Procesamiento del Lenguaje Natural, nº39 (2007), pp. 55-62 recibido 18-05-2007; aceptado 22-06-2007

Medidas internas y externas en el agrupamiento de resúmenes científicos de dominios reducidos *

Diego A. Ingaramo, Marcelo L. Errecalde
LIDIC, UNSL, Argentina
Avda Ejército de los Andes 950
San Luis (5700)
{daingara,merreca}@unsl.edu.ar

Paolo Rosso
DSIC, UPV, España
Camino de Vera s/n 46022
prosso@dsic.upv.es

Resumen: Los algoritmos de agrupamiento suelen evaluarse o utilizan en su funcionamiento distintas medidas *internas* (u *objetivas*) como el *índice de Davies-Boulding* o el *índice de Dunn*, que intentan reflejar propiedades *estructurales* del resultado del agrupamiento. Sin embargo, la presencia de estas propiedades estructurales no garantiza la *usabilidad* de los resultados para el usuario, una propiedad *subjetiva* reflejada por medidas *externas* como la *medida F* y que determinan hasta que punto los grupos obtenidos se asemejan a los que se hubieran logrado con una categorización manual real. En trabajos previos, se ha observado una correlación interesante entre la medida de *densidad esperada* (interna) y la tradicional *medida F* (externa) en tareas de agrupamiento con documentos del corpus standard RCV1. En este trabajo, analizamos si esta relación también se verifica en tareas de *agrupamiento de resúmenes en dominios muy restringidos*. Este tipo de tarea ha demostrado tener un alto grado de complejidad y por ello, un análisis de este estilo, puede ser útil para determinar cuales son las propiedades estructurales fundamentales a tener en cuenta a la hora de diseñar algoritmos de agrupamiento para este tipo de dominios.
Palabras clave: agrupamiento de resúmenes, dominios muy restringidos, medidas de evaluación

Abstract: Clustering algorithms are usually based (and evaluated) taking into account *internal* (or *objective*) measures such as the *Davies-Boulding index* or the *Dunn index* which attempt to evaluate particular structural properties of the clustering result. However, the presence of such structural properties does not guarantee the *interestingness* or *usability* of the results for the user, a subjective property usually captured by *external* measures like the *F-measure* that determine up to what extent the resulting groups resemble a real human classification. In previous works, an interesting correspondence have been observed between the (internal) *expected density* measure and the (external) *F*-measure in clustering tasks with documents from the standard corpus RCV1. In this work, we investigate if that correspondence also is verified in clustering on *narrow-domain abstracts* tasks. This is a challenging problem and we think that this kind of study can be useful for detecting which are the most relevant structural properties which should be considered when designing clustering algorithms for these domains.
Keywords: clustering of abstracts, narrow domains, evaluation measures

1. Introducción

El agrupamiento de textos consiste en la asignación no supervisada de documentos en distintas categorías. Si bien es común que este tipo de tareas se estudie utilizando colecciones de documentos standards, en muchos casos sólo están disponibles los *resúmenes* descriptivos (*abstracts*), como ocurre con muchas publicaciones científicas. La tarea de *agrupamiento de resúmenes*, presenta un desafío considerable debido a la baja frecuencia de ocurrencia de los términos en los documentos. Esta tarea se dificulta aún más, cuando los resúmenes abordan una temática similar, debido a que existe una intersección significativa en el vocabulario de los documentos. Esta tarea, conocida como *agrupamiento de resúmenes en dominios muy restringidos* (en inglés *clustering abstracts on narrow domains*) ha comenzado a ser abordada en distintos trabajos recientes que presentan distin-

* El trabajo fue financiado parcialmente por los proyectos de investigación TIN2006-15265-C06-04 y ANPCyT-PICT-2005-34015.

tas propuestas para enfrentar las complejidades propias de este tipo de dominio (Makagonov, Alexandrov, y Gelbukh, 2004), (Alexandrov, Gelbukh, y Rosso, 2005), (Pinto, Jimenez, y Rosso, 2006).

