

Parámetros en torno a LSA y la evaluación automática de respuestas abiertas: el caso de la dimensionalidad



Guillermo Jorge-Botana¹, Ricardo Olmos², José A. León²

¹ Universidad Complutense de Madrid, ² Universidad Autónoma de Madrid

Introducción y objetivos

- La reducción de las dimensiones se asume como el factor más identificativo de la técnica LSA. Wiemer-Hastings (1999) es el primero en cuestionar su eficiencia en todo tipo de corpus ya que encuentra que con una versión de LSA sin SVD (sólo cálculos de pesos de entropía y comparaciones geométricas), obtienen unos resultados que se aproximan mucho a la versión completa (con SVD).
- En estudios que han analizado corpus específicos de dominio han concluido que la reducción de dimensionalidad no resulta más eficiente que la aplicación del total de la dimensionalidad (Kontostathis et al., 2005; Silva et al., 2004; Kurby et al., 2003; Olde et al., 2002; Cox & Shahshani, 2001)
- Paradójicamente, en corpus específicos de dominio, ha sido habitual la dimensionalidad menor a 150 dimensiones (Wolfe et al., 1998; Foltz et al., 1996; Wiemer-Hastings, 1999), justificándose por la propia naturaleza y tamaño de los corpus e, incluso, se han llegado a emplear dimensionalidades en torno a 30 (Nakov, 2000; Nakov et al., 2001).

Método

- Participantes: 40 alumnos de 4º de psicología considerados como expertos y 41 alumnos de 1º de logopedia considerados como inexpertos sobre este tema.
- Responden a la pregunta ¿Qué es la "fobia social"? Y cuyas respuestas son evaluadas por 3 profesores expertos en psicología clínica, y se comparan con la evaluación del LSA.
- Evaluación del LSA: Se crean tantos espacios semánticos como posibles combinaciones de los parámetros. LSA extrae la semejanza entre la respuesta de cada alumno con un texto ideal redactado por un experto en la materia.
- Finalmente, se comparan las evaluaciones realizadas por el LSA con la de los expertos.

Corpus

- CORPUS:** 1.DSM, 2.DSM+CIE, 3. PSI 4.DSM+PSI, 5. DSM+CIE+PSI.
- Cinco textos diferentes compuestos por tres corpus básicos: **DSM, CIE, PSI**
 - **DSM:** Corpus extraído de la taxonomía DSM-IV y se trata de un texto en el que la información está jerarquizada, estructurada y con poca variabilidad de términos..
 - **CIE:** Corpus extraído de CIE-10 de trastornos mentales. Como DSM, es jerarquizado, estructurado, con poca variabilidad de términos y sin información tangencia de la materia.
 - **PSI:** Corpus extraído de textos de internet sobre psicología de temática muy variada y con muy poca estructura y jerarquía, yr más variabilidad en cuanto a sus términos.
 - Con estos tres corpus básicos se construyen los 5 textos que serán procesados. (DSM, 2.DSM+CIE, 3. PSI 4.DSM+PSI, 5. DSM+CIE+PSI). A cada uno de ellos se le extrae un índice de cohesión basado en las medidas de relación de orden 1 y orden 2 (Mill & Kontostathis, 2004; Kontostathis & Pottenger, 2006).

Corpus	Matrices					
	Tamaño del texto digitalizado	Nº Palabras	Nº caracteres (sin espacios)	Nº Términos Purga Partículas	Nº Documentos Purga Part+Verbos	Órdenes Índice Orden 1 Índice Orden 2
Dsm	113.581	846.932	4258,242	2616,242	16,58	4,8
dsm+cie	162.517	916.016	5416,446	3375,446	13,51	3,52
Psi	141.045	765.932	6844,717	4103,717	6,24	0,89
dsm+psi	254.627	1.407.865	8877,959	5335,959	7,59	1,38
dsm+cie+psi	303.565	1.684.094	9603,1163	5821,1163	7,43	1,36

Preproceso y Dimensionalidad

PREPROCESO: 1.ENTROPIA, 2.IDF, 3.NADA

Se tomarán dos posibles variantes en las que ambas siempre tomarán en cuenta el peso local. Entropía e IDF (ver, Navok et al., 2001). También la ausencia de preproceso.

DIMENSIONALIDAD: 1. D20, 2. D40, 3. D60, 4. D80 5. D100

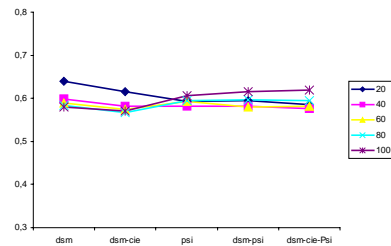
En lugar de tomar valores absolutos, tomamos valores porcentuales de acumulación de valores singulares como propuso Wild et al (2005). De esta forma, formamos 5 condiciones: 20 %, 40 %, 60 %, 80 % y 100 % .

