

Una aproximación a la recomendación de artículos científicos según su grado de especificidad*

An approach to the recommendation of scientific articles according to their degree of specificity

Antonio Hernández, David Tomás, Borja Navarro

Universidad de Alicante

Carretera San Vicente del Raspeig s/n - 03690 Alicante (Spain)

antoniojhb@gmail.com, {dtomas, borja}@dlsi.ua.es

Resumen: En este artículo se presenta un método para recomendar artículos científicos teniendo en cuenta su grado de generalidad o especificidad. Este enfoque se basa en la idea de que personas menos expertas en un tema preferirán leer artículos más generales para introducirse en el mismo, mientras que personas más expertas preferirán artículos más específicos. Frente a otras técnicas de recomendación que se centran en el análisis de perfiles de usuario, nuestra propuesta se basa puramente en el análisis del contenido. Presentamos dos aproximaciones para recomendar artículos basados en el modelado de tópicos (*Topic Modelling*). El primero de ellos se basa en la divergencia de tópicos que se dan en los documentos, mientras que el segundo se basa en la similitud que se dan entre estos tópicos. Con ambas medidas se consiguió determinar lo general o específico de un artículo para su recomendación, superando en ambos casos a un sistema de recuperación de información tradicional.

Palabras clave: recuperación de información, modelado de tópicos, sistemas de recomendación

Abstract: This article presents a method for recommending scientific articles taking into consideration their degree of generality or specificity. This approach is based on the idea that less expert people in a specific topic prefer to read more general articles to be introduced into it, while people with more expertise prefer to read more specific articles. Compared to other recommendation techniques that focus on the analysis of user profiles, our proposal is purely based on content analysis. We present two methods for recommending articles, based on Topic Modelling. The first one is based on the divergence of topics given in the documents, while the second uses the similarities that exist between these topics. By using the proposed methods it was possible to determine the degree of specificity of an article, and the results obtained with them overcame those produced by an information retrieval traditional system.

Keywords: information retrieval, topic modelling, recommender systems

1 Introducción

En los últimos años, se ha producido un aumento exponencial de la información digital que se genera y distribuye a través del World Wide Web, produciendo un incremento en

la dificultad de encontrar información inmediatamente acorde a las necesidades de los usuarios. Ante la necesidad de escudriñar este maremágnum de información digitalizada, es que surgen los sistemas de recuperación de información (RI) (Baeza-Yates y Ribeiro-Neto, 1999).

Estos sistemas reciben una consulta por parte del usuario. Como resultado obtienen una lista de documentos relevantes y ordenados siguiendo criterios que intenta reflejar en qué medida se corresponden con dicha pe-

* Queremos agradecer a los revisores sus valiosas sugerencias y comentarios. Este trabajo ha sido parcialmente financiado por los siguientes proyectos: AT-TOS (TIN2012-38536-C03-03), LEGOLANG-UAGE (TIN2012-31224), FIRST (FP7-287607), DIIM2.0 (PROMETEOII/2014/001) y por el Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I+D+i (CAS12/00113).

tición. Para ponderar esta relevancia se han propuesto numerosas medidas basadas en el contenido de los documentos, como TF-IDF (Salton y Buckley, 1988) y (Salton, 1991), los modelos de lenguaje (Ponte y Croft, 1998) y Okapi BM25 (Robertson et al., 1994).

Si bien estos sistemas han demostrado su efectividad para encontrar documentos relevantes que permitan satisfacer las necesidades de información expresadas por los usuarios mediante un conjunto de términos claves. Hay numerosas situaciones en las que la necesidad de información de un usuario no puede ser expresada ni satisfecha por estas vías. Un ejemplo de este tipo de situaciones se produce cuando un usuario pretende iniciarse en un nuevo campo del conocimiento o profundizar en un campo ya conocido.

Las medidas tradicionales de RI no resultan válidas por sí solas para determinar cómo de bueno es un documento a la hora de iniciarse en un tema. Hay aspectos como la necesidad de conocimientos previos, la diversidad de temas tratados en el documento, o la complejidad del discurso que no son tenidos en cuenta.

En este artículo presentamos una aproximación a la recomendación de artículos científicos según su grado de generalidad o especificidad. Para esta aproximación nos basamos en la noción que los usuarios más noveles en un tema preferirán leer artículos más generales para introducirse en el mismo, mientras que otros con un nivel más experto preferirán artículos más específicos.

