Measuring Adjective Spaces*

Timo Honkela, Tiina Lindh-Knuutila, and Krista Lagus

Adaptive Informatics Research Centre
Aalto University of Science and Technology
P.O. Box 15400, FI-00076 Aalto

Abstract. In this article, we use the model adjectives using a vector space model. We further employ three different dimension reduction methods, the Principal Component Analysis (PCA), the Self-Organizing Map (SOM), and the Neighbor Retrieval Visualizer (NeRV) in the projection and visualization task, using antonym test for evaluation. The results show that while the results between the three methods are comparable, the NeRV performs best of the three, and all of them are able to preserve meaningful information for further analysis.

1 Introduction

Large number of studies indicate that methods using co-occurrence data provide useful information on the relationships between the words, as words with similar or related meaning will tend to occur in similar contexts [1]. This intuition has been carefully assessed, in particular, for nouns and verbs. In this article, we study whether co-occurrence statistics provide a basis for automatically creating a representation for a group of adjectives as well. Further, we compare dimension reduction methods, in particular, the Principal Component Analysis [2], the Self-Organizing Map [3] and Neighbor Retrieval Visualizer (NeRV) [4] affect the quality of the final representation. We study the neighborhoods of the adjectives in the created vector space, and use antonyms pairs to evaluate the result.

Nouns and verbs have received much more attention than adjectives in language technology, knowledge engineering and related research fields. For instance, the nodes of ontologies are mainly entities labeled with nouns. In linguistics, verbs have been the subject of very active study as verbs usually serve as the ordering elements of a sentence. Adjectives are not very well described in ontologies, but Wordnet and EuroWordnet have considered including a small set of lexical conceptual relations that allow to encode adjectives.

The Self-Organizing Map has been earlier used in several studies to create word clusters automatically from statistical features obtained from corpora. In [5], analysis of 150 English word types was carried out using the self-organzing map. The resulting map divided into separate areas of verbs and nouns, with nouns further dividing into areas of animate and inanimate nouns. The verbs were studied in [6], where a verb map was created using features, such as case marking and adverbs. The resulting verb

* This work has been supported by the Academy of Finland and the Finnish Funding Agency for Technology and Innovation.

K. Diamantaras, W. Duch, L.S. Iliadis (Eds.): ICANN 2010, Part I, LNCS 6352, pp. 351–355, 2010.

clusters depicted organization related to emotional content. In addition, [7], an adjective map based on emotive aspects of words was created with manually provided features.

2 Methods

The word vector space model is a standard method for representing text data in numerical form. In the model, words are represented as feature vectors. Features of a word are often words that co-occur with it in a certain context or window [8]. We obtain feature vectors statistically from a corpus, using a window of a small size, and counting the co-occurrences of the feature words that appear with the target word in this window. The similarity of the words can be then measured as the Euclidean distance between them in the vector space.

The original dimensionality of the vector space is usually high and dimesion reduction methods are needed. This can be done by either feature selection, i.e. selecting a subset of the original features that give most information of the object in question, or by feature extraction, using features that are combinations of original dimensions. Both are frequently applied to word vector spaces.

Generally, a dimension reduction method is good, if the neighbors of the data points in the original space can be retrieved well based on the projected points in the visualization. We use three feature extraction methods that project the data to two dimensions for visualization purposes, the Principal Component Analysis [2] the Self-Organizing Map (SOM) [3] and the Neighbor Retrieval Visualizer (NeRV) [4]. In the following, we give the basic details of each method.

The PCA is an orthogonal linear transformation which transforms the possibly correlated data into new variables, in such a way that the greatest variance lies on the first principal component and most of the variance is contained in a few first principal components. This makes it a practical tool for dimension reduction, as the remaining components can be dropped with minimal loss of information.

The SOM is a classical unsupervised learning method which typically produces a two-dimensional discretized representation of the input space. It preserves the topological properties of the input space, which makes it an useful tool for visualizing high-dimensional data.

The novel NeRV method for nonlinear dimensionality reduction and data visualization [4] conceptualises the dimensionality reduction as an information retrieval problem and rigorously quantifies the goodness of the dimension reduction method in terms of precision and recall. The NeRV algorithm [9] is able to optimise the cost function (1) that allows an optimal balance between these two. The cost function is given as

$$E_{NeRV} = \lambda[E_i[D(p_i, q_i)]] + (1-\lambda)E_i[D(q_i, p_i)] \ , \tag{1}$$

where $E_i[D(p_i, q_i)]$ is the number of misses and $E_i[D(q_i, p_i)]$ the number of false positives. Minimising $E_i[D(p_i, q_i)]$ maximizes the recall, and minimizing $E_i[D(q_i, p_i)]$ maximizes the precision. The relative cost parameter λ can be used to focus to either of them.