Por otra parte, Stein (Stein, Meyer, y Wißbrock, 2003) destaca que las métricas tradicionales de validez de un agrupamiento (*índice de Davies-Boulding*, *índice de Dunn*, *densidad esperada* y otras), son medidas *internas* (u objetivas) que toman en cuenta distintas propiedades *estructurales* de los grupos obtenidos. Sin embargo, estas medidas no garantizan la calidad del agrupamiento de acuerdo a la clasificación que hubiera realizado un usuario ante la misma tarea. Este tipo de información suele estar expresada en medidas *externas* (o *subjetivas*) como la *precisión* o la *medida F*, y requieren para su cálculo de información sobre la clasificación real realizada por un humano. Un algoritmo de agrupamiento no tiene en general acceso a este tipo de información. Por ello, se suele tomar como referencia a las medidas internas, y confiar en que permitan predecir adecuadamente las medidas externas. Este es el caso de métodos como *MajorClust* (Stein y Niggemann, 1999), que aproxima la función de conectividad parcial o el algoritmo *AAT* (*Adaptive AntTree*) (Ingaramo, Leguizamón, y Errecalde, 2005b), (Ingaramo, Leguizamón, y Errecalde, 2005a) que utiliza el índice de Davies-Boulding en una etapa del algoritmo.

Respecto a las observaciones de Stein, éste analiza en que medida distintas medidas internas de un agrupamiento sirven para predecir la *usabilidad* del mismo (medidas subjetivas) usando en su estudio distintas muestras de un corpus etiquetado standard (RCV1). En este caso, se reportan resultados interesantes respecto a la correlación entre la medida de densidad esperada (interna) y la medida F (externa).

El objetivo de nuestro trabajo es determinar si esta correspondencia también se verifica en un dominio más dificultoso como lo es el agrupamiento de resúmenes en dominios muy restringidos. Esta información podrá ser utilizada en algoritmos de agrupamiento que explícitamente recurren a medidas internas (Ingaramo, Leguizamón, y Errecalde, 2005b), (Ingaramo, Leguizamón, y Errecalde, 2005a) para adaptarlos a las características de este tipo de dominios. En el trabajo experimental se consideran 3 corpora de resúmenes científicos en dominios muy específicos y un subconjunto de un corpus tradicional. En todos los casos, se utilizan distintas codificaciones de los documentos y distintos porcentajes del vocabulario. Los métodos de agrupamiento utilizados son k-means, MajorClust y un algoritmo de clustering "artificial".

El artículo está organizado de la siguiente manera. En la Sección 2 se resumen brevemente las particularidades que surgen en la tarea de agrupamiento de resúmenes en dominios muy restringidos. En la Sección 3 se describen algunas de las consideraciones realizadas por Stein respecto a las medidas internas y externas del agrupamiento y se detallan las medidas que utilizaremos en este trabajo. En la Sección 4 se describe el trabajo experimental y los resultados obtenidos. Por último se presentan las conclusiones y posibles trabajos futuros.

2. Agrupamiento de resúmenes en dominios reducidos

La categorización de textos es el agrupamiento de documentos con temáticas similares, y es una componente clave en la organización, recuperación e inspección de grandes volúmenes de documentos accesibles actualmente en Internet, bibliotecas digitales, etc. Distintos trabajos de investigación han abordado el problema de la categorización automática de textos en situaciones donde se cuenta con un esquema de clasificación predefinido y existe una colección de documentos ya clasificados. En estos casos, las técnicas de aprendizaje automático han demostrado una gran eficacia a la hora de obtener clasificadores con muy buenos desempeños en diversas colecciones de documentos (Sebastiani, 2002), (Montejo y Ureña, 2006).

Esta tarea de agrupamiento es más compleja cuando el proceso de formación de categorías es no supervisado y no se dispone de una colección de documentos etiquetados como referencia. En estos casos se introducen dificultades adicionales al caso supervisado como, por ejemplo, la correcta determinación del número de clases o la forma de evaluar el resultado del proceso de agrupamiento.

Si bien las técnicas de agrupamiento han sido aplicadas en reiteradas oportunidades a documentos completos provenientes de colecciones de acceso público, el acceso a muchas publicaciones científicas queda en muchos ca-

sos restringido a sus resúmenes (o *abstracts*). En estos casos, las técnicas de agrupamiento tradicionales suelen arrojar resultados inestables e imprecisos debido a las bajas frecuencias de ocurrencias de las palabras presentes en el resúmen y a la ocurrencia de frases comunes completas que no realizan ningún aporte al significado del documento (ej. "In this paper we present..."). Aqui es importante diferenciar:

- Resúmenes concernientes a temáticas bien diferenciadas (deportes, política, economía, etc).
- Resúmenes concernientes a un dominio muy restringido (*narrow domain*) donde todos los resúmenes abordan una temática similar y la intersección de sus vocabularios es muy significativa.

