Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Neurocomputing 70 (2006) 62–69
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Improving self-organization of document collections by semantic mapping

Renato Fernandes Corrêa[a,b], Teresa Bernarda Ludermir[b,*]

[a]Polytechnic School, Pernambuco University, Rua Benfica, 455, Madalena, 50.750-410 Recife—PE, Brazil
[b]Center of Informatics, Federal University of Pernambuco, P.O. Box 7851, Cidade Universitária, 50.732-970 Recife—PE, Brazil

Available online 22 August 2006

Abstract

In text management tasks, the dimensionality reduction becomes necessary to computation and interpretability of the results generated by machine learning algorithms. This paper describes a feature extraction method called semantic mapping. Semantic mapping, sparse random mapping and PCA are applied to self-organization of document collections using self-organizing map (SOM). The behaviors of the methods on projection of binary and tfidf document vector representations are compared. The classification error generated by SOM maps on text categorization of the K1 collection was used to compare the performance of the methods. Semantic mapping generated better document representation than sparse random mapping.

Keywords: Dimensionality reduction; Semantic mapping; Sparse random mapping; Self-organizing map; Document organization

1. Introduction

Nowadays, the ever increasing volume of documents in digital form available inside companies, institutions, military and government's sectors, and public available by means of the Internet, has allowed major access to information and knowledge acquisition. This huge amount of digital data has increased the velocity and quality of decisions making and generate new discoveries, methodologies, technologies, products, publications and need for more information. Thus even more sophisticated methods and systems to organize and allow flexible access to documents of large document collections are needed.

The systems implemented to assist users in finding documents that contain information relevant to their particular needs are denominated information retrieval systems (IRS) [24]. The IRS design and implementation are studied in information retrieval (IR) research area.

After the seminal works of Lin et al. [11] and Scholtes [17], many researches focus on the application of self-organizing maps (SOM) [9] in IRS implementation [3,10,12,15]. The SOM is used to generate document clusters and a two-dimensional graphical representation of document similarities and topics, called document map. The goals with the use of SOM are to provide IRS with the capacity of automatically organize or structure a corporate document base according to textual similarities, making easier the search of documents and discovery of related documents or topics. Nearby locations on the display contain similar documents, thus the maps are a meaningful visual background providing: an overview of the topics present in the collection and means to make browsing and content-addressable searches.

A problem that can make difficult the application of SOM and an others machine learning or data analysis algorithms to information retrieval of large documents collections is their computation complexity. In text management tasks, high-dimensional data vectors normally represent the documents; the length of these vectors is equal to the number of distinct terms in the vocabulary of the corpus. Thus to turn computational feasible the use of machine learning algorithms the dimensionality of vectors that represent the content of documents, called document vectors, must be reduced to few hundreds, turning essential the use of dimensionality reduction methods.

*Corresponding author. Tel.: +55 81 21268430; fax: +55 81 21268438.
E-mail address: tbl@cin.ufpe.br (T.B. Ludermir).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.07.007

In the especial case of SOM algorithm, this problem has been addressed by WEBSOM project [10]. WEBSOM is a method for organizing textual documents onto a two-dimensional map display using SOM. The main goal of the WEBSOM project was to scale up the SOM algorithm to be able to deal with large amounts of high-dimensional data. In a practical experiment were mapped 6.840.568 patent abstracts onto a 1.002.240-node SOM. This was possible due to the use of shortcuts in the creation of large maps from trained small ones and mainly due to the use of a dimensionality reduction method called random mapping (RM) [10].

We proposed a feature extraction method called semantic mapping (SM) [4] that analytically and experimentally has given superior performance than the use of sparse RM [10] and performance close to principal component analysis (PCA) [7,13,14] in dimensionality reduction of binary document vectors. However, in IR, documents are most of the times represented by vectors of weighted terms using the *tfidf* function [18].

The objective of this paper is to report a more deep analyses on how the methods behave in the application on different document vectors representation (binary and *tfidf* representations) to obtain several reduced dimensionality, extending the work done in [4].

This paper is organized as follows. In Sections 2 and 3 are given an overview of self-organizing of document collections and dimensionality reduction methods, respectively. In Sections 4 and 5, the RM method and the SM method are, respectively, described. Section 6 describes the methodology and results of the experiments on text categorization. The experimental results obtained are used to measure and compare the performance of dimensionality reduction methods. Section 7 contains the conclusions and future works.

2. Self-organization of document collections

The construction of IRS using SOM, involves four steps: document indexing, dimensionality reduction, construction of document map and construction of user interface (see Fig. 1).

