Presented in the Twenty-third Annual Meeting of the Society for Text and Discourse, Valencia from 16 to 18, July 2013

Giving an interpretation for the semantic dimensions in Latent Semantic Analysis

Olmos[1], R., Jorge-Botana[2], G., León[1], J.A., y Escudero[2], I.

1. Universidad Autónoma de Madrid;
2. Universidad Nacional Educación a Distancia (UNED)

Abstract

This research has three main objectives: (1) following the procedure for interpreting LSA space dimensions (Hu, et al, 2005), we implement a change of basis, from the original canonical basis into a basis whose vectors are real terms; (2) we show that a simple change of basis alone is not enough for this purpose; and (3) we present Gram-Schmitdt orthogonalization to partially correct this inefficiency. We used this procedure on a corpus taken from several Spanish newspapers. We chose essential terms from the original term matrix, such as "terrorism", "president", "police", etc. as vectors for the new basis so that, in the new term matrix generated after the change of basis, it is possible to say how much each word carries from these terms (how much each term in the semantic space carries from the terms in the basis). We conclude that Gram-Schmidt orthogonalization is a way to correct the change of basis.

1. **The starting point**

As is well known, the final products of the LSA procedure are three matrices (**X** ≈ **USV**´) given for the mathematical technique known as Singular Value Decomposition (Deerwester, Dumais, Furner, Landauer, & Harshman 1990; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). One of these matrices is the term matrix **U**, which is multiplied by the diagonal matrix **S** in order to weight each dimensions by a coefficient (the coefficients are singular values of **X**). The important fact is that we have a matrix **US** whose rows represent words in a semantic space resulting from the LSA process. Each of these vector-words is expressed in the canonical basis, which is orthogonal and meaningless. So the dimensions are not interpretable. They only serve to point a word in a *n*-dimensional space and to calculate distances, but they cannot be interpreted as topics. It is possible to interpret each word-vector by means of its semantic neighborhood, but it is not possible to interpret a word-vector by examining seeing its components along. The question is: how could this be made possible?

2. **The procedure**

   a. **Change of basis**

The procedure we use to change the original and arbitrary coordinates in a new space where the new coordinates are meaningful is based on some linear algebra methods. It is implemented in two steps. The first one is to perform a simple change of basis from the canonical basis to another basis. The vectors in the latter basis are real words from the semantic space, so we can generate every vector word in the semantic space using a few meaningful vectors. It is clear that this vector sect must be linearly independent in order to become a basis. We use this set to turn the meaningless dimensions in the original space into meaningful dimensions.

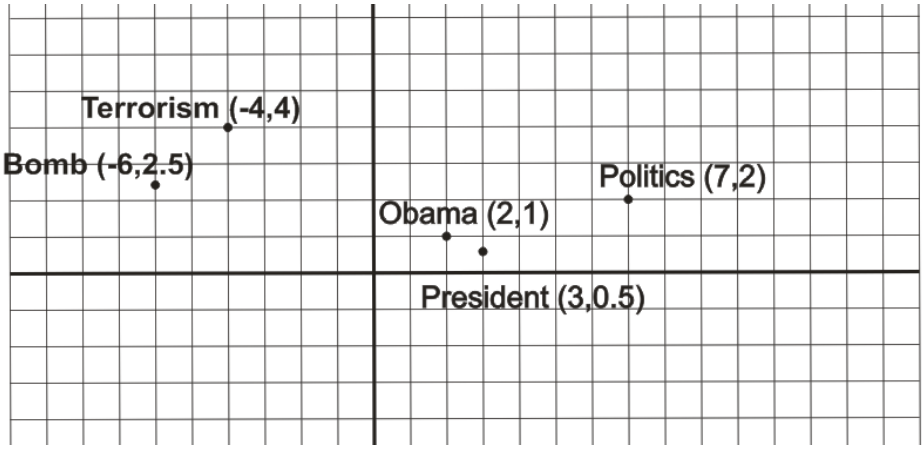
Let us call this new basis $\boldsymbol{\beta} = \{V_{family}, Vf_{ootball}, V_{war}, V_{maths}, \ldots, V_{n}\}$. If we sort $\boldsymbol{\beta}$ vectors as columns in a matrix **P**, using linear algebra, we can express the terms of the matrix **US** in

the new basis **β**, obtaining a new term matrix **C** whose dimensions are meaningful in theory (as meaningful as the vectors in the new basis **β**) . Each vector $\acute{c}_\iota$ in the new term matrix **C** can be calculated by multiplying $P^{-1}$by each vector $\acute{u}_\iota$in the old term matrix **US**. This possibility was suggested by Hu, et al. (2005):