Nuestras aproximaciones analizan el contenido de los artículos para determinar el grado de especificidad. Partiendo de la salida proporcionada por un sistema tradicional de RI, nuestra propuesta realiza una reordenación para ofrecer los artículos de más general a más específico. Utilizamos el modelo *Latent Dirichlet Allocation* (LDA) propuesto por Blei, YNg, y Jordan (2003) para el modelado de tópicos y planteamos dos aproximaciones. La primera se basa en la divergencia de tópicos entre artículos, mientras que la segunda se basa en la similitud de los mismos.

El resto del artículo está organizado de la siguiente manera: en la Sección 2 se aborda el estado de la cuestión de los sistemas de recomendación; la Sección 3 describe la aproximación propuesta; en la Sección 4 se describen los experimentos realizados y sus resultados; en la Sección 5 se detallan las conclusiones y

las propuestas de trabajo futuro.

2 Estado de la cuestión

Como estado de la cuestión asumimos los estudios sobre la recomendación de artículos científicos que se están desarrollando. En este estudio nos centramos en las aproximaciones, técnicas de representación del contenido y el análisis de la especificidad o generalidad de los artículos.

En la recomendación de artículos científicos, McNee et al. (2002) después de analizar más de 170 algoritmos basados en Filtrado Colaborativo (CF) y Basados en el Contenido (CB), obtiene los mejores resultados con los de CF para determinar la importancia de un artículo con sus citas a partir de las citas web de los artículos de Research.net.

Ekstrand et al. (2010), siguiendo la idea anterior, aplica los algoritmos de CF y los híbridos para generar una lista automática de artículos, evaluando estos algoritmos sobre los artículos del *ACM Computing Surveys*, que dan una visión general sobre un tema, concluyendo también que los CF son mejores en este tipo de recomendación.

El perfil de usuario es una de las características que más se analizan para recomendar artículos. Se puede crear a través de la valoración que le proporciona el usuario a un documento, señalando su interés en el mismo (Giugni y León, 2011). Así como examinando los perfiles de usuarios existentes para detectar temas en común entre ellos (Wang y Blei, 2011).

Para determinar el carácter general o específico de los artículos, hasta donde nosotros sabemos, Candan et al. (2009) aplica la teoría de conjunto para detectar cuando los temas de los foros de discusión son nuevos, generales o específicos. Siguiendo esta misma línea, Wang et al. (2010) aplica también la teoría de conjuntos a perfiles de tópicos. Estos perfiles son creados extrayendo las palabras claves del título y las 10 primeras oraciones de las noticias originales, incluyendo también los comentarios de los lectores.

Las aproximaciones descritas previamente determinan lo general o específico de un artículo de noticias de los foros de discusión. Estas aproximaciones únicamente aplican un proceso de extracción de tópicos basado en TF-IDF.

A diferencia de los trabajos anteriores. En nuestra aproximación obtenemos un listado

de artículos mediante el uso de un sistema tradicional de RI. Al contenido de los artículos se le aplica el modelo LDA para extraer los tópicos. Estos tópicos son analizados mediante su similitud y divergencia. Y finalmente obtenemos un listado reordenado de más general a más específico.

En la siguiente sección explicaremos en detalle nuestra aproximación.

3 Descripción de la aproximación

3.1 Modelado de tópicos

El modelado de tópicos tiene como fin encontrar a través de algoritmos estadísticos, los principales temas de colecciones de documentos.

Los recientes avances en este campo van más allá de analizar las colecciones masivas de documentos al análisis de colecciones de *streaming* de vídeos, de imágenes, así como encontrar patrones en datos genéticos, imágenes, redes sociales (Blei, Carin, y Dunson, 2010).

Existen diferentes tipos de modelos de tópicos, entre los que podemos encontrar, *Explicit Semantic Analysis*, *Latent Semantic Analysis*, *Latent Dirichlet Allocation*, *Hierarchical Dirichlet Process*, *Non-negative Matrix Factorization*. Como ya se mencionó en la introducción nosotros usamos el *Latent Dirichlet Allocation* (LDA).