The document indexing step consists on preprocess the text documents and represent them statistically. Generally, non-informative words are removed from initial vocabulary and word affixes are removed using a stemmer algorithm [18]. The isolated words without affixes are called terms. The documents are represented using the vector space model [16], i.e. the documents are represented by vectors, where terms are the indexes and the corresponding values represent the importance of a term to the semantics of a document. In IR literature, typically, the frequency of occurrence of a particular term in a document (*tf*) weighted by inverse document frequency (*idf*) function is used to approximate the importance of a term in a document, i.e. *tfidf* function [18]. Another alternative is to use binary document vectors to represent documents; in this case, each position in the vector indicates the presence or the absence of determined term in the document.

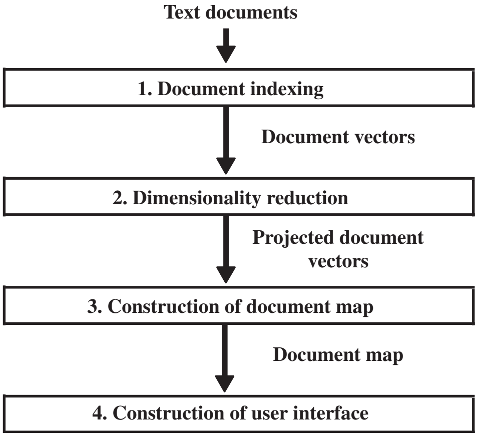
Text documents
1. Document indexing
Document vectors
2. Dimensionality reduction
Projected document vectors
3. Construction of document map
Document map
4. Construction of user interface

Fig. 1. Overview of IRS construction using self-organizing maps.

Dimensionality reduction step receive the document vectors generated in document indexing and apply some algorithm to reduce the number of dimensions or terms. These methods are discussed in Section 3.

The document vectors in reduced dimensionality are used as input to the step of construction of the document map. This step consists in training a SOM map with the input document vectors.

The last step is the construction of the user interface. The user interface must allow interactive browsing, content-addressable and keyword searches on the document map [10].

The construction of document map is the main step, due to the great influence on overall performance of the IRS. The map must organize the documents generating consistent clusters. In good quality document maps, similar document vectors must be mapped in the same node or neighboring nodes.

Although the frequent problem of inter-indexer inconsistency [18], as similar documents have great probability of belong to same category, the classification error or accuracy [10] on text categorization may be used as indicator of quality of document maps. Text categorization [18] is a process of classifying documents by grouping them into one or more existing categories according to the themes or concepts present in their content.

Document maps with minimal classification error are desired and considered of superior quality. The maps are desired because they represent the document similarity in a close way to the human being. The classification error in a test set is the best measure of the generalization of the cluster structure found by SOM, and can express better the quality of the document map.

The quality of the document map receive great influence of the document indexing and dimensionality reduction steps, given that if the semantic similarity of the documents is clearly expressed by the similarity of the document vectors, then best quality document maps are generated. Thus, in controlled experiments, the classification error in test set generated by document maps may be used as indicator of quality of the document representation generated by dimensionality reduction method and used in training of SOM map.

3. Dimensionality reduction methods

When the data vectors are high dimensional, it is computationally infeasible to use data analysis or pattern recognition algorithms that repeatedly compute similarities or distances in the original data space. For example, some neural networks, like most statistical models, cannot operate with tens of thousands of input variables [21]. In addition, to interpret the results and mining knowledge of models generated by machine learning algorithms from high-dimensional data is a difficult task.

Thus, dimensionality reduction methods are necessary to document representation in text management tasks where normally the vector space model is used to represent documents. Generally, the dimension of document vectors is equal to the size of corpus' vocabulary, about 50,000 words in large corpus [10].

The objective of use dimensionality reduction methods in self-organization of document maps is to reduce the dimensionality of the original vectors used in document representation, generating vectors with reduced dimensionality (projected vectors) that minimize the computational cost of SOM training, without compromise the quality of the generated document map.

There are many methods to dimensionality reduction. An orthogonal distinction may be drawn in terms of the nature of the resulting terms: term selection methods or term extraction methods [18].

The feature selection methods select a subset of the original set of features using a rank metric. Inverse document frequency (*idf*), chi-squared and information gain are examples of this kind of method [23].

The feature extraction methods extract a small set of new features generated by combinations or transformations of the original ones. These methods are based in generation of matrices of projection that multiplied by original feature vectors result in projected vectors with reduced dimensionality. The difference among those methods is the technique used to construct the matrices of projection.

