Chapter 6
A Geometric Algebra Based Distributional Model to Encode Sentences Semantics

Agnese Augello, Manuel Gentile, Giovanni Pilato and Giorgio Vassallo

Abstract Word space models are used to encode the semantics of natural language elements by means of high dimensional vectors [23]. Latent Semantic Analysis (LSA) methodology [15] is well known and widely used for its generalization properties. Despite of its good performance in several applications, the model induced by LSA ignores dynamic changes in sentences meaning that depend on the order of the words, because it is based on a *bag of words* analysis. In this chapter we present a technique that exploits LSA-based semantic spaces and geometric algebra in order to obtain a sub-symbolic encoding of sentences taking into account the words sequence in the sentence.

Keywords Semantic spaces · Sentences encoding · Clifford algebra

1 Introduction

Two rather orthogonal theories in Natural Language Processing are the symbolic [11] and distributional [25] paradigms: the former is compositional but only qualitative, the latter is non-compositional but quantitative [9].

A. Augello (✉) · G. Pilato
ICAR, CNR V.le delle Scienze - Ed.11, 90128 Palermo, Italy
e-mail: augello@pa.icar.cnr.it

G. Pilato
e-mail: pilato@pa.icar.cnr.it

M. Gentile
ITD, CNR Via Ugo La Malfa, 153, 90146 Palermo, Italy
e-mail: gentile@pa.itd.cnr.it

G. Vassallo
DICGIM Università di Palermo, V.le delle Scienze, Ed. 6, 90128 Palermo, Italy
e-mail: gvassallo@unipa.it

C. Lai et al. (eds.), *Distributed Systems and Applications of Information Filtering and Retrieval*, Studies in Computational Intelligence 515,
DOI: 10.1007/978-3-642-40621-8_6,

Distributional approaches try to quantify and categorize semantic correspondences between linguistic entities. The key idea is the *distributional hypothesis*, which states that words having similar meanings will occur in similar contexts [21]. This means that there is a correlation between distributional and meaning similarity, that makes it possible to estimate the latter starting from the former.

Algorithms that try to acquire distributional meaning can be divided in two categories: the first one includes all approaches that try to build distributional profiles for words based on which other words surround them, while the other one embraces the techniques that build distributional profiles based on in which text regions word occur [23].

The core of the distributional approach is that linguistic meaning is essentially differential, i.e. differences of meaning are mediated by differences of distributions, therefore the distributional methodology deals only with meaning differences or semantic similarity.

Usually the model that captures the pattern of distribution of single words across a set of contexts is a vector and the assessment of these models is often done by exploiting relations of semantic similarity between individual words.

Saussure gave the foundation of what developed later as structuralism; in a language signs are identified by their relation of difference; he emphasized that meaning arises from the differences between signifiers; these differences are of two kinds: syntagmatic and paradigmatic. The former deals with positioning and relate entities that co-occur in the text, the latter ones deal with substitution and relate entities that do not occur in the text.

According to Sahlgren [22], “A distributional model accumulated from co-occurrence information contains syntagmatic relations between words, while a distributional model accumulated from information about shared neighbors contains paradigmatic relations between words”.

Syntagmatic models collect text data in a words/documents co-occurrence matrix whose generic item is a function of the frequency of occurrence of a word in a document, while paradigmatic models collect text data in a words/words co-occurrence matrix whose generic item is a function of how many times words occur together within a context window.

In paradigmatic models the row and column vectors are different since row vectors model words appearing to the right of the other words, and the column vectors model words appearing to the left of the other words. The generated matrix is asymmetrical and is usually referred as “directional co-occurrence matrix”.

There are different techniques that exploit syntagmatic or paradigmatic models. Each of them exploits the fact that natural language elements, such as words, sentences, documents, are sub-symbolically represented as points in a high dimensional vector space, allowing the use of linear algebra in order to obtain pair-wise similarity scores. Such a space is usually named “semantic space”.

The peculiarity of semantic spaces is that this kind of structures are automatically induced by means of statistical analysis of large text corpora, usually without using any “a priori” knowledge.

One of the most used approaches for semantic space building is given by the Latent Semantic Analysis (LSA) paradigm. In particular LSA is based on a dimension optimization of the created space which highlights the latent indirect similarity relations among words and documents [15]. LSA, starting from a word-by-documents co-occurrence matrix, implements a syntagmatic use of contexts, and exploits the Truncated Singular Value Decomposition (TSVD) which approximates a paradigmatic use of contexts.