LDA es un modelo probabilístico generativo de tópicos. Los documentos de un corpus se representan como una combinación aleatoria sobre los tópicos latentes, donde cada tópico es caracterizado por una distribución de probabilidades sobre un vocabulario fijo de palabras. Para cada documento del corpus se generan palabras aleatoriamente siguiendo las siguientes etapas:

1. Seleccionar aleatoriamente una distribución de tópicos.
2. Para cada palabra del documento:
 - Seleccionar un tópico aleatoriamente sobre los tópicos generados en el paso 1.
 - Seleccionar una palabra aleatoriamente sobre su correspondiente vocabulario.

Al final de proceso obtendremos la probabilidad de pertenencia de cada palabra de cada documento a cada tópico (Blei, YNg,

y Jordan, 2003) y (Blei, Carin, y Dunson, 2010).

3.2 Descripción

Vamos a utilizar dos métodos para analizar el contenido de los artículos, uno basado en la divergencia de los tópicos de un artículo y otro basado en la similitud de tópicos entre artículos.

Para analizar la divergencia de los tópicos calculamos la desviación estándar entre ellos, siguiendo la intuición de que aquellos artículos cuya desviación sea menor tendrán una distribución de tópicos más equilibrado y por lo tanto tendrán un carácter más general, ya que no tendrán tópicos que destaquen marcadamente sobre el resto. Por otra parte, aquellos cuya desviación típica sea más elevada será porque existen tópicos más representativos que otros en el artículo, y que por lo tanto se trata de un artículo más específico.

En el caso de la similitud de tópicos entre artículos, emplearemos la similitud del coseno siguiendo la idea de que los artículos que presentan una similitud promedio más alta con otros artículos serán más generales, ya que tienen más tópicos comunes con otros artículos.

Los que tienen una similitud baja es porque tienen tópicos en común con sólo unos pocos y seguramente serán artículos más específicos.

Para determinar cuándo un artículo científico es general o específico. Seguimos los siguientes pasos que se detallan a continuación:

- Consulta: es el tema que se desea buscar en el corpus de artículos científicos.
- Indexación y recuperación de los artículos: usamos Lucene ¹ que es un motor de búsqueda de alto rendimiento escrito en Java y de código abierto. Entre sus características permite la indexación incremental de documentos, páginas web y contenidos de bases de datos, entre otras (McCandless, Hatcher, y Gospodnetic, 2010).

Los artículos recuperados se denotan por $P = (a_1, a_2, a_3, a_4, \dots, a_n)$ y están ordenados de mayor a menor relevancia. Esta relevancia viene dada por los parámetros por defecto con que se configura Lucene.

¹<http://lucene.apache.org/core/>

Por lo tanto, este sistema de RI nos sirve como un primer filtro para detectar dentro de un corpus los documentos más relevantes al área de estudio.

- **Análisis del contenido de los artículos:** para ello aplicamos la herramienta Mallet (McCallum, 2002). Esta herramienta está basada en el algoritmo LDA para el modelado de tópicos, permitiendo detectar y extraer tópicos del corpus de artículos. Los tópicos se almacenan en diferentes formatos y varios ficheros como resultado final. Nuestro análisis se centra en el fichero de pesos de los tópicos. Este fichero está conformado por una matriz donde los artículos están en la filas y los pesos de los tópicos de cada artículo en las columnas. En otras palabras, cada artículo estaría representado por los pesos de sus tópicos.
- **Resultados:** el listado de artículos recomendados se muestran según su generalidad (de más general a específico). También se puede configurar para que se muestre de lo más específico a general, es decir según su especificidad.

3.3 Aproximación #1. Divergencia

En esta aproximación analizamos como difieren entre sí los tópicos en un mismo artículo. Este análisis lo hacemos calculando la desviación estándar mediante la fórmula 1:

$$\sigma(P) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (a_i - \bar{a})^2} \quad (1)$$

Como resultado obtenemos un valor ponderado para cada artículo entre 0 y 1. Estos valores se interpretan de la siguiente manera:

- $\sigma(P) \approx 1$, se trata de un artículo específico ya que la importancia de los distintos tópicos en el documento varía mucho de unos a otros.
- $\sigma(P) \approx 0$, se trata de uno general ya que todos los tópicos tienen aproximadamente la misma relevancia en el artículo.