Latent semantic indexing (LSI) [5] and PCA [7,13,14] are feature extraction methods based in the estimation of principal components from term by document matrix. These methods often outperform the feature selection methods [18]. LSI and PCA have the potential benefit to detect synonyms as well as words that refer to the same topic, however their computation needs is hard to high-dimensional data.

The RM and SM are feature extraction methods and are described in the following sections.

4. Random mapping

The RM method [8] was generated and used in the context of WEBSOM project [10]. RM is a method generally applicable that approximately preserves the mutual similarities between the data vectors [8].

RM consists in construction and use of a matrix R of random values normally distributed, with the Euclidean length of each column normalized to unity. R multiplies each original n-dimensional data vector, denoted by x_j, forming the y_j d-dimensional representation of each one, i.e. the mapping is done taking

$$y_j = Rx_j.$$

The computational complexity of forming the random matrix O(nd), is negligible to the computational complexity of estimating the principal components, $O(Nn^2) + O(n^3)$ [8]. Here n and d are the dimensionalities before and after the RM, respectively, and N is the number of data vectors.

Analytically and experimentally, RM shows to be efficiently, producing results as good as PCA or the use of data vectors in the original space, when the final dimensionality is larger than 100 features [8,10]. Furthermore, the training SOM algorithm has low sensitivity to distortions of similarity caused by RM [8].

To increase the computation speed of RM, simplifications in the construction of the matrix R was suggested and evaluated experimentally [10]. One of these simplifications constructed R as a sparse matrix, where a fixed number of ones (or 5 or 3 or 2) were randomly generated in each column (determining in which extracted features each original feature will participate), and the others elements remained equal to zero. This sparse random mapping (SRM) method generated results close to the original method. The performance of SRM was directly proportional to the number of ones in each column (better performance was generated with 5 ones). The SRM was used successfully also in [1].

In experiments realized in [4], the SRM was used to dimensionality reduction of binary document vectors. The behavior of the method was not the same: the number of ones needed to generate best results was 2 and it was not proportional to the numbers of ones in each column; and SRM only generated acceptable document representations with the dimension of projection larger than 400.

Thus, experiments with other document representations are needed to discovery the true behavior of the method.

5. Semantic mapping

The SM method was developed in research for a better representation of documents in text categorization task and showed to extract more representative and better interpretable features than SRM [4].

The SM is a specialization of the SRM. SM incorporates semantics of the features, captured from data-driven form, in construction of new features or dimensions. This method consists of the steps listed in Fig. 2.

Initially, the N document vectors of a training set are considered as meta-features that describe semantically the original features (terms obtained from document indexing of the collection). In text categorization, this description has direct interpretation because the semantics or means of a term can be deduced analyzing the context where this is applied, i.e., the set of documents (or document vectors) where it occurs [19]. Semantic description of the features or terms corresponds to get the transpose of the matrix document by terms where each line represents a document vector; each line of the transposed matrix is a vector that describes a term, called term vector.

In the second step, term vectors are grouped in semantic clusters training a SOM map. In SOM maps, similar training vectors are mapped in the same node or neighboring nodes [11], as similar vectors represent co-occurrent terms, clusters of co-occurrent terms are formed. In text categorization, these clusters typically correspond to topics or subjects treated in documents and probably contain semantic related terms. The formed maps are called semantic maps. The number of nodes in semantic map must be equal to the number of extracted features wanted.

The construction of the matrix of projection, the third step, is done as follows: after the training of semantic map, each term vector is mapped in a fixed number of k nodes that better represent it, i.e., the k first nodes that has the closest model vector to the vector that describes the original feature; let n be the number of original features and d be the number of extracted features, the matrix of projection M was constructed with d lines and n columns, with m_{ij} equals to one if the original feature j was mapped into node i, zero otherwise. The position of the ones in the columns of the projection matrix indicates which extracted features each original feature will participate. While in SRM the position of the ones in each column of R is determined randomly, in the method of SM the position of the ones in each column of M is determined in accordance with the semantic clusters where each original feature was mapped. The set of matrices of projection generated by SM is a subset of that generated by SRM and RM, thus SM also approximately preserves the mutual similarities between the data vectors after projection to reduced dimension.