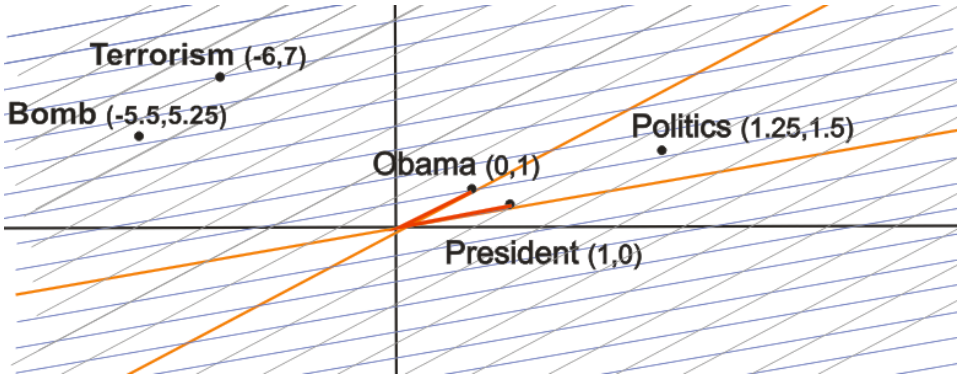
$$\acute{c}_\iota = P^{-1}\acute{u}_\iota$$

But given the nature of the vectors in the basis and the obliqueness ratio, there is a risk of distorting the old latent relations if we use change of basis alone. To solve the obliqueness problem we need to force an orthogonal basis and this is possible if a Gram-Schmidt algorithm is used (see for example Schneider, Steeg, & Young, 1987). But, before we present Gram-Schmidt orthogonalization, we would like to show graphically the problems associated with an oblique basis.

Let there be an original space where five terms are represented:



Suppose we choose a new basis β with a new β-coordinate vector formed by the terms President and Obama. Obviously, the new basis is oblique because Obama and President share a strong semantic relationship. The new basis and new β-coordinates are now represented as follows:

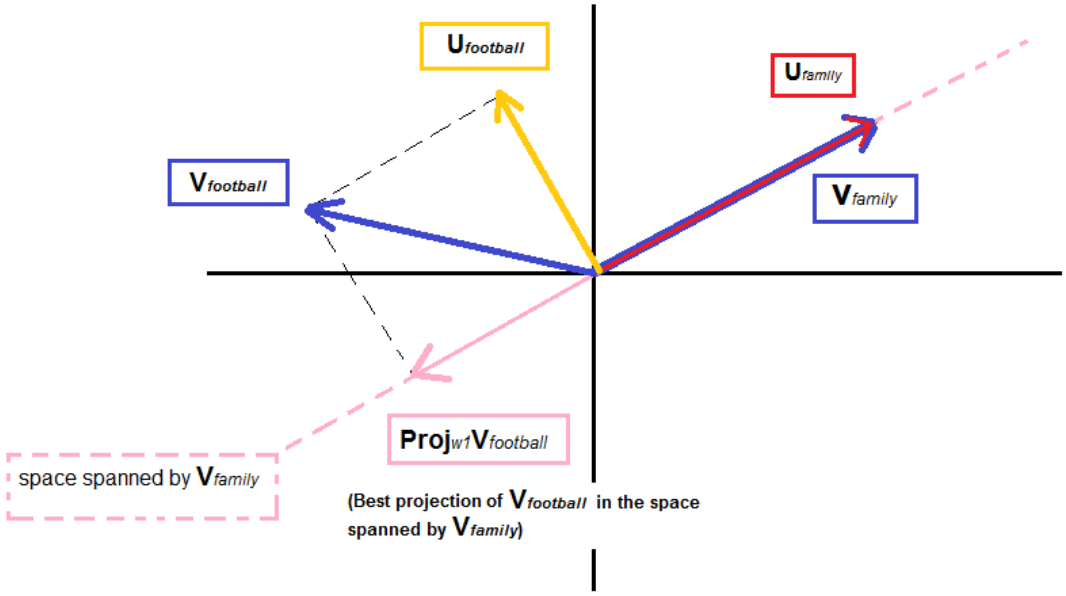


In the β-coordinate, Politics hardly has any from the Obama dimension or from the President dimension. However, Terrorism and Bomb have higher coordinates in the new space than Politics and this is counter-intuitive. Moreover, the norms for the terms Terrorism and Bomb are much larger than the norm vector Politics. The norm influences the cosines, so this is another way of seeing how an oblique space distorts the original cosines. These distorted coordinates to represent semantic similarities impose an orthogonal basis. This is how Gram-Schmidt works.