3.4 Aproximación #2. Similitud

En esta segunda aproximación, analizamos la similitud entre artículos. La similitud se calcula determinado el ángulo del coseno que forman entre sí los vectores de tópicos de los artículos a través de la fórmula 2:

$$Sim(a_i, a_{i+1}) = \frac{\sum_k^n a_{ik} \cdot a_{i+1k}}{\sqrt{\sum_k^n a_{ik}^2} \cdot \sqrt{\sum_k^n a_{i+1k}^2}} \quad (2)$$

Como resultado obtenemos un valor ponderado para cada artículo entre 0 y 1. Estos valores se interpretan de la siguiente manera:

- $\sigma(P) \approx 1$, se trata de un artículo general ya que los tópicos del artículo aparecen en muchos otros artículos.
- $\sigma(P) \approx 0$, se trata de uno específico ya que todos los tópicos del artículo aparecen en pocos.

En la siguiente sección se explican mediante experimentos los resultados de ambas aproximaciones.

4 Experimentos y resultados

4.1 Selección del corpus de evaluación

Para la comprobación de nuestra aproximación a través de los experimentos se decidió usar el repositorio de artículos científicos de la ACL Anthology². Ofrece de manera abierta más de 20.000 artículos científicos pertenecientes a los diferentes congresos organizados por la *Association for Computational Linguistics* (ACL) durante los últimos 40 años.

A través del proyecto ACL Anthology Network (Radev et al., 2013), se ofrece una versión en texto plano de todos los artículos de la ACL Anthology desde sus inicios hasta 2012. Este proyecto ofrece además información adicional como son la redes de citas del artículo y la de citas recibidas por el autor, etc.

4.2 Creación del corpus de referencia (*gold standard*)

Para poder evaluar el nivel de generalidad o especificidad de un artículo hemos creado un corpus de referencia (*gold standard*). Este corpus fue creado a partir de los 100 primeros artículos devueltos por Lucene para la consulta *question answering*. Estos 100 artículos fueron anotados manualmente por un experto en el tema. La anotación se realizó en una escala de 0 a 5 para determinar su grado de generalidad partiendo del siguiente criterio:

²<http://aclweb.org/anthology>

- 5: El artículo aporta una visión general sobre el área (survey), un artículo que se centra en mostrar el estado de la cuestión para *question answering*.
- 4: El artículo es una visión general sobre un subtema dentro del tema principal. Por ejemplo, un artículo que muestra todo el estado de la cuestión sobre la clasificación de preguntas aplicado a sistemas de *question answering*.
- 3: El artículo presenta una aproximación general al área. Por ejemplo, un artículo que describe una aproximación general a resolver el problema de *question answering*.
- 2: El artículo presenta una aproximación en un dominio o área concreta. Por ejemplo, un artículo que describe una aproximación al problema de la clasificación de preguntas en sistemas de *question answering*, o una aproximación general a *question answering* en un idioma concreto.
- 1: El artículo presenta un proyecto o recurso concreto. Por ejemplo, un artículo que presenta un corpus de preguntas para entrenar sistemas de *question answering*.
- 0: El artículo no tiene que ver con el área, es decir, Lucene nos ha devuelto algo que no tiene que ver con *question answering*.

Con esto se generó el orden de lectura de los artículos devueltos por Lucene según un experto.

4.3 Experimentos

Se realizaron dos experimentos con el fin de comprobar las ideas expuestas. Estos experimentos tuvieron como objetivos determinar la cantidad óptima de tópicos y evaluar las aproximaciones descritas anteriormente.

Para la evaluación usamos *Normalized Discounted Cumulative Gain* (nDCG) (ver fórmula 3). Esta medida combina la puntuación del documento (entre 0 y 5) con la posición en la que ha sido devuelto dentro de la lista de artículos recomendados, dando como única medida la ganancia acumulada sin importar el tamaño de la lista de documentos recuperados (Järvelin y Kekäläinen, 2002).

Como *baseline* para comparar el rendimiento de nuestra aproximación utilizamos la

salida original proporcionada por el motor de recuperación de información Apache Lucene (McCandless, Hatcher, y Gospodnetic, 2010).

En los experimentos propuestos se ha optado por mostrar en primer lugar aquellos de carácter más general.

A continuación se describen en detalles estos experimentos y sus resultados.

4.3.1 Experimento 1

El objetivo de este experimento fue determinar la cantidad óptima k de tópicos que se deben extraer del corpus de artículos para realizar su respectivo análisis de divergencia (σ) y similitud (Sim).

