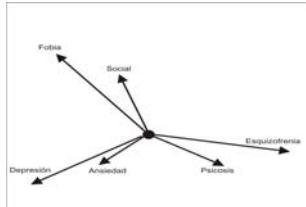


Introducción

- El Análisis Semántico Latente (LSA) es una herramienta informática que analiza relaciones semánticas entre diferentes unidades lingüísticas de forma completamente automatizada (Landauer y Dumais, 1997). El LSA crea un espacio semántico multidimensional en el que todas las palabras quedan representadas matemáticamente en forma de **vectores**.



Objetivos del estudio

- Analizar la fiabilidad del LSA en la evaluación de resúmenes utilizando tres nuevos algoritmos y compararlo con otro estándar como línea base.
 - ¿Mejora la fiabilidad de las evaluaciones con estos nuevos algoritmos en comparación con los tradicionales?
 - ¿Hasta qué punto evalúa correctamente el LSA resúmenes tan concisos y breves como los empleados en este estudio?
 - ¿Qué implicaciones tienen los algoritmos desde un punto de vista psicológico y matemático para el LSA?

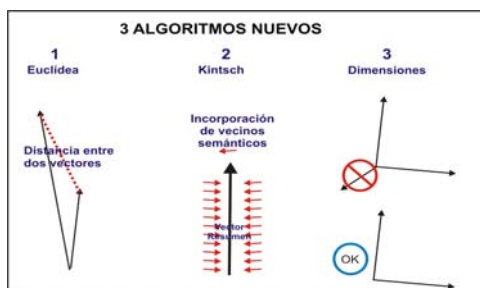
Algoritmos del estudio

Para evaluar un resumen con el LSA utilizamos dos métodos.

- (1) Comparamos la relación semántica entre el resumen y el texto con el coseno del ángulo formado entre dos vectores. Este método se conoce como RESUMEN-TEXTO. La lógica de este método es simple: cuanto mayor sea la similitud entre el resumen y el texto mejor es la calificación del resumen, ya que se infiere que mejor están representados los contenidos semánticos del texto en el resumen.
- (2) 6 expertos elaboraron lo que denominamos **RESÚMENES IDEALES**. Estos resúmenes se comparan con un resumen de cada estudiante a evaluar. La calificación es la media de la similitud semántica entre el resumen y los 6 resúmenes ideales (MÉTODO DEL EXPERTO).

Los cuatro algoritmos empleados son:

- Algoritmo estándar.** Los vectores que representan a los textos son los que se derivan del espacio semántico del LSA (coseno del ángulo).
- Algoritmo de predicación adaptado** (Kintsch, 2001). Al vector de cada resumen del estudiante, se le añaden los vectores de las 50 palabras semánticamente más relacionadas con dicho resumen. De esos 50 nuevos vectores se escogen aquellos 20 que a su vez estén más relacionados semánticamente con el texto que resumen. Esta idea supone añadir al resumen más información que la contenida explícitamente en él, como si añadiésemos conocimiento previo semánticamente relacionada al contenido de un resumen.
- Algoritmo mejores dimensiones.** Habitualmente, el LSA utiliza un vector con tantas dimensiones como haya quedado el espacio semántico reducido. Sin embargo, no todas las dimensiones tienen por qué ser buenas a la hora de evaluar un resumen (la idea se recoge de Hu, Cai, Franceschetti, Penumatsa, Graesser, Lowerse, McNamara y TRG, 2007). Este algoritmo va descartando paso a paso aquella dimensión que tras eliminarla del vector, hace que mejore más la fiabilidad del LSA respecto a los jueces humanos. Primeramente, se elimina la peor dimensión de las p originales. En un segundo paso se elimina la peor dimensión de las restantes $p-1$ dimensiones y así sucesivamente hasta eliminar el 20% de las dimensiones que peor funcionan. Con este algoritmo conseguimos tener en cuenta únicamente las dimensiones semánticas que contribuyan más a evaluar los resúmenes.
- Algoritmo distancia euclídea.** La medida utilizada prácticamente en todos los estudios con el LSA ha sido el coseno entre dos textos. En este algoritmo, hemos sustituido la medida del coseno por la medida de la distancia euclídea entre dos vectores (entre dos textos). La distancia euclídea depende del ángulo entre los dos vectores, pero también de la longitud de los vectores. En LSA se ha dicho frecuentemente (Rehder, 1998) que el módulo de un vector representa la cantidad de conocimiento del LSA que tiene del texto. La distancia euclídea nos permite evaluar resúmenes teniendo en cuenta no solamente el ángulo entre dos vectores, sino también considerando la cantidad de información que contiene un resumen.