b. **Gram-Schmitdt**

The Gram-Schmidt approach is a step-by-step method to re-orthogonalize a basis that is not orthogonal. Let us return to our example using the words in **β**. Given a non-orthogonal basis **β** $=\{\mathbf{V}_{\text{family}}, \mathbf{V}_{\text{football}}, \mathbf{V}_{\text{war}}, \mathbf{V}_{\text{maths}}, \ldots, \mathbf{V}_{\text{n}}\}$, for the current semantic space, the Gram-Schmidt algorithm builds an orthogonal basis $\{\mathbf{U}_{\text{family}}, \mathbf{U}_{\text{football}}, \mathbf{U}_{\text{war}}, \mathbf{U}_{\text{maths}}, \ldots, \mathbf{U}_{\text{n}}\}$ based in **β**. Notice that the vectors are this time labeled by **U**, because when orthogonalization is forced by means of Gram-Schmidt, if the word-vector $\mathbf{V}_{\text{football}}$ is used to generate a dimension for the new semantic space, Gram-Schmidt will choose a vector similar to $\mathbf{V}_{\text{football}}$, but not exactly $\mathbf{V}_{\text{football}}$. It choses $\mathbf{U}_{\text{football}}$. This is because the final basis is forced to be orthogonal.

Let us see the procedure and its consequences:



Step 1: Let (**U**family) = (**V**family)
so the vector **U**family is the same as **V**family.

Step 2: Let (**U**football) = (**V**football) – (**Proj**$_{W1}$ Vfootball)
where $\mathbf{W}_1$ is the space spanned by **V**family, and **Proj**$_{W1}$ **V**football is the orthogonal projection of **U**football, on $\mathbf{W}_1$.

Step 3: Let (**U**war) = (**V**war) – (**Proj**$_{W2}$ **V**war)
where $\mathbf{W}_2$ is the space spanned by {**U**family, **U**football} and **Proj**$_{W2}$ **V**war is the orthogonal projection of $\mathbf{V}_{war}$, on $\mathbf{W}_2$.

Step 4 Let (**U**maths) = (**V**maths) – (**Proj**$_{W3}$ **V**maths)
where $\mathbf{W}_3$ is the space spanned by {**U**family, **U**football, **U**war} and **Proj**$_{W3}$ **V**maths is the orthogonal projection of Vmaths, on $\mathbf{W}_3$.

Continue this process up to $\mathbf{U}_n$

The resulting set {**U**family, **U**football, **U**war, …, $\mathbf{U}_n$}, consists of a basis with *n* linearly independent re-orthogonalized real word vectors. The advantage of Gram-Schmidt is that orthogonalization of the basis preserves 100% of the cosines within the former latent semantic space in the new latent semantic space, avoiding distortion of the relations between the terms of the new term matrix **C**. In fact, it is a simple rotation between two orthonormal bases.

But when this procedure is used, the degree of de-virtualization of each substitute word {**U**family, **U**football, **U**war, …, $\mathbf{U}_n$}, should be taken into account for interpretation purposes.

When the orthogonalization process is performed, little errors are accumulated due to the movement magnitude of each projection. So, the more cycles (the more **U**s projected to **W**s), the greater the accumulated error. Due to this, it is important to set a threshold which indicates that after passing **U**s, its interpretation is not reliable. So, we can only interpret words in the current semantic space using the reliable ones. How can we find such a threshold?

A good measure is the correlation between the vectors in the former basis and the vectors generated by the Gram-Schmidt procedure. This shows to what extent the Gram-Schmidt preserves the characteristics of the vector of the word chosen to create the new basis. We can consider such a measure as the reliability of a term to represent that term. This is very important since we interpret all the words in a semantic space by believing whether a word is true or not. As a recommendation, values lower than 0.70 should prevent us from interpreting the new dimension by means of the chosen word, as this means that the vector generated by Gram-Schmidt shares less than 50% of the variance with the original word ($0.70 \approx 0.50^2$).