Para cumplir con este objetivo la secuencia de pasos que seguimos fue:

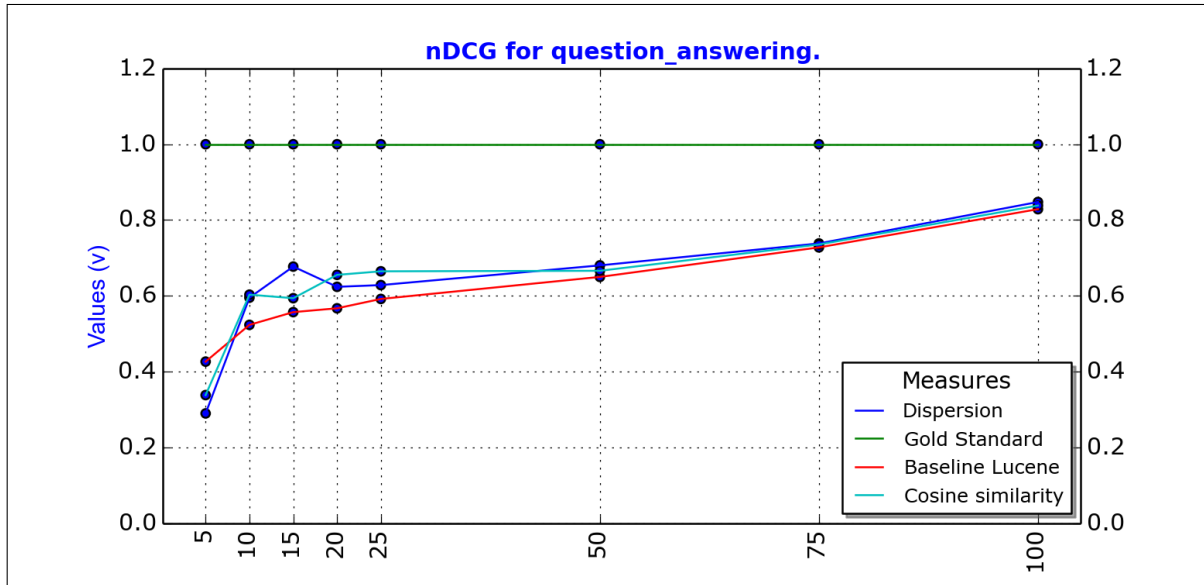
- Consulta, se realizó la consulta *question answering* sobre Lucene en el corpus de la ACL y se recuperaron 100 artículos, denotado por $P = (a_1, a_2, a_3, \dots, a_{100})$.
- Ordenación del corpus de referencia (CR), se ordenaron los artículos de mayor a menor puntuación, es decir, 5, 5, 5, 4, 4, 4, ..., 0, 0, 0 con el fin de obtener el mejor valor posible del sistema.
- Creación del corpus, se creó un corpus a partir de P (C_P).
- Extracción de tópicos, sobre el corpus C_P se experimentó con la obtención de distinto número de tópicos (5, 10, 15, 25, 50, 100, 150, 200, 250, y 300) para determinar la configuración óptima para nuestro sistema.
- Cálculo de medidas, para cada grupo de artículos creados por sus tópicos se aplicaron las dos aproximaciones definidas (desviación estándar de los tópicos de un documento y la similitud del coseno entre los tópicos de distintos documentos).
- Análisis de tópicos: se calculó $nDCG$ para cada uno de los conjuntos de tópicos propuestos (entre 5 y 300), a través de la fórmula 3:

$$nDCG_P = CR_{P_{[1]}} + \sum_{i=2}^{100} \frac{CR_{P_{[i]}}}{\log_2(i+1)} \quad (3)$$

Los valores de $CR_{P_{[1]}}$ se obtienen viendo la posición que tiene a_i en el corpus de referencia.

La tabla 1 muestra los resultados de $nDCG$ obtenidos para 5, 10, 15, 25, 50, 100,

nDCG	Número de tópicos									
	5	10	15	25	50	100	150	200	250	300
Sim	0,8328	0,8399	0,7995	0,7956	0,7970	0,8000	0,8056	0,7754	0,8343	0,8382
σ	0,7537	0,7654	0,8042	0,8384	0,8320	0,8207	0,8274	0,8333	0,8452	0,8476
M	0,7933	0,8027	0,8019	0,8170	0,8145	0,8104	0,8165	0,8044	0,8398	0,8424

Tabla 1: Resultados de $nDCG$ para cada grupo de tópicosFigura 1: Gráfico de $nDCG$ por cantidad de artículos.