La dificultad del agrupamiento en el último caso ya ha sido observada en distintos trabajos recientes (Alexandrov, Gelbukh, y Rosso, 2005), (Pinto, Jimenez, y Rosso, 2006) que proponen distintos enfoques para su abordaje. En (Makagonov, Alexandrov, y Gelbukh, 2004) por ejemplo, se utilizó una adecuada selección de las palabras claves y una mejor evaluación de la similitud entre documentos, experimentándose con dos colecciones de abstracts de las conferencias CICLing 2002 e IFCS 2000. En (Alexandrov, Gelbukh, y Rosso, 2005) se propone el uso del método MajorClust de Stein para el clustering de palabras claves y documentos, experimentándose con la misma colección CICLing mencionada previamente.

Recientemente, en (Jiménez, Pinto, y Rosso, 2005) un nuevo experimento con esta colección ha arrojado mejores resultados a partir del uso del método de punto de transición. Finalmente, en (Pinto, Jimenez, y Rosso, 2006), (Pinto et al., 2006) se muestra que esta técnica de selección de términos, puede producir un mejor desempeño que otras técnicas no supervisadas en colecciones de resúmenes.

Estos últimos trabajos comparten la conclusión de que puede haber una influencia significativa del tamaño del vocabulario en la medida F cuando se utiliza la técnica del punto de transición. Por este motivo, en este trabajo decidimos que el análisis de la relación de las medidas internas y externas tomaría en cuenta distintos porcentajes del tamaño de vocabulario, utilizando esta interesante técnica de selección de términos.

3. Medidas de evaluación de agrupamientos

El trabajo realizado por Stein en (Stein, Meyer, y Wißbrock, 2003) intentó determinar si las medidas de validez internas para un agrupamiento de textos se correspondian con los criterios utilizados por un usuario final, en relación a la misma tarea. Dentro de este marco se analizaron distintas medidas internas tradicionales como la familia de índices de Dunn y Davies-Bouldin y medidas basadas en densidad como la medida de conectividad parcial y la medida de densidad esperada. El análisis se realizó considerando que el criterio real del usuario estaba reflejado en la medida F (externa).

Para los experimentos se consideraron muestras de la colección Reuters Text Corpus Volume 1 (Rose, Stevenson, y Whitehead, 2002) y distintos algoritmos de agrupamiento como k-Means y MajorClust. Los resultados mostraron que las medidas internas tradicionales se comportan de manera consistente aunque los grupos encontrados no sean buenos en relación a la medida F. La medida de densidad esperada en cambio, tiene un mejor comportamiento que, de acuerdo a Stein, se debe a la independencia que tiene esta medida con respecto a la forma y a la distancia entre grupos y elementos de cada grupo. A continuación, se describen brevemente la medida de densidad esperada y la medida F analizadas en el trabajo de Stein.

3.1. Medida de densidad esperada

Se dice que un grafo ponderado $\langle V, E, w\rangle$ *no es denso* si $|E| = \mathcal{O}(|V|)$, y que es *denso* si $|E| = \mathcal{O}(|V|^2)$. De esta forma podemos calcular la densidad θ de un grafo mediante la ecuación $|E| = |V|^\theta$. Con $w(G) = |V| + \sum_{e\in E} w(e)$, la relación para grafos ponderados es:

$$w(G) = |V|^\theta \Leftrightarrow \theta = \frac{\ln(w(G))}{\ln(|V|)} \qquad (1)$$

θ puede usarse para comparar la densidad de cada subgrafo inducido $G' = \langle V', E', w'\rangle$ de G, y se dice que G' (no) es denso respecto a G si la relación $\frac{w(G')}{|V'|^\theta}$ es más chica (más grande) que 1.

Definición (Stein, Meyer, y Wißbrock, 2003): Sean $\mathcal{C} = \{C_1, .., C_k\}$ los grupos de un grafo ponderado $G = \langle V, E, w\rangle$ y sea $G_i = \langle V_i, E_i, w_i\rangle$ el subgrafo inducido de G respecto al cluster C_i. La *densidad esperada* $\overline{\rho}$ de un agrupamiento $\mathcal{C}$ es:

$$\overline{\rho}(\mathcal{C}) = \sum_{i=1}^{k} \frac{|V_i|}{|V|} \cdot \frac{w(G_i)}{|V_i|^{\theta}} \qquad (2)$$

Un mayor valor de $\overline{\rho}$ representa un mejor agrupamiento.