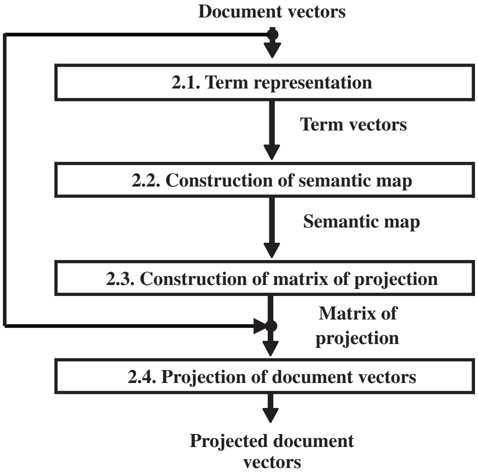
Document vectors
2.1. Term representation
Term vectors
2.2. Construction of semantic map
Semantic map
2.3. Construction of matrix of projection
Matrix of projection
2.4. Projection of document vectors
Projected document vectors

Fig. 2. Overview of semantic mapping method.

Finally, the mapping of the original document vectors to the reduced dimension is made multiplying the matrix of projection M by it. After the mapping, the generated vectors may be normalized in unitary vectors.

The computational complexity of the SM method is $O(nd(iN+k))$ that is the complexity of the construction of the semantic map with d units by SOM algorithm from n term vectors with N dimensions (number of document vectors in training set) for i epochs plus the superior complexity of the construction of the matrix of mapping with k ones in each column. This complexity is smaller than the complexity of PCA, and still linear to the number of characteristics in the original space as the RM.

As the extracted features by SM are, in theory, more representative of the content of the documents, beyond better interpretable that those generated by RM, then the representation of documents generated by SM is expected to improve the performance of the machine learning algorithms in relation to the use of the one generated by RM.

In experiments realized in [4], the SM performs better than SRM in projection of binary document vectors. This better performance is also confirmed empirically in the experiments related in the next session, where the methods were used to project *tfidf* document representation.

6. Experiments

In this session is presented the adopted methodology and the results of the experiments. The experiments consist of the application of SM, SRM and PCA to a problem of text categorization using SOM maps as classifier.

Classification error in text categorization was used as indicator of quality of the document representation generated by each method, i.e. the performance of each dimensionality reduction method. The classification error was evaluated in the same training and test sets of the K1 collection used in [4].

The performances achieved by SRM, SM and PCA are compared in projection of binary document vectors, measured in previous experiments [4], and in projection of *tfidf* document vectors, reported in this article. The goal is to analyze the behavior of each method in projection of different document representations.

In binary document representation, each position in the document vector indicates the presence or the absence of determined term in the document. In *tfidf* representation [18], the documents are represented by real vectors in which each component corresponds to the frequency of occurrence of a particular term in the document (*tf*) weighted by a function of the *idf*.

6.1. Document collection preprocessing

The documents categorized belong to K1 collection [2]. This collection consists of 2340 Web pages classified in one of 20 news categories at Yahoo: health, business, sports, politics, technology and 15 subclasses of entertainment (without subcategory, art, cable, culture, film, industry, media, multimedia, music, online, people, review, stage, television, variety).

The document vectors of the collection were constructed using the vector space model with term frequency. These vectors were preprocessed eliminating generic and non-informative terms [2]; the final dimension of the vectors was equal to 2903 terms.

After preprocessing, the document vectors were divided randomly for each category in half for training set and half for test set; the length of each set was 1170 document vectors.

The categories were codified and associated to document vectors as labels.

6.2. Methodology

The *tfidf* document representation was calculated as function of term-frequency document vectors as described in [18].

The performance of the projection methods for *tfidf* document representation was measured, in each dimension of projection (100, 200, 300 and 400), by the mean classification error generated by five SOM maps in the categorization of projected document vectors of a test set, trained with the respective projected document vectors of a training set.

The length of projected document vectors were normalized, thus the direction of the vector reflects the contents of the document [8].

The classification error for a SOM map is the percentage of documents incorrectly classified when each map unit is labeled according to the category of the document vectors in training set that dominated the node. Each document is mapped to the map node with the closest model vector in terms of Euclidean distance. The document vectors of the test set received the category assigned to the node where they were mapped. These SOM maps are denominated document maps.

To measure the performance of the methods SM and SRM in relation to the number of ones in each column, for each pair combining dimension and number of ones, were generated 30 matrices of projection for each method. The number of ones in each column in the projection matrix was: 1, 2, 3 and 5. In the case of SM, for each dimension, 30 semantic maps were generated, and for each one of these, 4 matrices of mapping with 1, 2, 3 or 5 ones in each column were generated. For the SRM in each dimension, 30 matrices of projection with only one in each column were first constructed, and from these matrices, others matrices were constructed successively adding randomly ones until reaching the number of necessary ones in each column.