Given the explanation above, and given that Gram-Schmidt is a sequential process, we can also take a basis where the first vectors are real words and the remaining vectors are abstract vectors in the canonical basis. In this way, we make errors accumulate in these abstract vectors, preserving real words from distortion.

3. Some results

The procedure was run on Gallito 2.0, a tool programmed and maintained by some of the authors of this paper (Jorge-Botana et al., 2012). In the following paragraphs, we present some informal examples of gradients generated using different procedures.

Gram-Schmidt change of basis. EXAMPLE

Figure 1 represents the coordinates of the term “Bomb” in the new latent semantic space after carrying out a change of basis using the Gram-Schmidt method. It is obvious that new space has a clear interpretation. It has a very large coordinate in the “attempt” dimension (with a reliability of .92) and middle-scale saturation in other ones such as “Madrid”, “Basque” (country), “terrorism”, also with acceptable reliabilities. To verify this hypothesis, we extract, for example, the first semantic neighbors in the semantic space of the term “Bomb”: the first 10 semantic neighbors include *explode*, *detonate* and *explosive*. In the new basis there is no

dimension which has exactly this meaning, but *attempt* is obviously semantically related to this term. Thus, the new coordinate system "understands" that the term Bomb must saturate in this dimension. Figure 1 also shows that errors (low reliabilities) have accumulated in the abstract dimensions of the new basis, that is to say, on the basis vectors that have been filled up with vectors from the canonical basis.

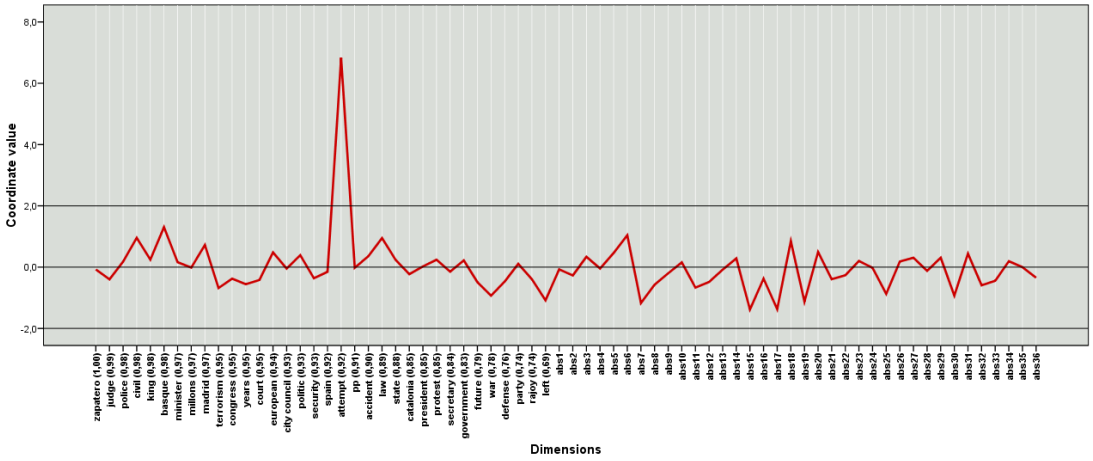


Figure 1

Figure 2 represents the coordinates of the term "Bomb" in the new latent semantics, but without performing Gram-Schmidt orthogonalization. The highest coordinates correspond to the man and woman dimensions. In any case, these dimensions are not clear enough to correctly represent the term Bomb. The initial conclusion is that the loss of orthogonality causes a loss in interpretability.