3.2. La medida F

La medida F combina las medidas de *precisión* y *recall*.

Definición: Sea D un conjunto de documentos, $\mathcal{C} = \{C_1, ..., C_k\}$ un agrupamiento de D y $\mathcal{C}^* = \{C_1^*, \ldots, C_l^*\}$ la clasificación real de los documentos en D. El *recall* de un grupo j en relación a la clase i, $rec(i, j)$ se define como $|C_j \cap C_i^*|/|C_i^*|$. La *precisión* de un grupo j respecto a la clase i, $prec(i, j)$ se define como $|C_j \cap C_i^*|/|C_j|$. La medida F combina ambas funciones de la siguiente manera:

$$F_{i,j} = \frac{2}{\frac{1}{prec(i,j)} + \frac{1}{rec(i,j)}} \qquad (3)$$

y la medida F global se define:

$$F = \sum_{i=1}^{l} \frac{|C_i^*|}{|D|} \cdot \max_{j=1,..,k}\{F_{i,j}\} \qquad (4)$$

En nuestro caso, es importante determinar si la correspondencia observada por Stein entre ambas medidas en la colección RCV1 se mantiene al agrupar resúmenes de dominios muy restringidos. Si ésto ocurre, se podrían adaptar para este tipo de dominios, algunos métodos de agrupamiento que explícitamente utilizan otras medidas internas. En caso contrario, se podría investigar si otras medidas internas se comportan mejor en estos casos.

4. Experimentos

4.1. Conjuntos de Datos

En los experimentos se utilizaron las 4 colecciones que se describen a continuación, que difieren fundamentalmente en la cantidad de documentos y el tipo de distribución entre los distintos grupos.

4.1.1. La colección *CICLing2002*

Este corpus se caracteriza por un reducido número de resúmenes (48) distribuidos manualmente y en forma balanceada en 4 grupos que corresponden a temáticas abordadas en la conferencia CICLing 2002. Es un corpus pequeño (23.971 bytes) con 3382 términos en total y un vocabulario de tamaño 953. La distribución de los resúmenes en los grupos se muestra en la Tabla 1.

Categoría	**Nro de resúmenes**
Lingüística	11
Ambigüedad	15
Léxico	11
Proc. de texto	11
TOTAL	**48**

Tabla 1: Distribución de *CICLing2002*

4.1.2. La colección *Hep-Ex*

Este corpus, basado en la colección de resúmenes de la Universidad de Jaén, España (Montejo, Ureña Lopez y Steinberg, 2005), está compuesto por 2922 resúmenes del área de física, originalmente guardados en los servidores del *Conseil Européen pour la Recherche Nucléaire* (CERN). Este corpus de 962.802 bytes de tamaño, con un total de 135.969 términos en total y un vocabulario de tamaño 6150, distribuye los 2922 resúmenes en 9 categorías de la manera que se muestra en la Tabla 2. Como se puede observar, tiene una mayor cantidad de grupos que en el caso de CICLing2002 y además es altamente desequilibrado, ya que uno de los grupos concentra casi el 90 % de los documentos.

Categoría	**Nro de resúmenes**
Resultados Experimentales	2623
Detectores y técnicas exp.	271
Aceleradores	18
Fenomenología	3
Astronomía	3
Transf. de Información	1
Sistemas No Lineales	1
Otros campos de la física	1
XX	1
TOTAL	**2922**

Tabla 2: Distribución de *Hep-Ex*

4.1.3. La colección *KnCr*

Esta colección es un subconjunto de la colección de textos científicos del área de me-

dicina de MEDLINE, restringida a aquellos resúmenes sobre temas vinculados al cáncer. Se compone de 900 resúmenes distribuidos en 16 categorías como se muestra en la Tabla 3. Este corpus tiene un tamaño de 834.212 bytes, con 113.822 términos en total y un vocabulario de tamaño 11.958. Estudios preliminares (Pinto y Rosso, 2006) demuestran la alta complejidad y el desafío que presenta esta colección.