150, 200, 250 y 300 tópicos estos valores fueron normalizados al dividir el resultado de DCG por el valor máximo posible obtenido al ordenar CR . Podemos observar que la mejor similitud entre artículos se logra con 10 tópicos y la divergencia entre tópicos en un artículo con 300.

En el caso de la similitud existe una diferencia de 0,0017 entre 10 y 300 tópicos. Al ser esta diferencia muy pequeña y el mayor valor de la media ($M=0,8429$) en los 300 tópicos. Se decidió fijar a 300 tópicos el valor óptimo para el segundo experimento.

Con este experimento concluimos que con el análisis de 300 tópicos por artículos tenemos en las primeras posiciones los artículos mejores anotados con respecto a CR_P .

4.3.2 Experimento 2

Este experimento tuvo como objetivo lograr una lista de artículos donde los primeros que se recuperen sean los más generales que aborden sobre el tema realizado en la consulta.

Partimos realizando una consulta con Lucene sobre *question answering*, recuperando los 100 artículos más relevantes y aplican-

do Mallet para extraer 300 tópicos para cada artículo (la mejor configuración según el experimento realizado en el apartado anterior).

Con estos 300 tópicos procedimos a realizar dos aproximaciones: la primera basada en el cálculo de la desviación estándar para analizar la divergencia de los tópicos en cada artículo, y la segunda en la similitud del coseno para ver la co-ocurrencia de tópicos entre artículos.

En este caso utilizamos la medida $nDCG@k$, que es la versión *cut-off* de $nDCG$ que presta más atención a los primeros resultados de la lista de artículos recuperados (Wang et al., 2013). El propósito de usar esta medida es valorar de manera más positiva aquellos sistemas que devuelven mejores artículos en las primeras posiciones de la lista. Pensando en la usabilidad de este tipo de sistemas, la satisfacción del usuario pasa por devolver los artículos que le resulten relevantes en las primeras posiciones.

Como se puede observar en la figura 1 a partir de 25 artículos se produce un aumento del $nDCG$, pero este aumento viene dado por

nDGC@k	Lucene	<i>Sim</i>	σ
5	0,4263	0,3380 (-20,7 %)	0,2896 (32,1 %)
10	0,5233	0,6030 (15,2 %)	0,5950 (13,7 %)
15	0,5569	0,5937 (6,6 %)	0,6771 (21,6 %)
20	0,5670	0,6552 (15,6 %)	0,6237 (10,0 %)
25	0,5917	0,6646 (12,3 %)	0,6283 (6,1 %)

Tabla 2: Resultados de nDGC@k

el propio funcionamiento de la medida: el corpus de referencia tiene valores más bajos en las últimas posiciones, incluso ceros, mientras que nuestra aproximación y el experimento de referencia *baseline* contienen artículos con una valoración más alta en esas posiciones.

En cualquier caso, como se comentó más arriba, lo que nos interesa es mirar los valores que se obtienen en las primeras posiciones, ya que a la hora de recomendar artículos a un usuario no es conveniente proporcionar un número elevado de ellos, ya que el usuario lo que espera son resultados válidos en las primeras posiciones.

Por esta razón, vamos a centrar el estudio en los 25 primeros artículos, aplicando $nDCG@k$, donde $k = \{5, 10, 15, 20 \text{ y } 25\}$.

En la tabla 2 se reflejan los resultados de aplicar $nDCG@k$ al *baseline* (Lucene), notando que nuestras aproximaciones en 10, 15, 20 y 25 artículos respectivamente supera siempre al *baseline*.

5 Conclusiones y trabajo futuro

En este artículo se han presentado dos aproximaciones basadas en modelado de tópicos para recomendar artículos científicos según su grado de especificidad. Primero analizamos la divergencia de los temas que se abordan en cada artículo, y seguidamente la similitud de los temas entre artículos.

Para poder evaluar nuestras aproximaciones tuvimos que crear nuestro propio corpus (100 artículos) partiendo de los más de 20.000 artículos de investigación que contiene la ACL. Nuestro corpus consistió en recuperar los primeros 100 artículos en el área de la “búsqueda de respuestas“ usando el motor de recuperación de información Apache Lucene.