Procedimiento

1. Tipificación de cada una de las evaluaciones extraídas tanto de LSA como de los jueces humanos. Esto homogeneizará todas las medidas.
2. Se resta cada puntuación que da cada uno de los espacios de LSA a un resumen, menos la puntuación dada por los jueces humanos a ese mismo resumen. El resultado se ofrece en valor absoluto, ya que no interesa si la diferencia es por defecto o por exceso, sino la magnitud de esta. Interesa una medida objetiva de bondad.
3. Se lleva a cabo un ANOVA sobre estos valores absolutos.
4. La medida introducida en este ANOVA es una medida objetiva de cuan distante está el criterio humano de cada una de las condiciones de LSA.

Resultados

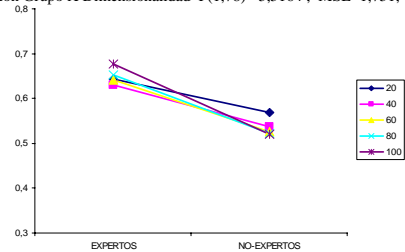
- Interacción Corpus X Dimensionalidad $F(1,78)=7,610$; $MSE=1,028$, $P=.007$



Corpus	Purga Partículas	Purga Part+Verbos	20%	40%	60%	80%	100%
dsm	4258,242	2616,242	26	64	110	166	242
dsm+cie	5416,446	3375,446	45	112	195	298	446
psi	6844,717	4103,717	61	150	268	414	717
dsm+psi	8877,959	5335,959	82	203	359*	562	959
dsm+cie+psi	9603,1163	5821,1163	101	251*	444*	695*	1163

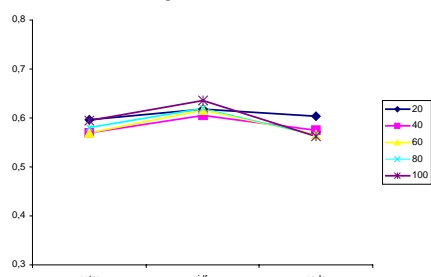
- La efectividad de la reducción de la dimensionalidad es sensible a las propiedades de los corpus lingüísticos. Un fenómeno que se observa es el corte de tendencia que se produce a partir del corpus PSI. Los corpus que carecen de información tangencial (DSM y DSM+CIE) provocan que ciertos términos prevalecen como términos cruciales (Franceschetti et al., 2001; Olde et al., 2002). La ocurrencia de estos términos lleva el peso de la evaluación por lo que no se hace tan necesario extraer información más fina sin necesidad de llevar a cabo por la reducción las dimensiones. Como consecuencia, la totalidad de las dimensiones (100%) es igual de efectiva que las demás. En los demás corpus (PSI, DSM+PSI y DSM+PSI+CIE), algunas dimensiones, (señaladas con asterisco), son significativamente más efectivas que la dimensionalidad total.

- Interacción Grupo X Dimensionalidad $F(1,78)= 5,5164$; $MSE=1,751$; $P=.021$



- Analizando el tipo de respuesta es posible que nos encontremos con que la reducción de dimensionalidad sea útil en la evaluación de ensayos elaborado por expertos, en la que no se valore únicamente la aparición o no de los términos cruciales de alta frecuencia. Por ejemplo, a la pregunta de que es la fobia social se responde con unos cuantos términos cruciales: "Un tipo de FOBIA o un MIEDO a la GENTE"

- Interacción Dimensionalidad X Preproceso $F(1,78)= 4,577$; $MSE=0,841$, $P=.050$



- La interacción muestra que si se tiene como propósito reducir la dimensionalidad para formar los espacios, resulta conveniente llevar a cabo previamente algún cálculo corrector que pondere la importancia de los términos en los documentos, como Entropía o IDF.

Conclusiones

- Es posible que en algunas ocasiones la reducción de dimensionalidad no aporta ninguna ventaja frente al empleo de la totalidad de las dimensiones y que en algunos estudios se haya dado por hecho una supremacía de la reducción de dimensiones cuando en realidad no existía.
- En esta misma línea y en contra de lo que se ha sugerido, el número de dimensiones óptimo en corpus específicos de dominio no tiene por que ser un número extremadamente bajo sino que en ocasiones, supera incluso las 300 dimensiones recomendadas para los corpus de dominio general.
- En el grupo de No-Expertos, la dimensionalidad total (100%) tiende a obtener iguales o mejores resultados que las reducciones de dimensionalidad más efectivas. Esto puede significar que en el grupo de los No-Expertos la efectividad de la evaluación viene dada por la ocurrencia de algunos términos cruciales de alta frecuencia sin requerir una representación fina de los demás términos.