Categoría	Nro de resúmenes
Sangre	64
Huesos	8
Cerebro	14
Pecho	119
Colon	51
Estudios Genéticos	66
Genitales	160
Pulmones	29
Hígado	99
linfoma	30
renal	6
piel	31
estómago	12
terapia	169
tiroide	20
otros	22
TOTAL	**900**

Tabla 3: Distribución de *KnCr*

4.1.4. La colección *5-MNG*

Las 3 colecciones previas corresponden a colecciones de resúmenes científicos en dominios muy específicos. Para poder comparar los resultados con una colección que no tuviera estas características, se generó un subconjunto de la colección de textos completos MiniNewsGroups [1], de manera tal que los grupos seleccionados correspondieran a temáticas bien diferenciadas. Esta colección, que denominamos *5-MNG*, está compuesta por 5 grupos de tamaño equilibrado de 100 documentos cada uno (ver Tabla 4).

4.2. Diseño Experimental

En el trabajo experimental se analizó si existe una correspondencia general entre la densidad esperada y la medida F evitando introducir distintos tipos de sesgos en factores como el tamaño del vocabulario utilizado o la codificación de los documentos. Por este motivo, se buscó obtener un muestreo representativo de resultados considerando distintos escenarios.

Para el caso de la codificación de los documentos, por ejemplo, se obtuvieron resultados considerando la mayoría de las 20 codificaciones SMART (Salton, 1971). Para la reducción del vocabulario, por su parte, los términos más relevantes fueron seleccionados mediante la técnica del *punto de transición*. Esta técnica ha demostrado tener un impacto significativo en la medida F en estudios recientes con este tipo de dominios (Pinto et al., 2006). Para cada uno de los corpus se consideraron los resultados obtenidos con los siguientes porcentajes de vocabulario: 2 %, 5 %, 10 %, 20 %, 40 %, 60 %, 80 % y hasta un 100 % (vocabulario total).

Como algoritmos de clustering se utilizaron los métodos k-means y MajorClust. En el primer caso se deben especificar el número de clusters requeridos y en el segundo caso no. También se implementó un algoritmo de *clustering artificial* del tipo del utilizado por Stein en sus experimentos. La idea en este caso es que, dado que se conoce la categorización de referencia $\mathcal{C}^*$, es posible generar agrupamientos artificiales $\mathcal{C}_1, \ldots, \mathcal{C}_n$ que difieren en el grado de ruido introducido en el agrupamiento. Este ruido es generado mediante el intercambio controlado de pares de subconjuntos de documentos entre los grupos que pueden variar desde un documento hasta el 50 % de los documentos de un grupo.

4.3. Resultados

En las Figuras 1, 2, 3 y 4 se muestran los resultados del agrupamiento artificial con las colecciones explicadas previamente. En todos los casos, los valores correspondientes al eje x representan las densidades esperadas $\overline{\rho}$ de los agrupamientos encontrados por este algoritmo, y los valores en el eje y son los valores de la medida F para cada agrupamiento. Debe-

Categoría	Nro de resúmenes
Gráficas	100
Motocicletas	100
Baseball	100
Space	100
Politica	100
TOTAL	**500**

Tabla 4: Distribución de *5-MNG*

[1] http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.html. 20 Newsgroups, the original data set. Ken Lang, 1993.

mos notar que además de los puntos correspondientes a los resultados del agrupamiento artificial, también se grafica una línea rotulada "*Curva ideal de la muestra*". Esta línea corresponde a la función lineal que pasa por los puntos $(\overline{p}_1, F_1)$ y $(\overline{p}_2, F_2)$ donde $\overline{p}_1$ y $\overline{p}_2$ son el mínimo y máximo valor de densidad esperada encontrado en los experimentos para este corpus y F_1 y F_2 son el mínimo y máximo valor de la medida F obtenidos para este corpus en nuestros experimentos. Esta función corresponde a un resultado idealizado donde la medida F se incrementaría linealmente de acuerdo al crecimiento de la densidad esperada. Dado que esta función sería un patrón deseable posible para la correlación entre ambas medidas, en todas las figuras subsiguientes, esta línea será tomada como referencia para comparar los resultados obtenidos con los distintos algoritmos de agrupamiento.

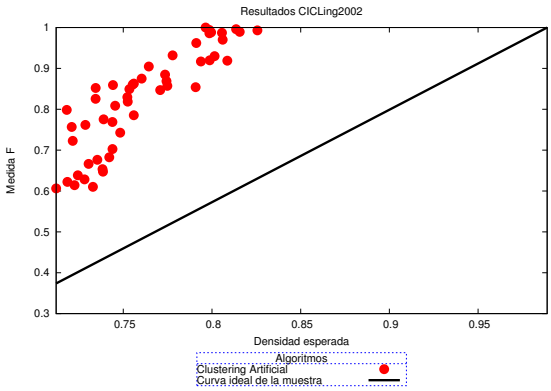

Figura 1: CICLing2002 (clustering artificial)

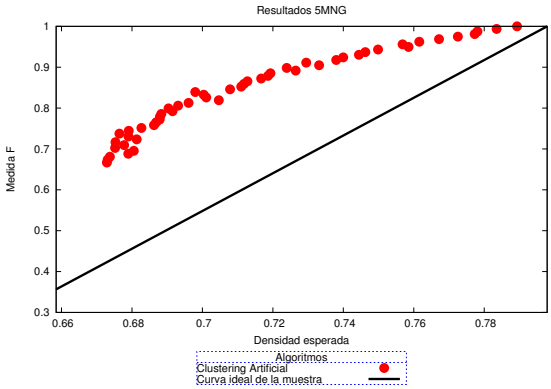

Figura 2: 5-MNG (clustering artificial)

En todas estas figuras se puede observar una buena correspondencia entre la medida de densidad esperada y la medida F cuando se introduce ruido gradualmente en el agrupamiento. Estos resultados se asemejan a los obtenidos por Stein con agrupamientos artificiales con RCV1. Sin embargo, en nuestro caso dos situaciones merecen atención. La primera es respecto a CICLing2002 (Figura 1) donde se observan variaciones significativas de F con pequeñas variaciones de la densidad. Esto parece indicar que cuando existen pocos grupos y pocos documentos por grupo la densidad esperada no provee una estimación muy estable de F. Esta inestabilidad no se observa en una colección con pocos grupos con textos completos como es el caso de 5-MNG (Figura 2) cuya curva tiene grandes similitudes con la curva ideal para este corpus. En el caso de Hep-ex (Figura 3) se observa que la medida F se mantiene casi inalterable respecto a las variaciones de la densidad esperada. Este comportamiento puede estar motivado por el hecho de que esta colección tiene un grupo que contiene el 90 % de los documentos y el clustering artificial parte del agrupamiento perfecto de los documentos. Es de esperar entonces, que si bien se incorpora paulatinamente ruido intercambiando documentos entre los grupos, el impacto que se

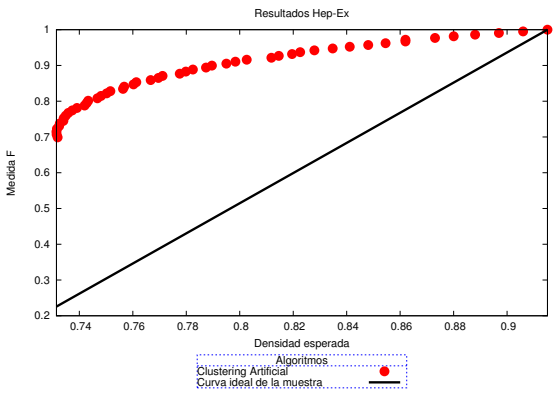

Figura 3: Hep-ex (clustering artificial)

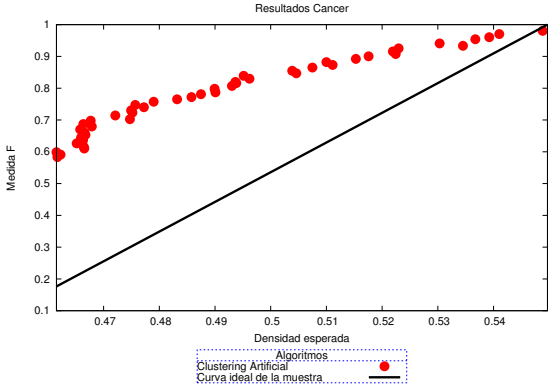

Figura 4: KnCr (clustering artificial)

tiene sobre la medida F no alcance a ser significativo. De esta forma, la medida F mantendrá alto sus valores independientemente de los valores de densidad esperada. La colección de resúmenes que muestra una mejor correspondencia entre la densidad esperada y la medida F es KnCr (Figura 4). En este caso, la curva obtenida tiene una semejanza a la curva ideal casi tan cercana como en el caso de 5-MNG.

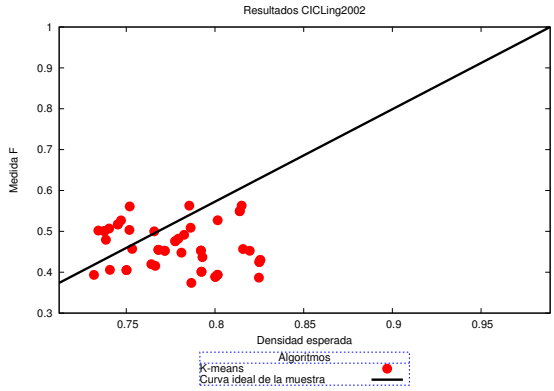

Figura 5: CICLing2002 (k-means)

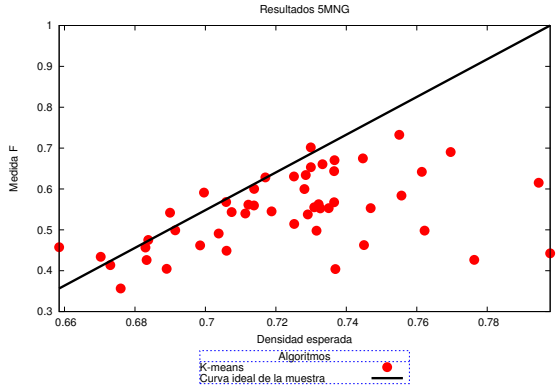

Figura 6: 5-MNG (k-means)

El segundo grupo de resultados se obtuvieron con el algoritmo k-means (con el número correcto de grupos) y se muestran en las Figuras 5, 6, 7 y 8. En los casos de Hep-ex y KnCr no se observa que un incremento en la densidad esperada implique un aumento de la correspondiente medida F. En el caso de 5-MNG en cambio, parece haber una relación más directa entre el crecimiento de la densidad esperada y el crecimiento de F. No obstante ésto, los valores de F comienzan a ser más inestables con valores de densidad superiores a 0.73. Considerando que en el caso de CICLing2002 tampoco se visualiza una relación clara entre la densidad y la medida F,

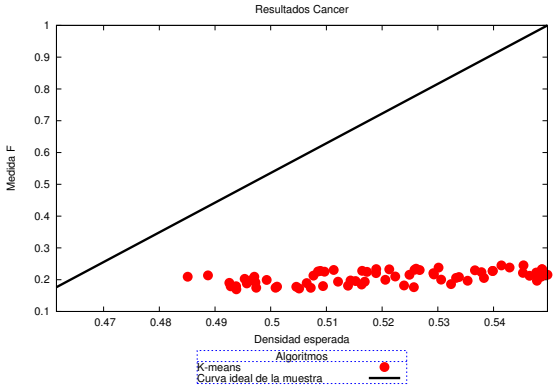

Figura 7: KnCr (k-means)

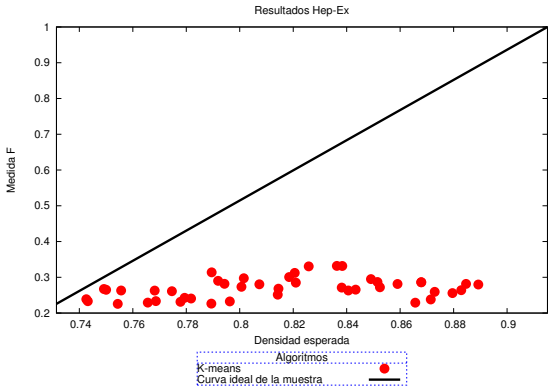

Figura 8: Hep-ex (k-means)

podemos inferir que si bien en un corpus con documentos completos y temáticas diferenciadas como 5-MNG, los resultados son consistentes con los obtenidos por Stein, en el caso de colecciones de resúmenes de dominios restringidos esta relación entre ambas medidas no parece verificarse.

Los resultados obtenidos con las colecciones de resúmenes no mejoraron cuando se utilizó un algoritmo como MajorClust que determina automáticamente el número de grupos que tendrá el resultado, ya que no cuenta con información sobre el número correcto de grupos como en los algoritmos previos. Como ejemplo representativo de estos resultados, en la Figura 9 se muestra el desempeño de MajorClust con la colección CICLing2002. Se puede observar que se tiene un rango más amplio de valores de densidad que con los dos algoritmos previos, debido a que la variación en el número de grupos hacen variar significativamente los valores de densidad. Sin embargo, con estos valores mayores de densidad esperada tampoco se percibe una mejora de la medida F.

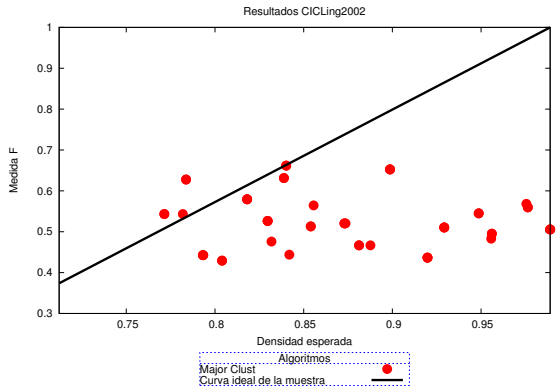

Figura 9: CICLing2002 (MajorClust)

5. Conclusiones y trabajo futuro

Los resultados obtenidos en este trabajo con la colección 5-MNG confirman las observaciones realizadas por Stein respecto a que la densidad esperada puede ser un buen indicador de la medida F cuando se agrupan documentos completos de temáticas disímiles. Sin embargo, esta relación entre ambas medidas no parece verificarse en tareas de agrupamiento de resúmenes de dominios muy reducidos. Estos resultados se constituyen en nuevos indicadores de la dificultad intrínseca de este tipo de dominios. Como trabajo futuro, sería interesante analizar el desempeño de otras medidas internas como el *índice de Davies-Boulding* o el *índice de Dunn*, en este tipo de dominios y su relación con la medida F. En base a estos estudios, sería factible incorporar la medida interna más adecuada en los algoritmos que las utilizan en alguna de sus etapas. De esta manera, se podría lograr un algoritmo de agrupamiento aceptable, adaptado a las características de este dominio tan dificultoso.

Bibliografía

Alexandrov, M., A. Gelbukh, y P. Rosso. 2005. An Approach to Clustering Abstracts. En *Proceedings of the 10th International Conference NLDB-05*, LNCS, páginas 275–285. Springer-Verlag.

Ingaramo, D., G. Leguizamón, y M. Errecalde. 2005a. Adaptive clustering with artificial ants. *Journal of Computer Science and Technology*, 5(04):264–271.

Ingaramo, D., G. Leguizamón, y M. Errecalde. 2005b. Clustering dinámico con hormigas artificiales. En *Proceedings of the CACIC 2005*.

Jiménez, H., D. Pinto, y P. Rosso. 2005. Uso del punto de transición en la selección de términos índice para agrupamiento de textos cortos. En *Procesamiento del Lenguaje Natural*, páginas 383–390.

Makagonov, P., M. Alexandrov, y A. Gelbukh. 2004. Clustering abstracts instead of full texts. En *Proc. of the TSD-2004*, páginas 129–135.

Montejo, A. y L. A. Ureña. 2006. Binary classifiers versus adaboost for labeling of digital documents. En *Procesamiento del Lenguaje Natural*, páginas 319–326.

Pinto, D., H. Jimenez, y P. Rosso. 2006. Clustering Abstracts of Scientific Texts Using the Transition Point Technique. En A. Gelbukh, editor, *Proceedings of the CICLing 2006*, volumen 3878 de *LNCS*, páginas 536–546. Springer-Verlag.

Pinto, D. y P. Rosso. 2006. Kncr: A short-text narrow-domain sub-corpus of Medline, TLH 2006.

Pinto, D., P. Rosso, J. Alfons, y H. Jiménez. 2006. A comparative study of clustering algorithms on narrow-domain abstracts. En *Procesamiento del Lenguaje Natural*, páginas 41–49.

Rose, T.G., M. Stevenson, y M. Whitehead. 2002. The reuters corpus volume 1: from yesterdays news to tomorrows language resources. En *Proceedings of the Third ICLRE*, páginas 29–31.

Salton, Gerard. 1971. *The Smart Retrieval System: Experiments in Automatic Document Processing*. Prentice Hall.

Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1–47.

Stein, B., S. Meyer, y F. Wißbrock. 2003. On Cluster Validity and the Information Need of Users. En *Proceedings of the 3rd IASTED*, páginas 216–221, Anaheim, Calgary, Zurich, Septiembre. ACTA Press.

Stein, B. y O.Ñiggemann. 1999. On the Nature of Structure and its Identification. volumen 1665 LNCS de *Lecture Notes in Computer Science*, páginas 122–134. Springer, Junio.