas

Aarkstore. (2010). *The Black Book Of Outsourcing: State Of The Outsourcing Industry 2010*. Retrieved from http://www.articlesnatch.com/Article/The-Black-Book-Of-Outsourcing--State-Of-The-Outsourcing-Industry-2010/1763430.

Aydin, M. N., & Bakker, M. E. (2008). Analyzing IT Maintenance Outsourcing Decision From a Knowledge Management Perspective. *Information Systems Frontiers, 10*(3), 293-305.

Basili, V., Craze, D., Jino, M., Mendonca, M., & Shull, F. (2007). *Extracting Information From Experimental Software Engineering Papers*. Paper presented at the XXVI International Conference of the Chilean Society of Computer Science, 105-114.

Biolchini, J., Mian, P., Natali, A., & Travassos, G. (2005). *Systematic Review in Software Engineering*. Rio de Janeiro: COPPE/UFRJ.

Brans, J. P., Mareschall, B., & Vincke, P. H. (1986). How to Select and How to Rank Projects: The PROMETHEE Method. *European Journal of Operational Research, 14*, 228-238.

Calvo-Manzano, J., Cuevas, G., Grossi, L., Mejia, J., & San Feliu, T. (2009). Outsourcing Contracts: A Systematic Review. *International Book Series "INFORMATION SCIENCE & COMPUTING", 11*, 107-115.

Calvo Manzano, J. A., Cuevas, G., Gasca, G., & San Feliu, T. (2009). State of the art for risk management in software acquisition. *SIGSOFT Softw. Eng. Notes, 34*(4), 1-10.

CMMI, P. T. (2010). *CMMI for Acquisition, Version 1.3* (CMU/SEI-2010-TR-032): Software Engineering Institute.

Corbet, & Associates. (2002). Outsourcing's Next Wave.

Dialani, M. (2010). *Worldwide and U.S. Research and Development/Product Engineering Services 2010–2014 Forecast: Will the Economic Downturn and Changing Customer Needs Enable Increased Outsourcing of These Services?* Retrieved from http://www.idc.com/research/viewdocsynopsis.jsp?containerId=225954&sectionId=null&elementId=null&pageType=SYNOPSIS.

Dickson, W. G. (1966). An Analysis of Vendor Selection Systems and Decisions. *Journal of Purchasing, 2*, 5-20.

Dorothy, L., & Dwayne, W. (2006). Bringing IT Back: An Analysis of The Decision to Backsource or Switch Vendors. *Decision Sciences, 37*(4), 605-621.

El-Emam, K., Hoaglin, D. C., Jones, P. W., Kitchenham, B. A., Pfleeger, S. L., Pickard, L. M., et al. (2001). Preliminary Guidelines For Emprical Research in Software Enginerring. *National Research Council of Canada*.

Garcia, F., Piattini, M., & Pino, F. (2008). Software Process Improvement in Small and Medium Software Enterprises: A Systematic Review. *Software Quality Journal, 16*(2), 237-261.

Gellings, C. (2007). *Outsourcing Relationships: The Contract as IT Governance Tool*. Paper presented at the 40th Annual Hawaii International Conference on System Sciences, 236c-236c.

Gill, S., Grewal, C. S., & Sareen, K. K. (2008). A Multicriteria Logistics-Outsourcing Decision Making Using The Analytic Hierarchy Process. *Int. J. Services Technology and Management, 9*(1), 1-13.

Huai, J. (2007). *An Incentive Model of IS Outsourcing Contract*. Paper presented at the International Conference on Wireless Communications, Networking and Mobile Computing, 6588-6592.

Lee, A. H. I. (2009). A Fuzzy Supplier Selection Model With The Consideration of Benefits, Opportunities, Costs and Risks. *Expert Syst. Appl., 36*(2), 2879-2893.

Mikhailov, L., & Singh, M. G. (2003). Fuzzy analytic network process and its application to the development of decision support systems. *IEEE Transactions on Systems Man and Cybernetics Part C Applications and Reviews, 33*(1), 33-41.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: Mc Graw-Hill.

Critérios Editoriais

A RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) é um periódico científico, propriedade da AISTI (Associação Ibérica de Sistemas e Tecnologias de Informação), que foca a investigação e a aplicação prática inovadora no domínio dos sistemas e tecnologias de informação.

O Conselho Editorial da RISTI incentiva potenciais autores a submeterem artigos originais e inovadores para avaliação pelo Conselho Científico.

A submissão de artigos para publicação na RISTI deve realizar-se de acordo com as chamadas de artigos e as instruções e normas disponibilizadas no sítio Web da revista (http://www.aisti.eu).

Todos os artigos submetidos são avaliados por um conjunto de membros do Conselho Científico, não inferior a três elementos.