Este listado de artículos en el orden que los recuperó Lucene fue nuestro *baseline* y para evaluar el nivel de generalidad o especificidad se etiquetaron de más general (5) a más específico (1), siendo nuestro corpus de referencia (gold standard).

Aplicando el modelado de tópicos hemos sido capaces de reordenar los documentos devueltos por un sistema de RI para que se muestren de manera más general a más específica.

El primer experimento consistió en determinar la mejor configuración posible en cuanto al número de tópicos a usar en el algoritmo LDA, obteniendo con 300 tópicos el resultado más óptimo.

El segundo experimento consistió en emplear nuestras dos aproximaciones (basada en la divergencia de tópicos en un documento y basada en la similitud del coseno entre tópicos de distintos documentos) para determinar la mejor reordenación posible de artículos, de más general a más específico.

Nuestras aproximaciones superaron al *baseline* de manera clara utilizando la medida $nDGC$, obteniendo una mejora del 21,6 % para $nDGC@15$ utilizando la medida de la divergencia de tópicos en un documento, y de un 15,6 % para $nDGC@20$ usando la similitud del coseno entre tópicos de distintos documentos.

Como trabajo futuro se plantea la forma de combinar la divergencia y similitud de los tópicos de los artículos para obtener mejores resultados a la hora de detectar la generalidad o especificidad de los artículos.

Bibliografía

- Baeza-Yates, R. y B. Ribeiro-Neto. 1999. *Modern information retrieval*, volumen 463. ACM press New York.
- Blei, D., L. Carin, y D. Dunson. 2010. Probabilistic topic models. *IEEE Signal Processing Magazine*, 27(6):55–65.
- Blei, D., A. YNg, y M. Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.
- Candan, K., E. Mehmet, T. Hedgpeth, J. Wook, Q. Li, y M. Sapino. 2009. Sea: Segment-enrich-annotate paradigm for adapting dialog-based content for improved accessibility. *ACM Transactions on Information Systems*, 27(3):1–45.

- Ekstrand, M., P. Kannan, J. Stemper, J. Butler, J. Konstan, y J. Riedl. 2010. Automatically building research reading lists. En *Proceedings of the fourth ACM conference on Recommender systems*, páginas 159–166. ACM.
- Giugni, M. y L. León. 2011. Clusterdoc un sistema de recuperación y recomendación de documentos basado en algoritmos de agrupamiento. *Telematique*, 9(2):13–28.
- Järvelin, K. y J. Kekäläinen. 2002. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446.
- McCallum, A. 2002. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu> (Consultado: 02 12 2014).
- McCandless, M., E. Hatcher, y O. Gospodnetic. 2010. *Lucene in Action: Covers Apache Lucene 3.0*. Manning Publications Co.
- McNee, S., I. Albert, D. Cosley, P. Gopalkrishnan, S. Lam, A. Rashid, J. Konstan, y J. Riedl. 2002. On the recommending of citations for research papers. En *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, páginas 116–125. ACM.
- Ponte, J. y B. Croft. 1998. A language modeling approach to information retrieval. En *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, páginas 275–281. ACM.
- Radev, D., P. Muthukrishnan, V. Qazvinian, y A. Abu-Jbara. 2013. The acl anthology network corpus. *Language Resources and Evaluation*, 47(4):919–944.
- Robertson, S., S. Walker, S. Jones, M. Hancock-Beaulieu, y M. Gatford. 1994. Okapi at trec-3. En *Proceedings of TREC*, volumen 3, páginas 109–126.
- Salton, G. 1991. Developments in automatic text retrieval. *Science*, 253(5023):974–980.
- Salton, G. y C. Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523.
- Wang, C. y D. Blei. 2011. Collaborative topic modeling for recommending scientific articles. En *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, páginas 448–456. ACM.
- Wang, J., Q. Li, Y. Chen, J. Liu, C. Zhang, y Z. Lin. 2010. News recommendation in forum-based social media. *The Journal of Information Science*, 180(24):4929–4939.
- Wang, Y., L. Wang, Y. Li, D. He, T. Liu, y W. Chen. 2013. A theoretical analysis of ndcg type ranking measures. *CoRR*, abs/1304.6480. <http://dblp.uni-trier.de/db/journals/corr/corr1304.html> (Consultado: 15 02 2015).