Vector-based models typically represent single words and do not take into account the grammatical structure of a sentence [14]. Therefore these models have a limited capability to model compositional operations over phrases and sentences.

In order to overcome these shortcomings, distributional methods have been lately extended in order to take into account also compositionality: these enhanced approaches have been named in literature of "distributional compositional semantics (DCS)" approaches.

Existing models are still arguable and provide general algebraic operators over lexical vectors. An overview of these methodologies which explains the benefits and limitations of different approaches about compositionality in distributional semantic models present in literature, including additive, multiplicative, mixture, tensor-based, and Structured Vector Space (SVS) models is given in [12].

Among the different approaches we recall here the work presented in [9], where a mathematical framework, based on the algebra of Pregroups, for a unification of the distributional theory of meaning using vector space models, and a compositional theory for grammatical types, has been introduced. The framework makes it possible to evaluate the meaning of a well-typed sentence from the meanings of its constituents.

Moreover in [6] a methodology based on Random Indexing and vector permutations has been proposed to encode several syntactic contexts in a single semantic space where a set of operations is defined. The technique exploits syntactic dependencies to perform some particular queries, such as the one for retrieving all similar objects of a verb, and it has been tested for semantic composition of short sentences and evaluated by using the GEMS 2011 dataset [13]. Finally, a distributional compositional semantic model based on space projection guided by syntagmatically related lexical pairs has been illustrated in [2]. Syntactic bi-grams are projected in a Support Subspace, in order to let arise the semantic features shared by the compound words and catch phrase-specific characteristics of the associated lexical meanings. The methodology relies on first selecting the most important components for a specific word pair in a relation and then modeling their similarity. This captures their meanings locally relevant to the specific context evoked by the pair. The approach is very effective for the syntactic structures of VO, NN and AdjN.

Recently we have proposed a sub-symbolic methodology for natural language sentences coding, exploiting Geometric Algebra (GA) rotation operators, named rotors [4, 20]. At a lexical-unit level the semantic coding is given by the vectors of an LSA space. At a words-pair level we associate to each bigram in a sentence an ad-hoc GA rotor. Finally at a sentence-level the whole coding is obtained by means of successive rotations of a standard basis in the semantic space, where each rotation is performed applying the rotor associated to the analyzed sentence bigram to the basis. Since this operation is non-commutative, word order is taken into account for the whole sentence encoding.

The approach has been here evaluated under the light of Compositional Distributional Semantic Models, and its performances have been evaluated by using the GEMS 2011 shared evaluation.

2 Semantic Rotors to Encode Sentences Semantics

The proposed methodology consists in an unsupervised procedure that injects information about the sentence structure and the semantics of its component words into a sub symbolic sentence coding.

The methodology is based on the following steps (see Fig. 1): the construction of a semantic space in order to extract a vector encoding of words belonging to a text corpus; the association of ad-hoc rotors to the sentence bigrams, and finally the coding of the sentence through the application of rotation operators (rotors) to a standard basis in the semantic space. Each rotation operator is dependent on the vector coding of the words composing the bigrams of the sentence. The rotation operator corresponds to a non-commutative operation represented by the clifford geometric product [17].

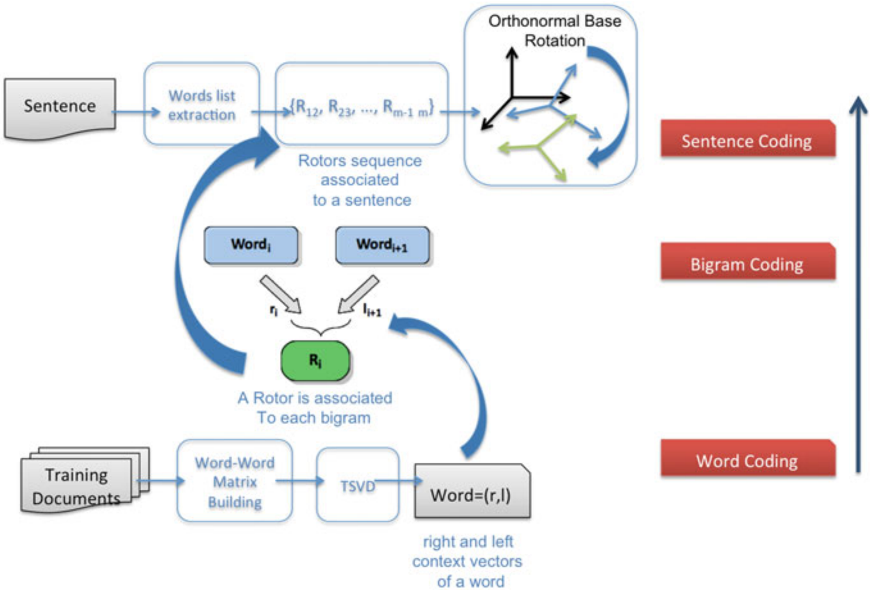


Fig. 1 Sentences encoding process

2.1 Words Coding in a Semantic Space

The first step aims at obtaining a semantic coding of the words composing a sentence. In particular a semantic space is obtained by means of Latent Semantic Analysis [16], a well established technique used to obtain a semantic representation of words. The strength of LSA is an induction-dimension optimization obtained through the truncated singular value decomposition (TSVD) that converts the initial representation of information into a condensed representation that captures indirect, higher-order associations between words [15]. In particular we consider the building of a word-by-word co-occurrences matrix, where its *(i, j)-th* entry of the matrix represents the number of times a bigram composed of the *i-th* word followed by the *j-th* word appears in a documents corpus inside a window of a fixed number of words.

An important characteristic is that the dimension of the matrix is determined only by the number of words included in the vocabulary and it is independent of the number of documents.

The resulting matrix, which is not symmetrical, is preprocessed substituting each entry of the matrix with the correspondent pointwise mutual information value .

The Pointwise Mutual Information (PMI) [8] between two words w_i and w_j is a co-occurrence metric, which allows to consider how likely it is to nd w_j in a document if that document contains w_i. The PMI normalizes the probability of co-occurrence of the two words with their individual probabilities of co-occurrence [7].

The PMI between w_i and w_j has been calculated as:

$$pmi = log_2 \frac{f(w_i, w_j)}{f(w_i) f(w_j)} \tag{1}$$

where $f(w_i, w_j)$ is the number of times that the ordered bigram $w_i - w_j$ occurs in the documents corpus considering a fixed size words window; $f(w_i)$ is the number of times that word w_i occurs in corpus; and similarly, $f(w_j)$ is the number of times that word w_j occurs in corpus. The weighted matrix is decomposed by means of truncated SVD, and the result is the following:

$$\mathbf{A} \approx \mathbf{A}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^T \ . \tag{2}$$

where $\mathbf{U}_k$, Σ_k and $\mathbf{V}_k$ are matrices that provide compressed information about the left and the right context of the word. In particular the *i-th* row of $\mathbf{U}_k$, multiplied by the square root of the Σ_{ii} element of Σ_k represents the *right context* of the *i*-th word, while the *i*-th row of $\mathbf{V}_k$, multiplied by the square root of the Σ_{ii} element of Σ_k represents the *left context* of the *i-th* word .

Therefore it is possible to associate to each word two different vectors in the generated semantic space: $\mathbf{l}_i$ and $\mathbf{r}_i$, the former representing the left context and the latter representing the right context of the word.

2.2 Bigrams Coding

A geometric algebra operator is associated to each bigram of a sentence. Given a bigram composed by the words w_i and w_j, let $\mathbf{l}_i$ and $\mathbf{r}_i$ be the left and right contexts of the word w_i and $\mathbf{l}_j$ and $\mathbf{r}_j$ the left and right contexts of the word w_j, a rotor represented as the following geometric product:

$$R_{ij} = \mathbf{r}_i \mathbf{l}_j = \mathbf{r}_i \cdot \mathbf{l}_j + \mathbf{r}_i \wedge \mathbf{l}_j \tag{3}$$

is associated to the bigram.

The geometric product is the combination of the classical dot product with the outer product ($\wedge$) and for this reason it is, in general, not commutative.

2.3 Sentence Encoding

The sentence encoding is obtained starting from a neutral, starting coding and applying, time to time, a non commutative operator dependent on the considered bigram. The starting coding is given by an orthonormal base of the semantic space: the canonical basis of k dimensions represented by the identity matrix. We call this starting coding $\mathbf{s}_0$, to consider it as the coding of an empty sentence.

The temporal sequence of words belonging to the sentence generates a rotation trajectory of an orthogonal basis in a semantic space. For a phrase of M words, and therefore of M-1 bigrams, we can associate M-1 rotors to the sentence, as Fig. 2 shows.

The sequence of these rotors will be applied to the original basis, transforming it M-1 times.

In particular we can analyze what happens in a specific step of the coding process.

Let $\mathbf{s}_{z-1}$ be the coding of the sentence after an analysis of $(z-1)$ bigrams. Let $z-th$ be the bigram composed of the words w_i and w_j.

We can associate to this bigram a rotor given by the geometric product between the right context $\mathbf{r}_i$ of the word w_i and the left context $\mathbf{l}_j$ of the word w_j.

Therefore we can perform a rotation of $\mathbf{s}_{z-1}$ in the $\mathbf{r}_i \wedge \mathbf{l}_j$ plane, obtaining the coding of the sentence at the $z-th$ step.

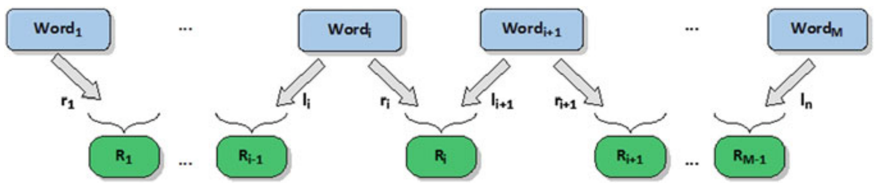
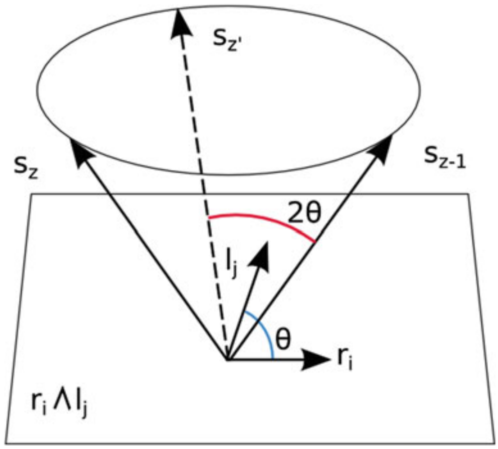


Fig. 2 Sequence of rotors associated to a sentence

Fig. 3 Coding of $\mathbf{s}_{z-1}$: double reflection in the plane specified by the rotor associated to the z_{th} bigram



The rotation [17] is performed making two subsequent reflections of $\mathbf{s}_{z-1}$ with respect to $\mathbf{r}_i$ and $\mathbf{l}_j$ vectors (see Fig. 3).

This operation is expressed in terms og geometric algebra by:

$$\mathbf{s}_z \mapsto R\ \mathbf{s}_{z-1}\ \tilde{R} = e^{(-\hat{B}\theta)}\ \mathbf{s}_{z-1}\ e^{(\hat{B}\theta)}\ . \tag{4}$$

where $\tilde{R}$ is the inverse of the R rotor [3], the unit bivector $\hat{B}$ represents the plane of rotation and the angle of rotation is 2θ. It is easy to demonstrate how the rotation operation is not commutative unless the rotation planes are completely orthogonal [24].

Each time a new bigram composing the sentence is analyzed, a new, intermediate encoding of the sentence that takes into account the sequence of the considered words is obtained. At the end of the procedure, the rotated basis can be represented by a vector of k^2 components, where k is the value chosen to truncate the SVD. The final coding is given by the orthogonal part of this vector with respect to the original basis. This allows to obtain a coding which is independent of the sentence length. It is important to point out that cyclical coding should not appear if the dimension of the semantic space is higher than the number of rotations associated to the sentence.

According to the non-commutative property of the rotation operation, given a list of rotors $\{R_1, R_2, ..., R_n\}$ corresponding to the bigrams in the sentence, the application of these rotation operations to the orthonormal base creates a coding that is function of the order in which these rotations are applied.

The final coding represents a synthesis of the word sequence history within the sentence and corresponds to its sub-symbolic coding.

3 Effects of the Rotation in Distributional Models Based on LSA

In a previous work we have highlighted how the semantic space obtained by means of LSA can be interpreted as a "conceptual" space. The axes of this space induced by the truncated SVD can be considered as the latent primitive concepts belonging to the training corpus and can be tagged with a set of words characterizing it [1].

The process of sentence encoding based on rotor operators allows to highlight "conceptual" relations that can arise between the primitive concepts as the bigrams composing the phrase are analysed.

We can make an analogy between the proposed model and a state transition system.

The rotated basis represents the "conceptual state" of the sentence. In particular the matrix associated to the rotated basis can be considered as the incidence matrix of a graph of connections among the "primitive concepts" of the space, where each node of the graph corresponds to a specific conceptual axis.

Let us suppose we have sentence s of M words. Let mq_0 the starting state of our coding process represented by the orthonormal unitary matrix, of dimension equal to the value chosen to the truncation parameter of the SVD.

Each time we consider a bigram of the sentence, we have a state transition function f given by the rotation operation, which brings to a new state that is the rotated basis represented by a orthonormal non unitary matrix mq. The result of each rotation leads to the induction of relations between the axes, represented as the connections between the nodes of the graph associated to state, as shown in Fig. 4. In fact, as shown in previous experimental results [4], the generic cell $mq_{i,j}$ of the matrix can be considered as representative of the relation between the *i-th* and the *j-th* conceptual axes. After the analysis of all the $M-1$ bigrams of the phrase the system reaches a final state F, representing the sentence encoding.

4 Experimental Results

This section reports some experimental results aimed at evaluating both the performance and the scalability of the proposed algorithm. The experimental phase has been performed according to the instructions of the GEMS 2011 shared evaluation [13].

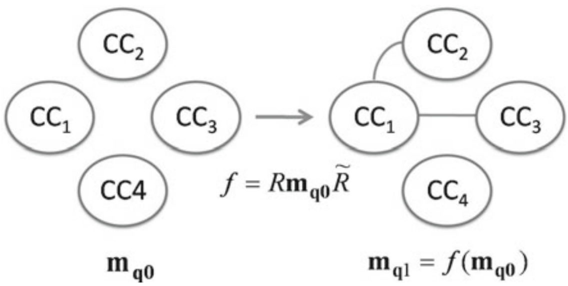


Fig. 4 Sentence encoding as a state transition process

In particular we used ukWaC [5] and TASA (see Acknowledgments) as source corpora for semantic spaces building. The test set, taken by GEMS 2011, consists in a list of two pairs of the following types: adjective-noun (AN), verb-object (VO) and compound nouns (NN), defined by Mitchell and Lapata [18, 19].

To each pair is associated a set of rates, ranging from 1 to 7, given by participants of a psycholinguistic experiment conducted by Mitchell and Lapata. For example the pair "result achieve"—"level reach" has a rate of 7, while "bus company"—"intelligence service" has a rate of 1. The total number of rates is 5833.

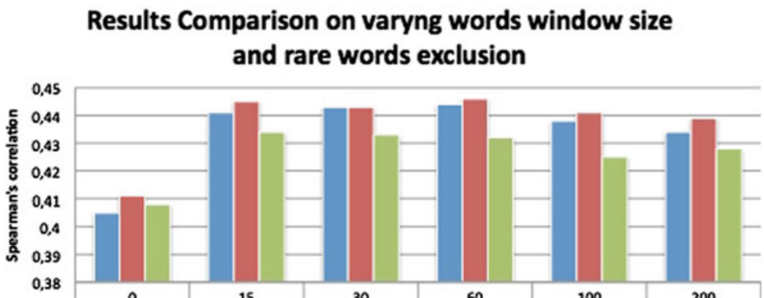
The system has been evaluated computing the scores obtained by the proposed algorithm for all of the adjective-noun combinations, verb-object combinations and compound nouns and therefore by calculating the Spearman correlation ρ between the obtained scores and all of Mitchell and Lapata's participant rates.

We have used documents of ukWack as source corpora to build the words co-occurrences matrix, where the elements of the matrix are weighted by means of the *pmi* score. The truncated SVD is performed with a factor $k = 100$.

We have considered a smaller subset of the ukWaC documents corpus, in order to quickly analyze the algorithm performance changes according to different values of the parameters. In particular the number of documents we have considered is 110165.

We have carried out different experiments changing parameters such as the window size, by taking into account or not the POS tags of the words, and removing words occurring in the corpus less than a given threshold. The results, shown in Fig. 5, show that the best results (evaluated over the all groups) are obtained by setting the words window equal to ± 4 and removing those words having a frequency lower than 60.

We have analysed the performance of the algorithm, by fixing one of the two parameters with the best value and changing the other. The following figures show



	0	15	30	60	100	200
3	0,405	0,441	0,443	0,444	0,438	0,434
4	0,411	0,445	0,443	0,446	0,441	0,439
7	0,408	0,434	0,433	0,432	0,425	0,428

Fig. 5 Performance changes on parameters varying: the numbers below each histogram represent the different the number of occurrences above which the words are selected; the numbers in the row of the table represent the different sizes of the words window: ± 3, ± 4 and ± 7

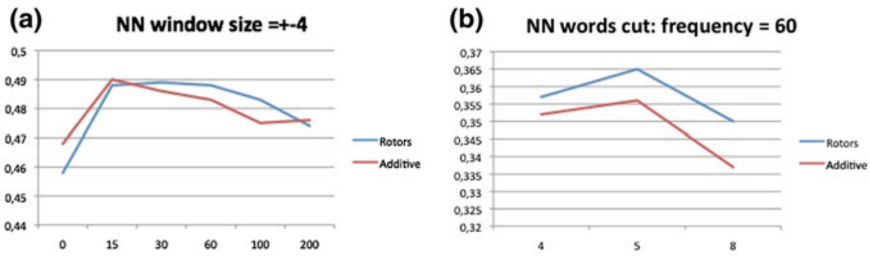


Fig. 6 Results obtained on the NN group fixing the size of the words window to its best value and changing the size of the cutting (**a**) fixing the size of the cutting to its best value and changing the size of the words window (**b**)

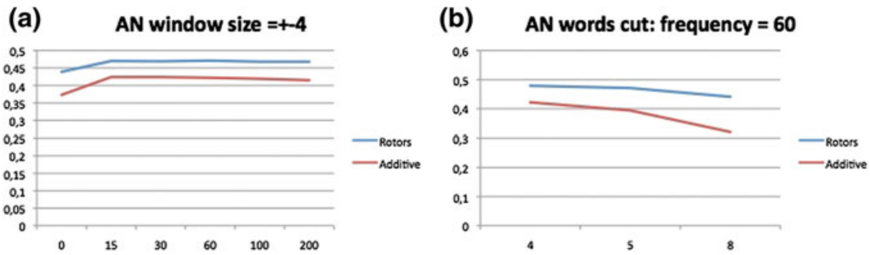


Fig. 7 Results obtained on the AN group fixing the size of the words window to its best value and changing the size of the cutting (**a**) fixing the size of the cutting to its best value and changing the size of the words window (**b**)

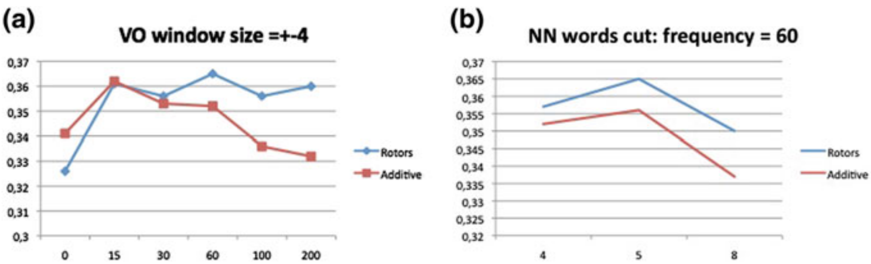


Fig. 8 Results obtained on the VO group fixing the size of the words window to its best value and changing the size of the cutting (**a**) fixing the size of the cutting to its best value and changing the size of the words window (**b**)

the results obtained on the different groups using the rotors-based or the additive operator.

The results reported a small but meaningful change on varying the bigrams window size, as shown in the left side of Figs. 6, 7 and 8: increasing the size of the window the results initially grow, they reach a peak and then they decrease. Moreover the trend obtained from the two operators are similar, however the rotors operator gets better results than the additive operator.

Table 1 Results obtained using the ukWaC corpus to build the semantic space

Space-operator	NN	AN	VO
LSA-Multiplicative	0.215	0.115	0.234
LSA-Additive	0.481	0.394	0.356
LSA-Rotors	0.488	0.471	0.365
Human Agreement	0.49	0.52	0.55

Table 2 Results obtained using the TASA corpus to build the semantic space

Space-Operator	NN	AN	VO
LSA-Multiplicative	0.044	0.090	0.052
LSA-Additive	0.184	0.285	0.189
LSA-Rotors	0.427	0.476	0.301
Human Agreement	0.49	0.52	0.55

Different conclusions can be made changing the value of words occurrences used to filter the set of terms to analyze.

The results are too much sensitive to the cuts and the trends of the two methods are very different. It should be noted for example as in the VO curve our algorithm increases the performance while the additive reaches a minimum by increasing the cut.

Table 1 summarizes the values obtained using the corpus ukWaC fixing the size of the words window to ± 4 and the value of words occurrences used to filter the terms to 60.

Table 2 show the results obtained using the documents of TASA as source corpora, weighting the co-occurrences matrix by means of the *pmi* score, and performing a truncated SVD with a factor $k = 100$.

The last row in the two tables shows the inter agreement among the participants, computed using the leave-one-out resampling according to [19]. According to [6] these values can be considered as upper bounds for our evaluation.

Moreover we have verified that both the models are significantly correlated with the human judgments ($p < 0.01$), and that the rotor model is significantly better ($p < 0.01$) than the standard additive model by using Fisher's z-transformation with the correction reported in [10] (p. 1071).

We carried out a comparison of the performance obtained with the different operators on the entire set of groups, using or not the information relating to the POS tags. The results shown in the Fig. 9 confirm what reported in literature: adding this information the unique words in the data increase, thus aggravating the sparse-data problem [22].

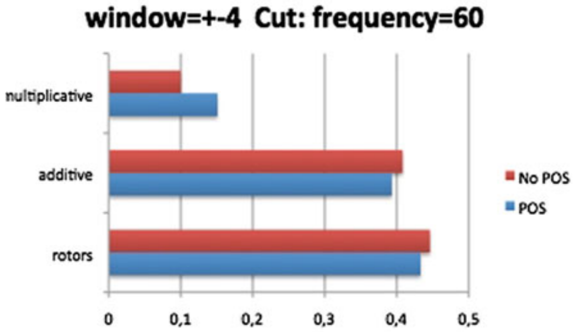


Fig. 9 Results obtained over the all groups using or not the POS tag information

5 Conclusion

In this work we have described a sub-symbolic methodology for sentences encoding. The methodology exploits the properties of Geometric Algebra operators, called rotors, to codify sentences by means of subsequent rotations of an orthogonal basis of a semantic space.

The methodology operates at three different levels: at a word level it is based on the building of an LSA semantic space, at a word-word level it associates ad-hoc semantic rotors to each bigram of a sentence, and finally at a sentence level, it applies the obtained rotors to perform the rotation of the basis. It is easy to show that this kind of coding:

- takes into account the semantics of the words composing it because the rotors are defined in a semantic space generated by LSA;
- it is a function of the words sequence into the sentence, thanks to the non-commutativity property of rotation;
- has a high enough dimensionality;
- is independent of the number of words belonging to the sentence.

The proposed approach has been evaluated according to the GEMS 2011 shared evaluation procedure. Experimental results show that the proposed approach is efficient and outperforms additive and multiplicative operators.

However the potential of the methodology become more evident on a test set of sentences longer than only two words as shown in our previous works [4, 20].

The method can be used to all traditional applications of classical LSA-based approaches, and has the advantage to be fully scalable, since the matrix which generates the coding of the sentences is a word-by-word matrix, and therefore its dimensions depend only on the vocabulary size, and not on the contexts used (i.e. sentences, or documents). On the other hand, Clifford rotors do not depend on the sentence length, being their application just a rotation of a basis in the semantic space.

Acknowledgments We are grateful to Professor Thomas Landauer, to Praful Mangalath and the Institute of Cognitive Science of the University of Colorado Boulder for providing us the TASA corpus. This work has been partially supported by the PON01_01687—SINTESYS (Security and INTElligence SYSstem) Research Project.

References

1. Agostaro, F., Pilato, G., Vassallo, G., Gaglio, S.: A sub-symbolic approach to word modelling for domain specific speech recognition. In: Proceedings of IEEE CAMP. International Workshop on Computer Architecture for Machine Perception, pp. 321–326 (2005)
2. Annesi, P., Storch, V., Basili, R.: Space projections as distributional models for semantic composition. In: Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing - Volume Part I, CICLing'12, pp. 323–335. Springer, Heidelberg (2012)
3. Artin, E.: Geometric Algebra. Interscience Tracts in Pure and Applied Mathematics. Wiley, New York (1988)
4. Augello, A., Vassallo, G., Gaglio, S., Pilato, G.: Sentence induced transformations in "conceptual" spaces. In: IEEE Computer Society on Proceedings of the 2008 IEEE International Conference on Semantic Computing, ICSC '08, pp. 34–41. Washington, USA (2008)
5. Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta, E.: The wacky wide web: a collection of very large linguistically processed web-crawled corpora. Lang. Resour. Eval. **43**, 209–226 (2009)
6. Basile, P., Caputo, A., Semeraro. G.: A study on compositional semantics of words in distributional spaces. In: 2012 IEEE Sixth International Conference on Semantic Computing (ICSC), pp. 154–161
7. Budiu, R., Royer, C., Pirolli, P.: Modeling information scent: a comparison of lsa, pmi and glsa similarity measures on common tests and corpora. In: Large Scale Semantic Access to Content (Text, Image, Video, and Sound), RIAO '07, pp. 314–332. Paris, France (2007). Le Centre de Hautes Etudes Internationales d'informatique documentaire
8. Christopher, D.: Manning and Hinrich Schütze. Foundations of statistical natural language processing. MIT Press, USA (1999)
9. Clark, Stephen, Coecke, Bob, Sadrzadeh, Mehrnoosh: Mathematical foundations for a compositional distributed model of meaning. Linguist. Anal. **36**(1–4), 345–384 (2011)
10. David, J.S.: Handbook of Parametric and Nonparametric Statistical Procedures. 3rd edn. Chapman & Hall/CRC (2003)
11. Dowty, D.R., Wall, R.E., Peters, S.: Introduction to Montague Semantics. vol. 11. D. Reidel, Dordrecht (1989)
12. Edward Grefenstette. Compositionality in distributional semantic models: A critical overview of current literature (2010)
13. GEMS Shared Evaluation. In: Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics, Association for Computational Linguistics. Edinburgh, UK (2011)
14. Hinrich, S.: Word space. In: Giles, L.C., Hanson, S.J., Cowan, J.D. (eds.) Advances in Neural Information Processing Systems 5, pp. 895–902. Morgan Kaufmann, San Francisco, CA (1993)
15. Landauer, T.K., Dumais, S.T.: Solutions to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol. Rev. 104 (1997)
16. Landauer, T.K., Foltz, P.W., Laham, D.: An introduction to latent semantic analysis. Discourse Process. **25**, 259–284 (1998)
17. Lounesto, P.: Clifford Algebra and Spinors. Cambridge University Press, Cambridge (1997)
18. Mitchell, J., Lapata, M.: Vector-based models of semantic composition. In: Proceedings of ACL-08: HLT, pp. 236–244. Columbus, Ohio (2008)

19. Mitchell, Jeff, Lapata, Mirella: Composition in distributional models of semantics. Cogn. Sci. **34**(8), 1388–1429 (2010)
20. Pilato, G., Augello, A., Gentile, M., Vassallo, G.: Geometric encoding of sentences based on clifford algebra. In: IC3K - KDIR 2012 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval - Special Session on Information Filtering and Retrieval: Novel Distributed Systems and Applications (DART). SciTePress, Barcelona (2012)
21. Rubenstein, Herbert, Goodenough, John B.: Contextual correlates of synonymy. Commun. ACM **8**(10), 627–633 (October 1965)
22. Sahlgren, M.: The distributional hypothesis. From context to meaning: Distributional models of the lexicon in linguistics and cognitive science (Special issue of the Italian Journal of Linguistics) **20**(1), 33–53 (2008)
23. Sahlgren, M.: The Word-Space Model. Ph.D. thesis, Stockholm University (2006). http://www.sics.se/mange/TheWordSpaceModel.pdf
24. Schoute, P.H.: Mehrdimensionale geometrie. Leipzig: G.J.Gschensche Verlagsha (Sammlung Schubert XXXV): Die linearen Rume, 1, 1902
25. Schütze, Hinrich: Automatic word sense discrimination. Comput. Linguist. **24**(1), 97–123 (March 1998)