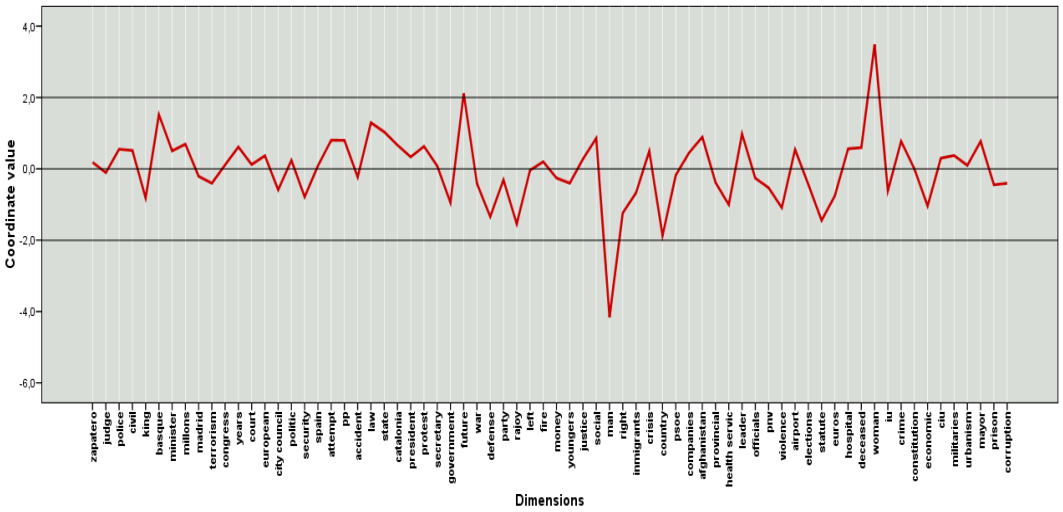


Figure 2

Figure 3 represents the coordinates of the term Bomb in the old latent semantics. It is clear that there is no particularly high coordinate . In this semantic space, as is well known, dimensions lack an interpretation: they are pure abstractions, derived from the mathematical application of Singular Value Decomposition.

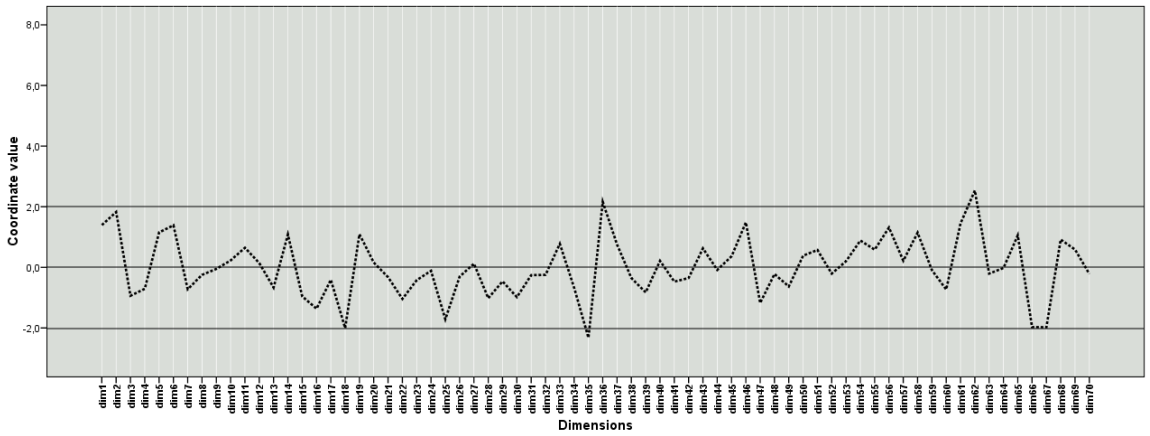


Figure 3

References

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., y Harshman, R. (1990). Indexing By Latent Semantic Analysis. *Journal of the American Society For Information Science* , 41, 391-407.

Hu, X., Cai, Z., Wiemer-Hastings, P., Graesser, A. C., y McNamara, D. (2005). Strengths, Limitations, and Extensions of LSA. In D. McNamara, T. Landauer, S. Dennis, and W. Kintsch, editors, *LSA: A Road to Meaning*, Erlbaum, Mahwah, NJ.

Jorge-Botana,G., Olmos,R.,& Barroso, A. (2012). Gallito (Version 2.0.1)[NLP Software]. Retrieved from http://www.elsemantico.es/descargas-eng.html

Landauer, T. K., y Dumais, S. T. (1997) A solution to Plato's problem: the Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Landauer, T. K., Foltz, P. W., y Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes,* 25, 259-284.

Schneider, D.M., Steeg, M., & Young, F.H. (1987). Linear Algebra: A Concrete Introduction. 2nd Edition. Publisher, Simon & Schuster Books.