Em cada número da revista são publicados entre cinco a oito dos melhores artigos submetidos.

Criterios Editoriales

La RISTI (Revista Ibérica de Sistemas y Tecnologías de la Información) es un periódico científico, propiedad de la AISTI (Asociación Ibérica de Sistemas y Tecnologías de la Información), centrado en la investigación y en la aplicación práctica innovadora en el dominio de los sistemas y tecnologías de la información.

El Consejo Editorial de la RISTI incentiva autores potenciales a enviar sus artículos originales e innovadores para evaluación por el Consejo Científico.

Lo envío de artículos para publicación en la RISTI debe hacerse de conformidad con las llamadas de los artículos y las instrucciones y normas establecidas en el sitio Web de la revista (http://www.aisti.eu).

Todos los trabajos enviados son evaluados por un número de miembros del Consejo Científico de no menos de tres elementos.

En cada número de la revista se publican cinco a ocho de los mejores artículos enviados.

Chamada de Artigos

Encontra-se aberto até 20 de Abril de 2012 o período de envio de artigos para o nono número da RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação), o qual será publicado durante o próximo mês de Junho de 2012.

Este número é dedicado a eGovernment, eDemocracy e eParticipation e pretende integrar contribuições originais e relevantes nas diferentes dimensões e vertentes destas temáticas. Os tópicos recomendados incluem os listados abaixo. No entanto, também serão bem-vindos outros tópicos relacionados com estas temáticas mas aqui não incluídos:

- Arquiteturas de suporte e interoperabilidade;
- Gestão de identidade, privacidade e segurança;
- Serviços multicanal e serviços centrados no cidadão;
- Reutilização de informação, qualidade da informação e ontologias;
- Infraestruturas partilhadas e computação em nuvem;
- Compras públicas eletrónicas;
- Serviços móveis, sensíveis ao contexto e baseados na localização;
- Voto eletrónico e marketing político eletrónico;
- Comunidades virtuais, redes sociais e participação política *online*;
- Transparência e *accountability*;
- Modelos de maturidade e *benchmarking*.

Os artigos devem ser escritos em Português ou Espanhol. Para informações sobre dimensão, normas de formatação e processo de submissão, agradecemos a consulta do Portal da RISTI: http://www.aisti.eu

Llamada de Artículos

Se encuentra abierto hasta al día 20 de Abril de 2012 el período de envío de artículos para el noveno número de la RISTI (Revista Ibérica de Sistemas y Tecnologías de la Información), el cual será publicado durante el próximo mes de Junio de 2012.

Este número se dedica a las temáticas de eGovernment, eDemocracy y eParticipation. Pretende integrar contribuciones originales y relevantes en las diferentes dimensiones y aspectos de estos temas. Los asuntos recomendados incluyen los abajo listados, pero también serán bienvenidos otros asuntos relacionados con la temática y aquí no incluidos:

- Arquitecturas de apoyo y interoperabilidad;
- Gestión de la identidad, privacidad y seguridad;
- Servicios multicanal y servicios centrados en los ciudadanos;
- Reutilización de la información, calidad de la información y ontologías;
- Infraestructuras compartidas y computación en nube;
- Contratación pública electrónica;
- Servicios móviles, sensibles al contexto y basados en la localización;
- Voto electrónico y marketing político electrónico;
- Comunidades virtuales, redes sociales y participación política en línea;
- Transparencia y *accountability*;
- Modelos de madurez y *benchmarking*.

Los artículos deben ser escritos en portugués o español. Para obtener información sobre longitud, reglas de formato y proceso de envío, por favor consulte el Portal de la RISTI: http://www.aisti.eu

Os associados da AISTI recebem a RISTI gratuitamente, por correio postal. Torne-se associado da AISTI. Preencha o formulário abaixo e envie-o para o e-mail aisti@aisti.eu

Los asociados de la AISTI reciben la RISTI por correo, sin costo alguno. Hazte miembro de la AISTI. Rellena el siguiente formulario y remítelo al e-mail aisti@aisti.eu

Formulário de Associado / Formulario de Asociado

Nome/Nombre: _____

Instituição/Institución: _____

Departamento: _____

Morada/Dirección: _____

Código Postal: _____ Localidade/Localidad: _____

País: _____

Telefone/Teléfono: _____

E-mail: _____ Web: _____

Tipo de Associado e valor da anuidade:

☐ Individual - 25€

☐ Instituição de Ensino ou I&D/Institución de Educación o I&D - 250€

☐ Outro (Empresa, etc.) - 500€

NIF/CIF: _____

Data/Fecha: __/__/_____ Assinatura/Firma: _____

aisti
Associação Ibérica de Sistemas e Tecnologias de Informação

risti

Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Apoio

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR