Expert Systems With Applications 131 (2019) 71–80

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Could LSA become a “Bifactor” model? Towards a model with general and group factors

Guillermo Jorge-Botana[a,*], Ricardo Olmos[b], José María Luzón[a]

[a] Universidad Nacional de Educación a Distancia, c/ Juan del Rosal, 10. 28023 Madrid, Spain
[b] Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, c/ Iván Pavlov, s/n., 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 8 July 2018
Revised 22 April 2019
Accepted 22 April 2019
Available online 23 April 2019

Keywords:
Latent semantic analysis
Bifactor model
Distributional semantics
Inbuilt-Rubric method
Rotation
Text assessment

ABSTRACT

One insufficiently grounded criticism made against Latent Semantic Analysis is that it is impossible to semantically interpret its dimensions. This is not true, as several studies have transformed the latent semantic space to interpret them, by means of some methods. One of them is the Inbuilt-Rubric method. Rather than grouping concepts around dimensions, as in Exploratory Factor Analysis based rotation methods, the Inbuilt-Rubric is a method that perform an “a priori” imposition of concepts onto the latent semantic space. It uses a confirmatory strategy. This study seeks to propose solutions for two limitations found in the current Inbuilt-Rubric methodology: one solution is inspired by Bifactor Models and the management of common variance of the concepts involved; and the other one is based in randomizing the sequence to perform the process. Both methods outperform the current Inbuilt-Rubric version in relevant content detection. The reported improvements can be incorporated into expert systems that use Latent Semantic Analysis and Inbuilt-Rubric in relevant content detection or text classification tasks.

1. Introduction

Latent Semantic Analysis (LSA) is a computational model for the extraction of meanings which has been extensively used in the last two decades (Evangelopoulos, 2013; Tonta & Darvish, 2010). Basically, the LSA model obtains a representation of the meanings of text units (words, paragraphs, documents, etc.) using a reduced number of dimensions, usually about 300. To do so, in the LSA process a linguistic corpus which can include hundreds of thousands of semantic contexts (paragraphs, sentences, or documents) must be analyzed. By means of the mathematical technique called singular value decomposition (SVD), words and texts are vectorially represented in a high dimensional space which is usually known as the *latent semantic space (called “latent” as the vectors of the basis are meaningless)*. On the basis of this lexicon representation, LSA is used on many studies as a semantic model that is able to successfully simulate properly human tasks (Bhatia, 2017; Denhière & Lemaire, 2004; Günther, Dudschig, & Kaup, 2016; Huettig, Quinlan, McDonald, & Altmann, 2006; Kintsch, 1998; Kintsch & Mangalath, 2011; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998; Yeari & van den Broek, 2015, 2016). The capacity of these interesting properties of LSA has also been used in applied fields integrated in many expert and intelligent systems. Some examples are text categorization and classification systems (Thorleuchter, & Van den Poel, 2013; Zhang, Yoshida, & Tang, 2011), automatic summarization (Kireyev, 2008; Ozsoy, Alpaslan, & Cicekli, 2011), Business Analytics (Marcolin, Becker, Wild, & Schiavi, 2018) Chatbots with semantic Intelligence (Ranoliya, Raghuwanshi, & Singh, 2017; Thomas, 2016), semantic change monitoring (Balbi & Esposito, 1998; Contreras Kallens & Dale, 2018; Landauer, Kireyev, & Panaccione, 2011), voice recognition correction (Bellegarda, 2000; Lim, Ma, & Li, 2005; Shi, 2008), Call Routing Systems (Chu-Carroll & Carpenter, 1999; Jorge-Botana, Olmos, & Barroso, 2012; Serafín & DiÉugenio, 2004) and intelligent systems for the detection of the presence or absent of relevant content in texts, especially in student-constructed responses (Magliano & Graesser, 2012) being the later especially important, because the purpuse of this paper is analyze a method of this last group.

But although LSA can be considered a mature technique, one insufficiently grounded criticism made against LSA as oppose to other models, such as for example Topic Models (Steyvers & Griffiths, 2006), is that because its dimensions are originally latent, it is impossible to interpret its meanings. Hence, LSA has been usually regarded as a non-interpretable representation model (see a review of the confused aspects of LSA use in Jorge-Botana, Olmos, & Luzón, submitted). However, this is not true, as several

* Corresponding author.
E-mail addresses: gdejorge@psi.uned.es (G. Jorge-Botana), ricardo.olmos@uam.es (R. Olmos), jmluzon@psi.uned.es (J.M. Luzón).

https://doi.org/10.1016/j.eswa.2019.04.055
0957-4174/

recent studies have transformed the latent semantic space in order not only to find the meaning of its dimensions, but also use the LSA-based *n*-dimensional spaces more flexibly in expert systems which perform relevant content detection or text classification. These studies include two main streams. The first represented by those studies that use techniques proper to Factor Analysis (FA) to identify the meaning of some of the dimensions of the LSA space (Evangelopoulos, 2013; Evangelopoulos & Visinescu, 2012; Evangelopoulos, Zhang, & Prybutok, 2012; Kundu, Jain, Kumar, & Chandra, 2015; Visinescu & Evangelopoulos, 2014). These studies apply rotations to the original LSA space (e.g., Varimax, Equamax and Quartimax) to find a simple solution in which each dimension brings together words that identify it. This process makes it possible to extract a vector with semantically identified dimensions from the projection of new texts onto the rotated LSA semantic space (what is known as a folding-in projection, see Dumais, 2004), and thus to measure how much of each dimension can be found in each projected new text (saturations in factor analysis). These studies follow the logic of exploratory factor analysis, in the sense that identification of the dimensions depends on the usual rotation methods (e.g. varimax) aimed at providing the rotated space with parsimonious, simple solutions (Thurstone, 1931). However, a second stream of studies focuses, rather than on exploratory identifying concepts with dimensions, on the *a priori* imposition of meanings onto some dimensions of the latent semantic space. These studies differ from the previous ones, among other things, in that they follow a strategy that is basically (more) confirmatory. Thus, the contents of dimensions are imposed *a priori*; that is, the meanings imposed to the dimensions depend on specific tasks the expert system must deal with. More specifically, the method used by authors following this second line is called the *Inbuilt-Rubric method* (Martínez-Huertas et al., 2018; Olmos, Jorge-Botana, León, & Escudero, 2014; Olmos, Jorge-Botana, Luzón, Martín-Cordero, & León, 2016), as it originates in its first studies in automatic evaluation of student-constructed responses. It is important to point out that Inbuilt-Rubric methods have been used in real educational contexts with automatic writing tutoring by means of the G-Rubric online solution (see www.grubric.com), a prompt-based expert system for training in academic writing through feedback originating in various sources in real time (see Santamaría, Hernández, Sánchez-Elvira, Luzón, & Jorge-Botana, 2017). Even though more details about the Inbuilt-Rubric method will be given later, we will now just say that this method ensures that the first *k* dimensions of the semantic space (in other words, the *k* first vectors of the basis of the space) represent or "acquire" meaning (the concepts of a rubric), which makes it possible to identify, based on those concepts, any text unit that is projected onto that space. The advantages of imposing meanings rather than identifying them by means of more exploratory strategies are potentially two: the first one is the classification reference that is created *ad hoc*, in which its first *k* dimensions specifically represent the classification or categorization concepts needed in the expert system. This means that the Inbuilt-Rubric method provides great versatility for systems used for classification or content detection, and, above all, for systems for the automatic evaluation of academic texts, where it is ultimately most frequently employed. From a certain point of view, it is a pseudo-faceted form of classification that makes analysis from different point perspectives possible, serving as a prism to activate the semantic information required by a given task (e.g. detecting relevant content or evaluating texts under a rubric). Conversely, the exploratory methods for the identification of non-latent dimensions have the drawback that they depend on algorithms for the minimization or maximization of some criterion, so they cannot be easily adapted to specific tasks or requirements. Such identification strongly depends on the solution found. The second advantage is its versatility, as this method can impose that those dimensions that represent oblique concepts in the original latent space be orthogonally processed in the new semantic space with meaning (it behaves as a zoom effect).

In sum, even though both procedures - exploratory identification by means of a simple solution and the Inbuilt-Rubric method - use a linear algebra-based mathematical apparatus, in the case of Inbuilt-Rubric it is not exploratory so much as impositional, as it directly imposes meanings on some of the dimensions of the semantic space, the meanings needed to detect in the classification system or the meanings needed to be included in an content rubric. Since this study seeks to solve two limitations found in the current Inbuilt-Rubric method, the next section is dedicated exclusively to show the details of the procedure of Inbuilt-Rubric remarking some limitations to be solved.

2. Steps of current Inbuilt-Rubric method

The Inbuilt-Rubric procedure has three steps:

Step 1. Selection of a new algebraic basis to substitute the latent (meaningless) basis of the original space: the change in the basis of the LSA space was tentatively suggested as a potential extension by Hu, Cai, Wiemer-Hastings, Graesser, and McNamara (2007). This idea was later developed using the Inbuilt-Rubric procedure (Olmos et al., 2014; Olmos et al., 2016). In this first step, a basis other than that of the original latent space is selected whose first *k* vectors represent each of the rubric concepts. This new basis is now called **P**. The basis of the original latent space is the standard basis that will be called $\mathbf{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e_n}\}$, *n* being the dimensionality of the original latent space (usually, as stated, 300 dimensions). Each of those *k* vectors in **P** is calculated as follows: a group of experts previously identify the basic concepts that a rubric must include. In order to take that rubric to the new semantic space, a number of descriptors must be identified in the original semantic space. For example, if a topic is about fauna and botany, experts establish that the concept *insects* is a basic type of content that must be included in the rubric, the following terms could be identified as good representatives of that concept {“fly”, “diptera”, “bee”, “ant”, “chitin”, butterfly”}. Then the vector for the concept *insects* must be included in the new basis (as a new algebraic basis is being generated). One simple way to achieve this is to add the vectors of those descriptors previously identified (some vectors that represent some types of contents can also be removed to prevent the risk of polysemous meanings in concepts) To continue with our example, the vector for the new basis that represents the concept *insects* is then calculated as the sum of the vectors for the selected descriptors: $\boldsymbol{V}_{insects} = \boldsymbol{V}_{fly} + \boldsymbol{V}_{diptera} + \boldsymbol{V}_{bee} + \ldots + \boldsymbol{V}_{chitin}$. Then, to obtain the *k* vectors for the new basis, this procedure is followed with all the basic contents identified by the experts to constitute the rubric. For example, if another basic content referring to the concept *mammals* is identified, the $V_{mammals}$ vector will be composed of the sum of the vectors that represent the descriptors for the *mammal*s concept. Following this procedure for all the concepts in the rubric, finally, we will have the **P** basis with *k* vectors representing each of those concepts. To ensure that the new basis **P** has the same number of vectors (*n*) as the basis of the original latent space **E**, the basis **P** comprising those *k* vectors is filled with part of the standard basis **E**, as many vectors from the standard basis as are required to complete it, namely, a set of *m* vectors from the standard basis ($m = n - k$). These *m* vectors are taken randomly and with no repetitions. Preserving the same dimensions, the standard basis **E** and the new basis **P** will be two possible bases for a same space, the first one with all dimensions latent (meaningless), and the second one with the first *k* dimensions having the meaning of the concepts in the rubric. In this respect, $\mathbf{P} = \{\boldsymbol{V}_{insects}, \boldsymbol{V}_{mammals}, \ldots, \boldsymbol{e_m}, \boldsymbol{e_{m-1}}, \boldsymbol{e_{m-2}}, \ldots, \boldsymbol{e}_1\}$ the vectors with the letter *e* being some vectors from the standard basis **E**.

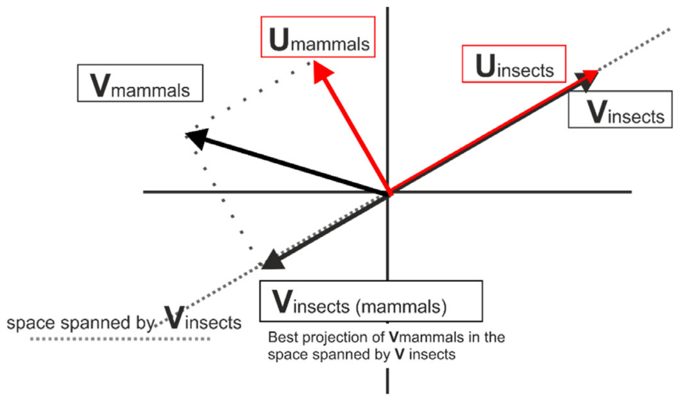

Fig. 1. The Gram-Schmidt method orthonormalizes an oblique basis in a sequential process in which next vector to be orthonormalized is projected onto the subspace generated by the sequence vectors that have already been orthonormalized. For example, the first vector in the sequence, the one that represents "insects" is taken and the second vector of the sequence, that for "mammals" is projected onto the subspace it generates. This projection is represented as $V_{\text{insects(mammals)}}$ and the orthogonal "mammals" vector is found. In this way, the basis constituted by insects and mammals is turned into an orthonormal basis. To expand this orthogonal basis with the next vector in the oblique basis, it is projected onto the subspace generated by "insects" and "mammals". Unlike what can be seen in this instance of a representation of two dimensions, in spaces with a high number of dimensions the movement experienced by vectors is not great, and allows for orthonormalization with no significant loss of meaning.

Step 2. The second step has to do with the orthogonalization of the vectors in basis **P**. It should be pointed out in advance that the new basis **P** is oblique, as it is very likely that the k vectors, despite being linearly independent (a *sine qua non* condition to be a basis) will not be orthonormal. Thus, this second step consists in re-orthonormaliznormalizing **P** by means of the Gram-Schmidt method, so that the new space with meanings retains the orthonormal properties of the original semantic space, and retains the same distances between terms. These properties are highlighted, for example, in the paper by Visinescu and Evangelopoulos (2014). This re-orthonormalization process is sequential (see Fig. 1). Nonetheless, forcing the vectors in **P** to be orthonormal distorts a small proportion of the meaning which they carry (the change experienced for it to be orthogonal to the previous vectors). However, if the fundamental contents of the rubric are not excessively semantically linked, it can be expected that the original meaning of those first k vectors will not be too distorted. Previous papers recommend that the correlation between a vector k and the same vector k orthogonalized, not be lower than 0.70 (see Olmos et al., 2014 for details on the re-orthonormalization stage).

Step 3. This is a change of basis - the final step. Once the basis **P** has been orthonormalized with the k vectors that represent the rubric, together with the m vectors in the standard basis, the new vector space with meaning (which is designated by the letter **N**) is achieved as the product of the new basis **P** on the original semantic space (which is usually named **U**). More specifically, the transformation of U into N is carried out by means of the formula:

$$\boldsymbol{N} = \boldsymbol{P}^{-1}\ \boldsymbol{U}^{\boldsymbol{T}} \quad (1)$$

Thus, the result of the application of these three steps described is the new semantic space **N**, in which the first k dimensions represent concepts in the rubric. Fig. 2 shows how D1, D2, D3, and D4 (shaded) are the dimensions that represent the concepts in the rubric {“insects”, “mammals”, …}. Any text projected onto **N**, e.g. a student’s answer, will be a vector whose four first scores identify and quantify the extent to which they contain the concepts represented in D1, D2, D3, and D4. Thus, generically, it can be considered as an *ad-hoc* expert system for relevant content detection.

This study seeks to solve two limitations found in the current Inbuilt-Rubric methodology shown above. In the next paragraphs, such two limitation will be presented with some detail and two tentative solutions will be presented.

3. Limitations of current Inbuilt-Rubric method and solutions

Two limitations can be identified in the details of the current Inbuilt-Rubric method. The first one has to do with what is described in the second case regarding Gram-Schmidt re-orthonormalization of the k vectors that constitute the **P** basis. As described, this step forces these vectors to be orthonormal, which means that the movement process aimed at achieving the independence of some vectors from others will generate idiosyncratic meanings that sacrifice the common meaning among the vectors (in the usual factor analysis language, this would be saying that the common variance among vectors is removed). In other words, given the orthonormality of some vectors with respect to others, the dimensions generated by them in the new space **N** only include what is specific to each concept, thus preventing them from reflecting topics that also belong to those concepts and are common to all. The loss of common variance due to re-orthogonormalization is transferred to the dimensions that remain latent (that is, the m abstract dimensions used to complete space **N**).

The second significant limitation to be overcome is that caused by the sequential nature of the re-orthonormalization process using Gram-Schmidt. As it is a sequential process, relocation of the first vector of the set of k vectors is not required to force it to be orthogonal. This first vector is the root of the process. Upon the subspace generated by that first vector will be projected the vector for the second concept, and the vector for that second concept, orthogonal to the first one, will be extracted (see Fig. 1). Thus, this second concept undergoes a slight movement with respect to the nonre-orthonormalized. same vector (that movement represents a proportional meaning change), as do the successive vectors for the concepts that are re-orthogonalized later. Even though the movement from the second vector-concept (and successive vector-concepts) can be averaged and controlled (ensuring that it does not cross a given threshold), this second vector-concept and the successive ones are penalized with respect to the first one, which remains unchanged, which makes the process somewhat arbitrary as the first vector-concept is the one that would be most reliable, again in psychometric terms, and the reliability of the successive vector-concepts is variable .

To overcome the first limitation, the proposal followed here is to introduce in the model an additional dimension that serves as a factor bringing together the common variance, like general factors in factor models. In the same way as for the dimensions for the $\boldsymbol{k}$ concepts, the goal can be to ensure that this general factor will also be orthogonal to all of them. In this way a configuration of the dimensions in which both the k concepts and the dimension of the general factor are orthogonal would be obtained. Along these lines, in recent years there have been many studies highlighting the advantages of using bifactor models (Reise, Moore, & Haviland, 2010). Bifactor models have been used to identify structures underlying the data. These models try to adjust data to a structure in which there are: (a) Certain factors called group factors that represent the specific variance of the observed indicators. These group factors are orthogonal and have zero correlation to each other. This means that they represent a specific variance that cannot be accounted for by a general factor. (b) A general factor that represents the common variance to all observed indicators but which is also orthogonal to the group factors, so that both sources of variance, the groups and the general factor, are clearly distinct and broken down. Bifactor models have been used, among

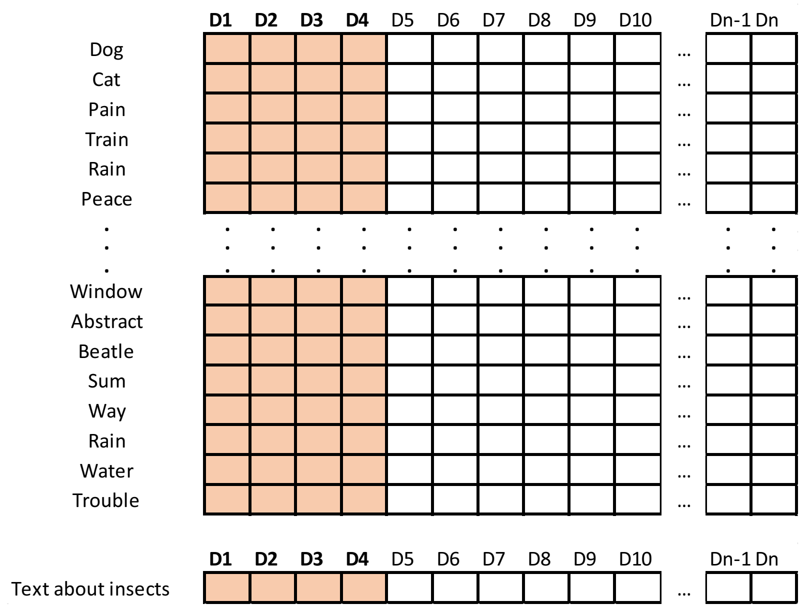

Fig. 2. New semantic space, after the first three steps described for the Inbuilt-Rubric method, in which the first k dimensions (shaded) represent concepts in the rubric. These concepts have in turn been built with descriptors (words) that represent each of them. For example, D1 represents the concept insects, and has been built with the sum of the vectors for the following descriptors: {"insects", "fly", "diptera", "bee", "ant"}. If a text on animals in which insects were part of the topic were to be projected via a folding-in, the vector for that text would have a high score in D1.

other things, to verify the adjustment to unidimensionality in TRI studies that are not strictly unidimensional (Reise, Morizot, & Hays, 2007), or to verify the existence of a structure that actually breaks unidimensionality in factor studies, taking the irrelevance of the groups and the substantive nature of the general factor as a null hypothesis (Reise et al., 2010). As was previously stated, there is a similarity between the bifactor model and Inbuilt-Rubric work as regards the orthogonal nature of the dimensions of concepts. The re-orthogonalization process means that the dimensions extracted in the new non-latent space **N** share no common variance and are idiosyncratic distillations of the meanings. However, the functional difference between a bifactor model and the Inbuilt-Rubric model is that in the latter the common variance from all concepts is not identified. Conversely, this common variance is forced to rest on the m abstract dimensions with which the basis is completed - dimensions that can be termed out-topics. Even though this procedure works well as a proxy or a sensor of the presence of concepts in the text, it is in a certain way a fallacy to believe that the common part is an out-topic. Thus, taking the bifactor model as a reference, a common factor that captures that common variance is to be generated. In this case, we firstly intend to impose that the first dimensions of the semantic space represent the meaning of a set of concepts established *a priori*; and, secondly, we intend to extract an additional dimension that represents a meaning common to all those concepts. Moreover, all dimensions with meanings and a general factor will remain orthogonal in the model, making that an LSA semantic space functionally behaves like a bifactor model and acquires its main advantages. This includes -which is of crucial interest in our study - the possibility of obtaining a general factor that is orthogonal to the other factors and, like a substrate, represents the common part of the concepts. So the common part ceases to be regarded as an out-topic. To do this, we must add to the k vectors that constitute the **P** basis an additional vector which represents the common variance of the concept-vectors. Further details are given below.

Regarding the second limitation observed, the sequential nature of the orthonormalization of the k vectors of the basis **P** to obtain the new semantic space **N**, a mechanism can be included to overcome this sequential nature - namely, performing the re-orthonormalization of the k vectors successively in a different order$\{(\boldsymbol{V}_{insects}, \boldsymbol{V}_{mammals}, \boldsymbol{V}_{birds}), (\boldsymbol{V}_{birds}, \boldsymbol{V}_{insects}, \boldsymbol{V}_{mammals}), \ldots, (\boldsymbol{V}_{insects}, \boldsymbol{V}_{birds}, \boldsymbol{V}_{mammals})$ so that the k final vectors that will constitute the new basis P are an average of each of the k vectors resulting from the successive re-orthonormalizations. This improvement process will be tested in this study, and the analysis of its results will be added to that already explained as a solution to the first limitation: functional homologation with the bifactor model.

4. Method

As stated, the goal of this study is to test the attempted solutions for the limitations described in the previous section: the lack of a common factor and the sequential nature of the re-orthogonalization process. To overcome the first limitation, in previous section we have proposed introducing a vector that represents the general factor of the other k vectors of the new basis **P**. And to solve the second limitation, we have proposed removing the sequential nature of the process by including the mechanisms described: the re-orthonormalization of the k vectors successive times in a different order. To test both solutions, three different procedures or versions of the Inbuilt-Rubric method will be examined: the first procedure (first version) is the original method, that is, as it was implemented at the start and which in this study will be regarded as the baseline (see the procedure). The second version of the method will try to solve the problem of sequentiality. To this end, it introduces the removal of sequentiality in re-orthonormalization as proposed. The third version, in addition to including the non-sequential re-orthonormalization mechanisms, adds one further dimension to the k dimensions described above, in order to serve as a common factor, that is, representing the common variance to all the concepts imposed on the other k vectors in the new basis. These will be the three procedures whose performance will be analyzed and compared on the basis of three evaluation tasks, comparing them with the evaluation conducted by humans, as described below.

4.1. Procedures for the generation of the Inbuilt-Rubric spaces

As stated, three procedures will be compared: a baseline, which is the sequential procedure or the first version (the one used in prior studies to date); two procedures as solutions to the two problems previously posed; sequentiality (or the second version) and the absence of a factor that brings together the variance common to the concepts (or the third version).

Sequential procedure (version 1): As explained, this is the procedure already described in previous papers. In it, the procedure for the re-orthonormalization of the basis depends on the order in which the experts generate the rubric. With this procedure, the first re-orthonormalized non-latent dimension has a reliability of 1 (the correlation between the first vector selected for the basis correlates 1 to the same vector in the re-orthonormalized basis). But when the following basis vectors are sequentially re-orthogonalized, reliability is no longer one, but rather each new re-orthonormalized vector is distorted. Even though in the procedure reported in the previous studies it is recommended that the correlation between the vectors before and after re-orthonormalization not be lower than 0.70, this distortion can penalise each vector in an increasing and sequential way.

Random sampling procedure (version 2): To solve the order-dependency of the first condition, the *random sampling* method generates *k!* different sequences with the basis *K* vectors, where *k* is the number of non-latent vectors for the new basis, so that the re-orthonormalization is performed taken all the possible orders of the *k* vectors into account. However, to prevent the random combinatorial process from being unmanageable, *k!* can only reach a maximum of 120. At this point different orders are randomly displayed). Finally, the final *k* vectors are the average (component by component) of each of the *k* vectors resulting from each of the successive re-orthonormalizations. This ensures that order-dependency is avoided and that in no case *k* vectors prevail over others.

Bifactor-based procedure (version 3): Once the *k* vectors that represent the rubric concepts and constitute the basis have been established, a factor analysis is conducted by means of the Principal Axis Method, in which the solution of one general factor is extracted from all those *k* vectors. Then the vector of scores for that general factor is included in the set of the *k* vectors of the *P* basis (*k* has one more vector, the last vector representing the general factor). The procedure to re-orthogonalize this basis is the same as that described for the *random sampling*, with the difference that the general factor is added to the procedure as one additional basis vector of *k*.

At this point, a warning should be made. As stated, in order to extract a general factor from the vectors from the set *k*, the Principal Axis Method was employed, which generated excessive co-linearity problems in the re-orthogonalization process, which led to dividing the descriptors from each concept (descriptors whose vector sum generate each *k* vector in the basis **B**) into two groups. For example, descriptors from *insects* {“fly”, “diptera”, “bee”, “ant”, “chitin”, butterfly”} are divided into {“fly”, “diptera”, “bee”} and {“ant”, “chitin”, butterfly”}. The descriptors in each group are added and two vectors are obtained. Following the example, we obtained $\boldsymbol{V}_{insects1} = \boldsymbol{V}_{fly} + \boldsymbol{V}_{diptera} + \boldsymbol{V}_{bee}$ and $\boldsymbol{V}_{insects2} = \boldsymbol{V}_{ant} + \boldsymbol{V}_{chitin} + \boldsymbol{V}_{butterfly}$

In this way, for each concept, rather than one single vector, two vectors will be extracted, one for each group of descriptors. For example, if a rubric with four concepts (each one with its descriptors) is proposed initially and each concept is represented by a vector, eight vectors are obtained, two per concept (for every group of descriptors for the same concept). Then the factor analysis by Principal Axis Method is performed using these eight vectors and the common factor is obtained. The vector that represents the general factor, together with the *k* vectors (4 in the example) will be the part of the new basis that represents the concepts (remember that it will be completed with part of the standard basis). To avoid co-linearity problems, other divisions could have been made, for example by extracting the common factor from all the descriptors independently of concepts groups. However, in this study we used a two-split strategy.

4.2. Comparisons between human judges and LSA procedures

The three procedures to be compared were tested in three real summary evaluation tasks in the academic field, as it is there that the Inbuilt-Rubric method and its results have been reported (Martínez-Huertas et al., 2018; Olmos et al., 2014; Olmos et al., 2016). In addition, it is a very frequent task in the literature when it comes to comparing different procedures within automatic evaluation systems (Dronen, Foltz, & Habermehl, 2014; Foltz, Laham, & Landauer, 1999; Foltz, Streeter, Lochbaum, & Landauer, 2013; Haley, Thomas, Petre, & De Roeck, 2007; Jorge-Botana, León, Olmos, & Escudero, 2010; Jorge-Botana, Luzón, Gómez-Veiga, & Martín-Cordero, 2015; Magliano & Graesser, 2012; Olmos, León, Jorge–Botana, & Escudero, 2009). Each of the three evaluation tasks comprised an activity in which a text had to be read and summarized. The topics of the texts to be read were Fauna and flora, Human communication, and Psychobiology. The texts were read and summarized by university students. The Flora and Fauna text was about 620 words long, while the Human Communication and Psychobiology text was about 1000 words long. All students were asked to write their summary on a text editor and upload it to the platform provided (they were expressly asked to write them on MS-Word as they would be easier to turn into plain text for later processing). The summaries had the following average word counts (the word count only includes the words represented in the semantic space): Flora and fauna 40.6 words, Human communication 58.8 words, and Psychobiology 70.6 words.

Two evaluations of each text were performed: one carried out by human judges and the other by LSA procedures. The first one was an assessment by three human judges for each rubric concept. The judges were instructed to give a score for the extent to which each rubric concept was present in the student’s text (the average per concept was calculated on the basis of the three judges’ scores). The human evaluation was regarded as the gold standard against which to compare the automatic evaluation by the LSA system (in its three procedures: sequential, random sample, and bifactor based). As regards LSA, a corpus based on a sample of articles taken from the online encyclopedia WIKIPEDIA was used, and, for every type of text to be summarized, a rubric and a number of descriptors for each concept in that rubric are created (see the example of the concept “insects” in the introduction). The corpus was processed using the standard LSA procedure (character cleanup, stop lists, lemmatization, identification of entities - animal kingdom becomes Animal Kingdom-, calculation of log-entropy and SVD), and it served to generate the original latent space of 300 dimensions, 39,566 represented terms, and 404,436 paragraphs (or documents). The changes corresponding to each of the three Inbuilt-Rubric procedures were performed on this original latent space. Thus, for each version or condition, the *k* vectors for the bases *b* were calculated using the sum of the descriptors for each concept, and in the case of the condition based on the bifactor, one more vector representing the general factor was also extracted. The LSA form of evaluation was as follows: each summary was projected (via folding-in) onto the semantic space generated for each Inbuilt-Rubric condition: that is, the space in which its first k dimensions represent the rubric concepts (as well as a general factor in the third condition for generation of the space based on the bifactor). In this way, the score for the projection of each summary in each one of those first k dimensions identified each of the summaries on the basis of the concepts of the rubric (including the general factor in the third condition). Both the LSA processing to create the original latent space and the Inbuilt-Rubric procedures required were performed using the Gallito Studio® application (Jorge-Botana, Olmos, & Barroso, 2013).

4.3. Data analysis

Three analyses were conducted. The first was a verification of validity. In order to obtain more validation evidence of the capacity of the computational methods to mimic human assessment, the first analysis consisted of three confirmatory factor analyses where the elements of the rubric assessed by humans and the computational methods were taken as observed indicators and the concept factors were inferred from them (i.e. concept factors as the latent variables). It should be noted that these factorial models were intended to test that each pair of elements assessing the same conceptual axis in the rubric (human and computational model pair) were an observed indicator for the same latent factor (i.e. the conceptual factor). That is, human scores and LSA scores assessing the same concept were expected to converge in a unique latent factor. Scores assessing different concepts for both LSA and human judges were also expected to be part of different latent factors. It should be pointed out that a random sample and a sequential sample were used as the computational Inbuilt-Rubric procedures, and the bifactor-based procedure was not used as it has a general factor and this causes an intrinsic noise in a factor solution (e.g. it would be similar to including an entire scale as well as the items of which it comprises in a factor analysis). As the second analysis, Pearson correlations between the human score (taken as the criterion) and the total score for each LSA procedure were established. The higher these correlations, the greater the validity of the criterion for the LSA procedure. The third analysis was intended to verify that the introduction of a general factor by means of the Bifactor-based procedure does not remove the capability of prediction of the quality of a student's text from the specific dimensions. In other words, introducing a general factor does not blur the metrics sensitive to the appearance of indicators for each specific concept. To this end, multiple regression models were created in which each human score for a concept was expected to be predicted by the LSA score for that concept, not for the LSA scores for the other concepts. This was done in the Random Sampling and Bifactor-Based procedures, for, as shown below, the Random Sampling procedure behaves better than the Sequential procedure, and the Bifactor-Based procedure introduces the general factor, the effect of which we want to verify. In the Bifactor-based case, it is expected that, when entering the general factor in the model, the LSA score in the concept to be predicted will remain significant, that is, the general factor will serve as an added value to increase prediction. For the regression the Step Wise method was used, in order to obtain only models created using statistically significant predictors. Moreover, for the Bifactor-based procedure, the regression model was forced to include the general factor as a predictor, in order precisely to verify its effect on the other predictors.

5. Results

5.1. Confirmatory factor analysis

In the human communication text, the sequential LSA procedure did not converge and thus this procedure did not mimic the human assessments well. However, the factor analysis for the

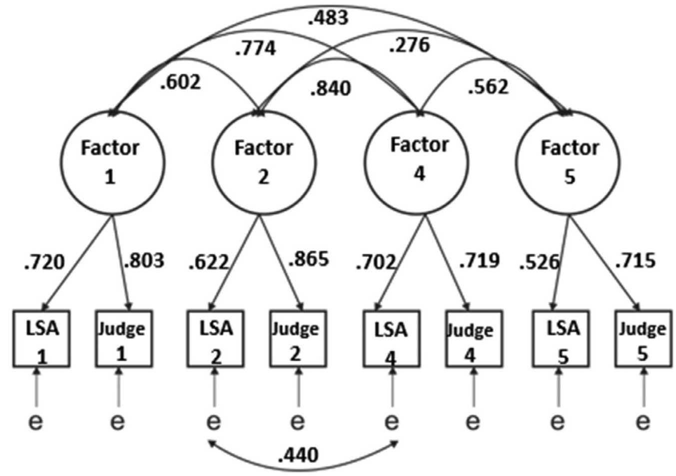

Fig. 3. Standardized factor solution for the Human communication text. LSA comprises the observed computational scores for each concept (yielded by the random sample method). Judge 1 means the average judges' score for concept 1. Judge 2 means the average judges' score for concept 2, and so on.

human communication text using the random sample computational method fitted well. It was necessary to remove the third conceptual axe because this latent factor correlated more than one with two other conceptual factors. This implies an excessive overlap in the content of this and other conceptual factors. After removing this conceptual factor, a fourth factor solution was adjusted (one residual correlation between two observed computational indicators was freed). The model fitted the data well: $N = 100$; $\chi^2(13) = 17.955$, $p = 0.159$, RMSEA = 0.062 90% CI 0.000–0.125; *CFI* = 0.979; *TLI* = 0.954; *SRMR* = 0.047. Fig. 3 shows the four-factor solution for the human communication text. The standardized factor solution shows positive correlations between the latent conceptual factors (range 0.276 and.840). The standardized factor loadings (all above 0.500) also show that the observed indicators are reliable measures for the concept factors.

Secondly, the factor solution for the flora and fauna text was performed. This time, the four conceptual factors were retained in the model. Both computational LSA procedures, the sequential and the random sample procedures, fitted equally well, but both factor models fitted the data more modestly than in the case of the previous text. As both LSA procedures fitted equally well, we provide Random Sample goodness of fit indices. CFI and SRMR showed good results, although TLI and RMSEA did not. The goodness of fit indices for the random sample method were: $N = 77$; $\chi^2(14) = 30.593$, $p = 0.006$, RMSEA = 0.124 90% CI 0.063–0.184; *CFI* = 0.944; *TLI* = 0.887; *SRMR* = 0.045. One modification index (10.474) suggested freeing a residual correlation between a computational observed indicator (LSA4) and a human Judge indicator (Judge 3), but this makes no theoretical sense. Fig. 4 shows the standardized four-factor solution. As in the previous Confirmatory Factor Analysis, the standardized factor loadings show highly reliable measures of the concept factors (range 0.624 to 0.942), and the factor correlations ranged between 0.237 and 0.851.

As regards the psychobiological text, both computational LSA procedures, the sequential and the random sample procedure, fitted equally well, but the Random Sample procedure fitted slightly better (all the goodness of fit indices were better in Random Sample procedure). However, in both LSA procedures one of the conceptual factors (the second concept) was removed due to the existence of a Heywood case (a correlation between the second and fourth latent factor above one). The standardized factor loading of an observed indicator for the second concept factor (the human judge indicator) was the only one that was less than 0.30. For the

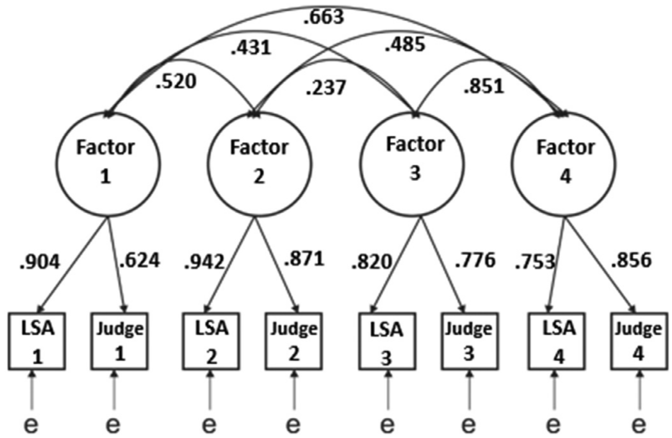

Fig. 4. Standardized factor solution for the Flora and fauna text (random sample method). LSA comprises the observed computational scores (yielded by the random sample method). Judge 1 means the average judges' score for concept 1. Judge 2 means the average judges' score for concept 2, and so on.

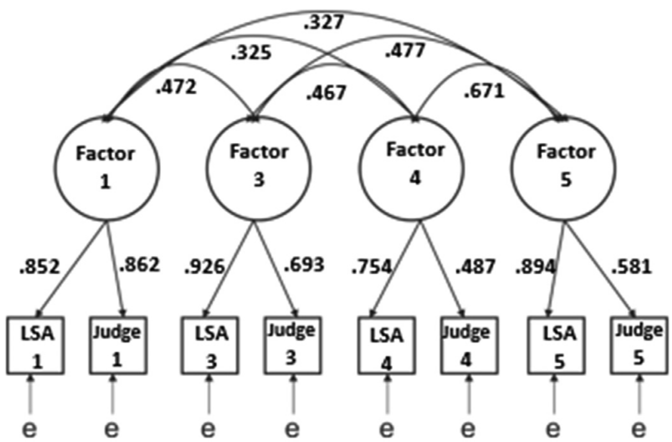

Fig. 5. Standardized factor solution for the Psychobiological text. LSA comprises the observed computational scores (yielded by the random sample method). Judge 1 means the total judges' score for concept 1. Judge 2 means the total judges' score for concept 2, and so on.

sake of parsimony, we only report and comment on indices from the Random Sample LSA procedure. When the confirmatory analysis was conducted on four of the concept factors, the model fitted the data well: $N = 100$; $\chi^2(14) = 21.458$, $p = 0.090$, RMSEA = 0.073 90% CI 0.000 – 0.131; $CFI = 0.969$; $TLI = 0.937$; $SRMR = 0.044$. Fig. 5 shows the four-factor solution. The standardized factor solution shows positive correlations between the concept factors (range 0.325 and.671). The factor loadings were high and significantly different from zero.

The confirmatory factor analysis shows that computational models to construct meaningful dimensions mimic human assessment well, but we have found that the Random Sample method yields purer and clearer factor solutions, which means better mimicking.

5.2. Pearson correlations analysis

Thirdly, correlations analysis helps us to see which LSA procedure has the higher criterion validity. The criterion used was the human total score for quality (regarded as the sum the scores for all concepts). Table 1 shows the Pearson correlations between the human expert scores for student summaries and the three proposed LSA procedures. The magnitude of the correlations is large (range between 0.618 and 0.810) and significant ($p < 0.001$). It is worth noting that the bifactor method provides the highest correlations within each of the texts, followed by the random sample method and the sequential method. Thus, it seems that the two novel methods capture a greater variance of the human quality scores than the sequential method.

5.3. Multiple regression

It is important to state that, on the basis of the results of the Confirmatory Factor Analyses, the concepts that did not preserve the LSA-Human factor structure were excluded from the regression analysis. In the case of the Human Communication text, concept 3 was excluded, and in the case of the Psychobiological text, concept 2 was excluded. Tables 2–4 show the results of the multiple regressions in six confusion matrices (two per LSA procedure for each text). Each row in each table shows of the judges' score for a concept is predicted by all the LSA dimensions. For example, in the first row of Table 2A (Human Communication and Random Sample LSA Procedure), the human score for concept 1 is predicted by dimension 1 and to a lesser extent by dimension 5. This is correct to the extent that the human score for concept 1 is predicted on the basis of LSA dimension 1. In general, the diagonal cells in the confusion matrix should have the highest values. The result shows how this is mostly the case. Each dimension representing a concept displays a tendency to predict the human score for that concept. Moreover, when introducing a general factor by means of the Bifactor-based procedure (Tables 2B, 3B; and 4B), this capacity is not blurred, but even seems to become clearer, yielding fewer deviations than in the Random sample method (Tables 2A, 3A, and 4A). This is important as the goal of this third analysis is precisely this: to verify and, if applicable, confirm that the introduction of a general factor does not detract from the predictive capability of specific dimensions.

6. Conclusions and future work

There are some very good examples of text categorisation and content detection systems with LSA (Thorleuchter, & Van den Poel, 2013; Wang, Peng, & Liu, 2015; Zhang et al., 2011) but all of them exploit the original space of the model, in which dimensions are meaningless or latent. Nonetheless, recent studies have proposed some procedures to exploit a transformed and meaningful space instead of the original one. This is the case of Inbuilt-Rubric methodology (Olmos et al., 2014; Olmos et al., 2016). In this study we have put to the test the solution for two limitations found in its current version. Basically, the current Inbuilt-Rubric procedure selects a new basis, alternative to the standard basis, to express the original latent space in another in which some vectors have the meanings of the concepts in a rubric. Then, after a re-orthonormalization process, the original standard basis is changed into that new basis. This generates a new space in which the first dimensions have the meanings of the concepts in that rubric. The two limitations found in that version are the following: The first one has to do with the current version in which, due to the re-orthonormalization process, the meaning that would represent a common variance for all the rubric concepts is sacrificed, with this common meaning being left as out-topic. The second limitation has to do with the sequentiality of the re-orthonormalization process, which penalizes some dimensions representing rubric concepts. As a solution to the first limitation, we propose extracting a new dimension from the rubric to represent a general factor to the concepts in the same rubric. This is done by obtaining the variance represented by the general factor in a factor analysis. As for the second limitation, sequentiality, we propose an improved procedure that randomly extracts different sequences to extract the order in which the vectors in the new basis are re-orthonormalized. The goal of this paper was to verify the degree of efficiency of both

Table 1
Correlations between the human expert score and the Inbuilt-Rubric methods.

	Sequential method	Random sample method	Bifactor-based method
Human communication text ($N = 100$)	.618**	.727**	.762**
Flora and fauna text ($N = 77$)	.777**	.783**	.810**
Psychobiological text ($N = 100$)	.717**	.724**	.779**

Note: significant values $^{**}P<0.001$, $^{*}P < 0.005$.

Table 2
Human communication text.

	A. Random sample method				B. Bifactor-based method				
	Dim 1	Dim 2	Dim 4	Dim 5	Dim 1	Dim 2	Dim 4	Dim 5	BF
Judge 1	.498**	–	–	–	.402**	–	–	–	.257*
Judge 2	.235**	.361**	.232**	–	–	.259**	–	–	.326**
Judge 4	.316**	–	.336**	–	–	–	.294*	–	.281*
Judge 5	–	–	–	.293**	–	–	–	.333**	–

Note: Blank cells are non-significant values. *Note:* significant values $^{**}P<0.001$, $^{*}P < 0.005$.

Table 3
Flora and fauna text.

	A. Random sample method				B. Bifactor-based method				
	Dim 1	Dim 2	Dim 3	Dim 4	Dim 1	Dim 2	Dim 3	Dim 4	BF
Judge 1	.569**	–	–	–	.381**	–	–	–	.483**
Judge 2	–	.820**	–	–	–	.869**	–	–	–
Judge 3	–	–	.647**	–	–	–	.730**	–	–
Judge 4	.248**	–	.327**	.351**	–	–	.335**	.177$^{\#}$	.375**

Note: Blank cells are non-significant values. *Note:* significant values $^{**}P<0.001$, $^{*}P < 0.005$, $^{\#}$marginally-significant.

Table 4
Psychobiological text.

	A. Random sample method				B. Bifactor-based method				
	Dim 1	Dim 3	Dim 4	Dim 5	Dim 1	Dim 3	Dim 4	Dim 5	BF
Judge 1	.692**	–	–	–	.529**	–	–	–	.285**
Judge 3	–	.591**	–	–	–	.446**	–	–	.369**
Judge 4	–	–	.267**	–	–	–	.359**	–	.258**
Judge 5	–	.246**	–	.473**	–	–	–	.484**	–

Note: Blank cells are non-significant values. *Note:* significant values $^{**}P<0.001$, $^{*}P < 0.005$.

solutions, and different strategies have been followed to this end. Firstly, confirmatory factor models were tested to check that the scores for each concept in the LSA and human rubrics could be included in latent factors. These confirmatory models also served to identify the rubric concepts that distorted the metrics of the LSA procedures. Secondly, the correlation coefficients between total LSA scores and total judges' scores were calculated. In addition, regression analyses were performed to verify whether each human score for a concept was predicted by its corresponding LSA dimension. These regression analyses also served to verify that the introduction of a general factor as a new rubric concept would not decrease the sensitivity of the LSA metrics for each concept.

In general, the results have shown that the solutions proposed to overcome the limitations of the current Inbuilt-Rubric method, namely, the sequential nature of the re-orthonormalization process and the lack of a measure of the common variance among concepts, provide better results than the original version of Inbuilt-Rubric. Examining each procedure separately, it can be said that the Random Sample procedure is useful to find, by means of a Confirmatory Factor Analysis, a Human-LSA factor structure, thus providing the LSA metrics with construct validity. This procedure also serves to discard any concepts that do not fit the factor structure (e.g. detecting convergence problems or Heywood cases, such as correlations above 1 or negative variance). Both issues are important in a potential quality control, and deficiencies can be corrected by monitoring scores in this way (Bejar, 2011). However, the procedure that includes both solutions - the Bifactor-based procedures - is the one that displayed the best performance when assessing global text quality, and, as shown by the data, their sensitivity to detect specific concepts is not decreased when a general factor is introduced. In fact, sensitivity seems even to improve. Thus, it should be highlighted that introducing the general factor as another rubric concept makes it possible to include and evaluate the content portion which the original version of the Inbuilt-Rubric method wasted under the term "out-topic"; as well as doing so without having a negative impact on the sensitivity of the other concepts (probably even improving them). We believe that this study constitutes a bridge between classic use of LSA, Factor Analysis models, and the processes for the refinement of evaluation instruments within psychometric theory.

In terms of its direct applicability to real context of academic assessment, it is important to say that the current version of Inbuilt-Rubric (Olmos et al., 2014; Olmos et al., 2016) has proven to be very efficient in assessment on real platforms where thousands of exercises on many topics written by students in learning centers at many levels are evaluated (see the performance of G-Rubric www.grubric.com in Santamaría et al., 2017). These platforms are in general relevant content detection expert systems which evaluate the presence/absence of a topic in texts (sometimes called Topic Labeling). The results of this study provide a better control of the entire Inbuilt-Rubric process, which will provide better sensitivity to these systems to detect not only specific contents relevant to the system task but also to provide overall quality measurement of texts. These improvements pave the way for several real applications, such as virtual assistants, expert systems to control the Peer Correction process, automatic feedback providers, academic recommendation systems, Computer-Based Examination for massive assessment, etc.

Some remarks can be made regarding use of the Inbuilt-Rubric technique as a generic classification tool. Inbuilt-Rubric is a very versatile and simple technique in environments where there is no possibility of supervised training (where no training labeled data are available), for example, supervised neural networks. Since categories are imposed and different models can be tested, an early content detector system can be deployed. In addition, Inbuilt-Rubric meaningful dimensions can even be also used in future studies in the feature extraction phase in further supervised training, as it has been observed that resuming the textual information in a vector with topic components is a better procedure than "bag of words" pre-processing (Mandl, 1999). Inbuilt-Rubric is also very versatile where the topics of the classification are continuously changing. Imagine for example an expert system classifying news items into weekly topics, or an enterprise with temporary names and formats forcommercial offers. In such scenarios, labeling a sample *by hand* to perform a supervised training every time topics change is not possible. It will require quick changes in the classification concepts and easy ways to deploy different models. Another configuration of Inbuilt-Rubric for future research could involve introducinga summarization phase before the Inbuilt-Rubric classification. In other words, an extractive expert system that proposes the most relevant sentences or words from a text in order to prepare the classification (Kireyev, 2008; Ozsoy et al., 2011). In turn, this extractive process could also be filtered toward the Inbuilt-Rubric meaningful dimensions so that the words or sentences of the summarization would have to be connected to the classification concepts of the system. That would attune the process to the extraction of the most important meanings to be classified. In any case, the reported improvements can be very useful where an expert system uses Latent Semantic Analysis and Inbuilt-Rubric in such scenarios.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

Balbi, S., & Esposito, V. (1998). Comparing advertising campaigns by means of textual data analysis with external information. In S. Mellet (Ed.), *Actes des 4es journées internationales d'Analyse statistique des données textuelles* (editor). UPRESA Nice: 39-47.

Bejar, I. I. (2011). A validity-based approach to quality control and assurance of automated scoring. *Assessment in Education, 18*(3), 319–341.

Bellegarda, J. R. (2000). Exploiting latent semantic information in statistical language modeling. *Proceedings of the IEEE, 88*(8), 1279–1296.

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review, 124*(1), 1.

Chu-Carroll, J., & Carpenter, B. (1999). Vector-based natural language call routing. *Computational Linguistics, 25*(3), 361–388.

Contreras Kallens, P., & Dale, R. (2018). Exploratory mapping of theoretical landscapes through word use in abstracts. *Scientometrics, 116*, 1641. https://doi.org/10.1007/s11192-018-2811-x.

Denhière, G., & Lemaire, B. (2004). A computational model of children's semantic memory. In K. Forbus, & D. Gentner y T. Regier (Eds.), *Proceedings of the 26th annual meeting of the cognitive science society* (pp. 297–302).

Dronen, N., Foltz, P. W., & Habermehl, K. (2014). Effective sampling for large scale automated writing evaluation systems. arXiv preprint arXiv:1412.5659

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology., 38*, 188–230. doi:10.1002/aris.1440380105.

Evangelopoulos, N. E. (2013). Latent semantic analysis. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*, 683–692.

Evangelopoulos, N., & Visinescu, L. (2012). Text-mining the voice of the people. *Communications of the ACM, 55*(2), 62–69.

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems, 21*(1), 70–86.

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). The intelligent essay assessor: Applications to educational technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 1*(2). http://imej.wfu.edu/articles/1999/2/04/index.asp.

Foltz, P., Streeter, L., Lochbaum, K., & Landauer, T. (2013). Implementation and applications of the intelligent essay assesor. In M. D. Shermis, & J. Burstein (Eds.), *Handbook of automated essay evaluation*. NY: Routledge.

Günther, F., Dudschig, C., & Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *The Quarterly Journal of Experimental Psychology, 69*(4), 626–653.

Haley, D. T., Thomas, P., Petre, P., & De Roeck, A. (2007). Seeing the whole picture: Comparing computer assisted assessment systems using LSA-based systems as an example. Technical Report Number 2007/07. Open University.

Hu, X., Cai, Z., Wiemer-Hastings, P., Graesser, A. C., & McNamara, D. (2007). Strengths, limitations, and extensions of LSA. In T. K. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *The handbook of latent semantic analysis* (pp. 401–426). Mahwah, NJ: Erlbaum.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. (2006). Models of high--dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica, 121*(1), 65–80.

Jorge-Botana, G., León, J. A., Olmos, R., & Escudero, I. (2010). Latent semantic analysis parameters for essay evaluation using small-scale corpora. *Journal of Quantitative Linguistics, 17*(1), 1–29.

Jorge-Botana, G., Luzón, J. M., Gómez-Veiga, I., & Martín-Cordero, J. I. (2015). Automated LSA assessment of summaries in distance education: some variables to be considered. *Journal of Educational Computing Research, 52*(3), 341–364.

Jorge-Botana, G., Olmos, R., & Barroso, A. (2013). Gallito 2.0: A natural language processing tool to support research on discourse. In *Proceedings of the 13th annual meeting of the society for text and discourse*.

Jorge-Botana, G., Olmos, R., & Barroso, A. (2012). The construction-integration framework: A means to disminish bias in LSA-based call routing. *International Journal of Speech Technology., 15*(2), 151–164.

Jorge-Botana, G., Olmos, R., & Luzón, J.M. (submitted). Bridging the theoretical gap between lexical representation models without the pressure of a ranking: Some lessons learnt from LSA. Cognitive Proccessing.

Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge University Press.

Kintsch, W., & Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science, 3*(2), 346–370.

Kireyev, K. (2008). Using latent semantic analysis for extractive summarization. In *Proceedings of text analysis conference, TAC 2008* November 17-19.

Kundu, A., Jain, V., Kumar, S., & Chandra, C. (2015). A journey from normative to behavioral operations in supply chain management: A review using latent semantic analysis. *Expert Systems with Applications, 42*(2), 796–809.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284.

Landauer, T. K., Kireyev, K., & Panaccione, C. (2011). Word maturity: a new metric for word knowledge. *Scientific Studies of Reading, 15*(1), 92–108. doi:10.1080/10888438.2011.536130.

Lim, B. P., Ma, B., & Li, H. (2005). Using semantic context to improve voice keyword mining. In *Proceedings of the international conference on Chinese computing (ICCC 2005* 21–23 March.

Marcolin, C., Becker, J. L., Wild, F., & Schiavi, G. (2018). Business analytics in tourism: Uncovering knowledge from crowds. *15th international conference on information systems and technology management (CONTECSI'18)At*.

Magliano, J. P., & Graesser, A. C. (2012). Computer-based assessment of student constructed responses. *Behavior Research Methods, 44*(3), 608–621.

Mandl, T. (1999). Efficient preprocessing for information retrieval with neural networks. In *Presented at the 7th European congress on intelligent techniques and soft computing* (pp. 13–16).

Martínez-Huertas, J. A., Jastrzebska, O., Mencu, A., Moraleda, J., Olmos, R., & León, J. A. (2018). Analyzing two automatic assessment LSA methods (Inbuilt-Rubric vs. Golden Summary) in summaries extracted from expository texts. *Psicología Educativa, 24*(2), 85–92.

Olmos, R., León, J., Jorge–Botana, G., & Escudero, I. (2009). New algorithms assessing short summaries in expository texts using latent semantic analysis. *Behavior Research Methods, Instruments and Computers, 41*(3), 944–950.

Olmos, R., Jorge-Botana, G., León, J. A., & Escudero, I. (2014). Transforming selected concepts into dimensions in latent semantic analysis. *Discourse Processes, 51*(5–6), 494–510.

Olmos, R., Jorge-Botana, G., Luzón, J. M., Martín-Cordero, J. I., & León, J. A. (2016). Transforming LSA space dimensions into a rubric for an automatic assessment and feedback system. *Information Processing and Management, 52*(3), 359–373.

Ozsoy, M. G., Alpaslan, F. N., & Cicekli, I. (2011). Text summarization using Latent Semantic Analysis. *Journal of Information Science, 37*, 405–417.

Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university related FAQs. *International conference on advances in computing, communications and informatics (ICACCI*. Udupi: ICACCI.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559. http://doi.org/10.1080/00223891.2010.496477.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31.

Santamaría, M., Hernández, M., Sánchez-Elvira, Luzón, J. M., & Jorge-Botana, G. (2017). Using semantic technologies for giving a formative assessment and supporting scoring in large courses and MOOCs: First experiences at UNED (2015-2017). *The online, open and flexible higher education conference*.

Serafín, R., & Di´Eugenio, B. (2004). FLSA: Extending latent semantic analysis with features for dialogue act classification. In *Proceedings of ACL04, 42nd annual meeting of the association for computational linguistics*.

Shi, Y. (2008). *An investigation of linguistic information for speech recognition error detection*. Baltimore County, Baltimore: University of Maryland.

Steyvers, M., & Griffiths, T. (2006). Probabilistic topic models. In T. Landauer, . D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Latent semantic analysis: A. Road to meaning*. Laurence Erlbaum.

Thomas, N. T. (2016). An e-business chatbot using AIML and LSA. In *2016 International conference on advances in computing, communications and informatics (ICACCI)* (pp. 2740–2742). doi:10.1109/ICACCI.2016.7732476.

Thorleuchter, D., & Van den Poel, D. (2013). Technology classification with latent semantic indexing. *Expert Systems with Applications, 40*(5), 1786–1795.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38*(5), 406–427. http://dx.doi.org/10.1037/h0069792.

Tonta, Y., & Darvish, H. R. (2010). Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics, 4*(2), 166–174.

Visinescu, L., & Evangelopoulos, N. (2014). Orthogonal rotations in latent semantic analysis: An empirical study. *Decision Support Systems, 62*, 131–143.

Wang, J., Peng, J., & Liu, O. (2015). A classification approach for less popular webpages based on latent semantic analysis and rough set model. *Expert Systems with Applications, 42*(1), 642–648.

Yeari, M., & van den Broek, P. (2015). The role of textual semantic constraints in knowledge-based inference generation during reading comprehension: A computational approach. *Memory, 23*, 1193–1214. doi:10.1080/09658211.2014.968169.

Yeari, M., & van den Broek, P. (2016). A computational modeling of semantic knowledge in reading comprehension: Integrating the landscape model with latent semantic analysis. *Behavior Research Methods, 48*(3), 880–896.

Zhang, W., Yoshida, T., & Tang, X. J. (2011). A comparative study of TF*IDF, LSI and multi-words for text classification. *Expert Systems with Applications, 38*(3), 2758–2765.