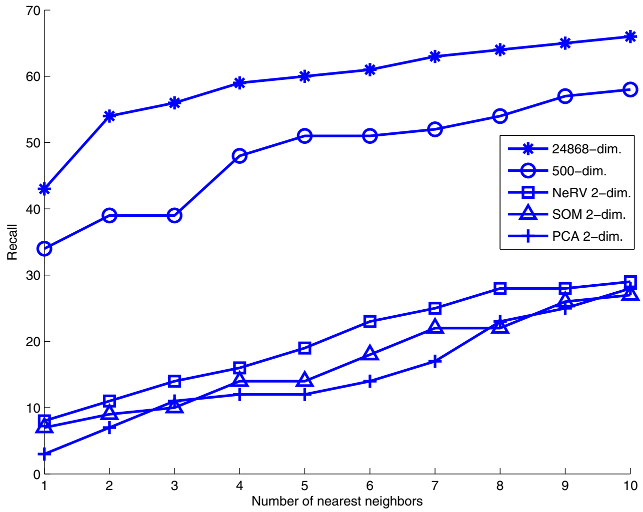

Fig. 1. Comparison of different methods used to create an adjective space based on word co-occurrence statistics. The y-axis shows the percentage of antonyms successfully found for the 72 test adjectives among the *n* nearest neighbors and the x-axis provides different values of *n*, from 1 to 10. The high-dimensional spaces are marked with an asterisk (24868 dimensions) and a circle (500 dimensions). The methods for creating the 2-dimensional projection from the 500-dimensional data are the NeRV (marked with a square), the SOM (marked with a triangle), and the PCA (marked with a plus sign).

3 Experiments

The objective of the experiments was to study the effect of the dimension reduction on the data and see whether there are differences in the dimension reduction methods. The text collection used in the experiments was extracted from English Wikipedia articles. The statistics of the two closest context words were collected for each of the 72 adjectives included in the analysis. For each adjective, a 24868-dimensional feature vector was created. The original feature dimensionality contains all the words that occur in the collection over 100 times. We then reduced the dimensionality of matrix by feature selection: Only the 500 words that occur most frequently with the 72 adjectives are included. Further, we use the matrix with 500-dimensional feature vectors to project the data into two dimensions using the PCA, the SOM and the NeRV. The PCA was implemented using standard Matlab functionalities, the SOM with its common functionalities using the SOM Toolbox. The NeRV is implemented in the dredviz software package developed for information visualization.[1]

[1] The software package, developed in the Adaptive Informatics Research Centre, Aalto University School of Science and Technology, is available at
`http://www.cis.hut.fi/projects/mi/software/dredviz/`

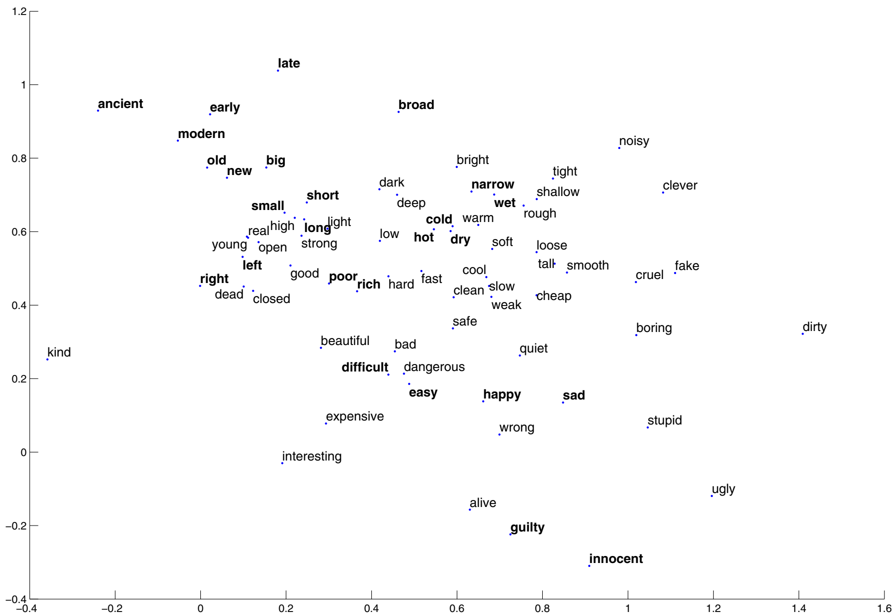

Fig. 2. The set of adjectives used in the study projected into a 2-dimensional space using the Neighbor Retrieval Visualizer (NeRV) method. The words in bold have the antonym in their local neighborhood.

There does not seem to exist a consensus among linguists on how adjectives should be divided into categories in general. We then considered two alternatives, i.e., synonyms and antonyms. Antonyms are words that have an opposite meaning, and synonymous words have a same or almost the same meaning. They both offer a means for evaluating the ordering of the obtained vector space. As synonymity is not clearly defined, we used the antonyms for which the definitions are clearer. Each adjective had an antonym in the set: long–short, good–bad, etc. We then calculated the recall, that is checked whether the antonym could be found within the $[1, 2, \ldots 10]$ nearest neighbors. The result of this experiment is presented in Fig. 1.

The fact that the high-dimensional spaces provide higher percentages than the 2-dimensional spaces is understandable. Lower-dimensional spaces are often used to reach lower computational complexity. Moreover, a 2-dimensional space is particularly useful in visualization. Thus, it is to be noted that the nearest neighbors in the case of SOM, NeRV and PCA were calculated in the 2-dimensional space. The NeRV method seems to reach better performance than the SOM which again exceeds the performance of the PCA method. Figure 2 shows the results for the NeRV. We can see that many antonym pairs are located close to each other in the visualization. The pairs ancient-modern, old-new and early-late form their own cluster in the top-left corner of the figure. Another group is formed by the pairs hot-cold, dry-wet in close vicinity with cool and warm. Other pairs close to each other include poor-rich, innocent-guilty, happy-sad and difficult-easy.

4 Conclusions and Discussion

In this article, we studied the representation of the adjectives based on the short context they appear in. The results, based on an antonym test, show that the context provides a reasonably good means for automatically extracting meaningful relationship between adjectives. While the NeRV performs best in this setting, based on this evaluation, all methods tested preserve meaningful information for further analysis.

The antonym test is a very simple evaluation method which only measures one side of the problem: whether the words with opposite meaning can be found close to each other in the word vector space. A closer look at Fig. 2 reveals that there are several other words for which the meaning is close or which seem to belong to a same category as well. To obtain concrete results, though, we would need linguistically or semantically defined groupings of adjectives and lists of adjectives in each group - which we could then use as a basis of more thorough evaluation.

Acknowledgements

We wish to thank Jaakko J. Väyrynen and Sami Virpioja who have helped us by gathering the basic data used in the experiments and by providing assistance in the data analysis.

References

1. Firth, J.R.: Papers in linguistics 1934-1951. Oxford University Press, Oxford (1957)
2. Jolliffe, I.T.: Principal Component Analysis. Springer, Heidelberg (2002)
3. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (2001)
4. Venna, J., Peltonen, J., Nybo, K., Aidos, H., Kaski, S.: Information retrieval perspective to nonlinear dimensionality reduction for data visualization
5. Honkela, T., Pulkki, V., Kohonen, T.: Contextual relations of words in Grimm tales, analyzed by self-organizing map. In: Proceedings of ICANN 1995, Nanterre, France, vol. II, pp. 3–7. EC2 (1995)
6. Lagus, K., Airola, A.: Semantic clustering of verbs – analysis of morphosyntactic contexts using the SOM algorithm. In: Acquisition and Representation of Word Meaning: Theoretical and computational perspectives, Pisa, Roma. Linguistica Computazionale. Instituti Editoriali E Poligrafici Internazionali, vol. XXII-XXIII, pp. 263–287 (2005)
7. Honkela, T.: Adaptive and holistic knowledge representations using self-organizing maps. In: Proceedings of Int. Conference on Intelligent Information Processing, IIP 2000, pp. 81–86. IFIP (2000)
8. Schütze, H.: Word Space. In: Advances in Neural Information Processing Systems, NIPS Conference, vol. 5, pp. 895–902. Morgan Kaufmann Publishers Inc., San Francisco (1993); Journal of Machine Learning Research 11, 451–490 (2010)
9. Venna, J., Kaski, S.: Nonlinear dimensionality reduction as information retrieval. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics, AISTATS 2007 (2007)