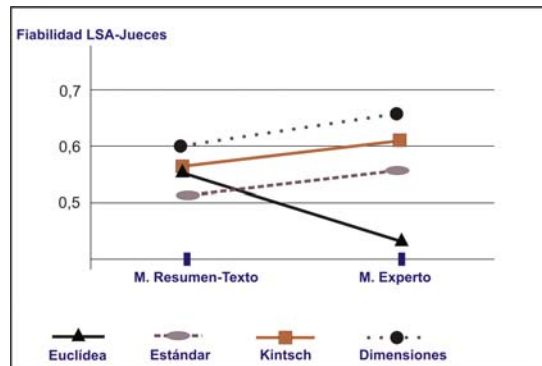


Procedimiento

- El material del estudio contó con 192 resúmenes de estudiantes de primaria y de ESO de un texto expositivo "Los árboles estranguladores", que cuenta cómo es el proceso de supervivencia de unas plantas epifitas en las selvas tropicales. El texto tenía exactamente 500 palabras y se impuso que los resúmenes tuvieran un máximo de 50 palabras.
- Los resúmenes fueron evaluados en una escala del 1 al 10 por cuatro jueces entrenados para ello. Los jueces tuvieron en cuenta que el resumen tuviese los contenidos fundamentales del texto y la coherencia con la que se expresan dichos contenidos.
- El corpus utilizado en este estudio para entrenar al LSA fue específico de dominio. Contó con 372 documentos de contenidos similares al texto y 5995 palabras distintas. Las palabras están lematizadas, de manera que cualquier forma verbal queda en espacio semántico como un mismo vector y tampoco se distinguieron singulares y plurales o masculinos y femeninos en vectores distintos. Con esto se consigue minimizar el efecto de la gramática sobre el LSA.
- Para analizar los resultados se llevó a cabo un ANOVA de 2 factores completamente aleatorizado. La variable dependiente fue las fiabilidades LSA-jueces. Los factores fueron: 1) Los cuatro algoritmos comparados (estándar, predicación de Kintsch adaptado, mejores dimensiones y distancia euclídea; y 2) Los dos métodos de evaluación (Resumen-Texto y Expertos).

Resultados

- No encontramos efecto del método ($F(3, 24) = 0.01, MSE = 0.02, p = 0.94$)
- Encontramos efecto del algoritmo ($F(3, 24) = 11.19, MSE = 0.02, p < 0.05$)
- Encontramos efecto de la interacción ($F(3, 24) = 6.32, MSE = 0.02, p < 0.05$)



- Lo primero que se puede apreciar en el gráfico de perfil es que con la salvedad del algoritmo de la distancia euclídea en el método del Experto, todas las fiabilidades LSA-Jueces superan el 0,50. El algoritmo de las mejores dimensiones supera incluso la fiabilidad 0,60 en ambos métodos.
- Además, se aprecia que el algoritmo de las mejores dimensiones funcionó mejor que los restantes tanto en el método Resumen-Texto como en el de los Expertos. La distancia euclídea funcionó relativamente bien en el método Resumen-Texto pero baja mucho su rendimiento en el método de los expertos. El algoritmo estándar y el adaptado de Kintsch no difirieron significativamente, sin embargo puede deberse a una escasa potencia del análisis debido al tamaño de la muestra, ya que la tendencia es que el algoritmo de Kintsch también mejora el algoritmo estándar.

Conclusiones

- El mejor algoritmo fue el de las mejores dimensiones. El algoritmo implica reducir el número de dimensiones que van a participar en la evaluación de los resúmenes, eliminando para ello el 20% de las dimensiones que peores fiabilidades LSA-jueces arrojen. Este algoritmo supera el valor 0,6 en fiabilidad LSA-jueces, lo cual es un valor meritorio habida cuenta de que la brevedad de los resúmenes representa una dificultad muy seria para el LSA. Cuanto más breve sea un resumen más le limita al LSA la información semántica contenida en el vector. Los mejores resultados en evaluación de textos escritos se consiguen con longitudes que superan las 200 palabras (Rehder, 1998), los nuestros nunca superaron las 50 palabras.
- Por otra parte, el algoritmo adaptado de Kintsch también supera el valor 0,6 con el método de los expertos. Aunque no fue significativamente mejor que el algoritmo estándar, parece que existir una tendencia a mejorar sus resultados. La falta de potencia de este análisis (sólo se cuenta con 4 jueces) es posible que haya impedido detectar diferencias significativas. La virtud de este algoritmo es que tiene más verosimilitud psicológica al basarse en el modelo de comprensión de Kintsch (1998).
- El algoritmo de la distancia euclídea resulta el peor de todos, sobre todo con el método de los expertos. Esta medida, habitualmente descuidada por los teóricos del LSA, ha funcionado correctamente en otros estudios (Jorge-Botana, León y Olmos, en prensa), por ejemplo para comparar textos entre expertos y novatos. La virtud de esta medida está en que incorpora simultáneamente la medida del coseno y de la longitud del vector.