The PCA method involves the use of singular value decomposition (SVD) method [6] in the extraction of the principal components of the matrix of correlation of the terms in the training set. The correlation matrix was calculated on a *tfidf* matrix of terms by documents. The components are ordered in such way that the first ones describe most of the variability of the data. Thus, the last components can be discarded. Given that the components were extracted, four matrices of projection were constructed, one for each dimension, taking the 100, 200, 300 and 400 first components, respectively.

The matrices of projection generated by the three methods had been applied on *tfidf* document vectors, thus forming the projected vectors in the reduced dimensions. The projected vectors of the training and test sets had been normalized, and were used to construct the document maps and to evaluate the performance of methods, respectively.

The algorithm used for training SOM maps was batch-map SOM [10] because it is quick and have few adjustable parameters. The SOM maps used to construct the semantic maps and document maps had a rectangular structure with a hexagonal neighborhood to facilitate visualization. The Gaussian neighborhood function was used. For each topology, the initial neighborhood size was equal to half the number of nodes with the largest dimension plus one. The final neighborhood size was always 1. The number of epochs of training was 10 in rough phase and 20 in the fine-tuning phase. The number of epochs determines how mild the decrease of neighborhood size will be, since it is linearly decreasing with the number of epochs. The dimensions of document maps were 12×10 units (as suggested in WEBSOM project [8]) with the model vectors with 100, 200, 300 and 400 features. Because there is no prior knowledge in word clustering, the semantic maps had the most squared possible topologies: 10×10, 20×10, 20×15 and 20×20, with the model vectors with 1170 features. For all SOM maps topology, it was used the same randomly initialized configurations obtained in the previous experiments [4] using the som_randinit function of somtoolbox.

6.3. Results

The first step was the evaluation of the number of ones needed in each column of the matrices of projection generated by SRM and SM in order to minimize the mean classification errors in the test set. The *t*-test of combined variance [20] was used to compare the performances of the methods with different numbers of ones. The *t*-test was applied on the average and the standard deviation of the classification errors achieved by each method in the test set. The results are presented in Tables 1 and 2.

In Tables 1 and 2 the following codification of the *P*-value in ranges was used [22]: "$\gg$" and "$\ll$" mean that the *P*-value is lesser than or equal to 0.01, indicating a strong evidence of that a system generates a greater or smaller classification error than another one, respectively; "<" and ">" mean that the *P*-value is bigger than 0.01 and minor or equal to 0.05, indicating a weak evidence that a system generates a greater or smaller classification error than another one, respectively; "∼" means that the *P*-value is greater than 0.05 indicating that it does not have significant difference in the performance of the systems.

Table 1 shows that in the experiments using *tfidf* document representation, as in previous experiments using binary document representation [4], the SM generates one better representation of documents in all the dimensions when 2 ones was used in each column of the projection matrix. However, in contrast to experiments with binary document representation, the differences among the use of 2, 3 or 5 ones is not significant, thus the number of ones in each column of the matrix of projection of SM has small influences in its performance for values greater or equal to 2. In experiments using the binary document representation [4], the use of 1 or 5 ones generated superior classification errors, the use of 3 ones generated classification errors superior or equivalent to the use of 2 ones, and the number of ones in each column of the matrix of projection of SM method had great influences in its performance, since the majority of the evidences were strong ones.

Table 1
Results of *t*-test on SM number of ones in each column of the matrices of projection

Dimension	2–1	3–1	3–2	5–1	5–2	5–3
100	≪	≪	∼	≪	∼	∼
200	<	<	∼	∼	∼	∼
300	≪	<	∼	<	∼	∼
400	≪	≪	∼	≪	∼	∼

Table 2
Results of *t*-test on SRM number of ones in each column of the matrices of projection

Dimension	2–1	3–1	3–2	5–1	5–2	5–3
100	∼	∼	∼	∼	∼	∼
200	∼	∼	∼	∼	∼	∼
300	∼	∼	∼	∼	∼	∼
400	∼	∼	∼	∼	∼	∼

In Table 2, SRM shows to be not sensible to the number of ones in each column of the projection matrix, since in all the times the results had been equivalents. The insensibility of SRM to the number of ones is a fact already expected due to the purely random nature of SRM in extract features, however, comparing with the results generate with the use of binary document representation [4], the *tfidf* document representation increase the insensibility of the method to the number of ones. In experiments using binary document representation, SRM shows to generate better representation of documents in all the dimensions when 2 ones was generated in each column, minimizing the classification errors, but most of the time the performance with different number of ones were equivalents.

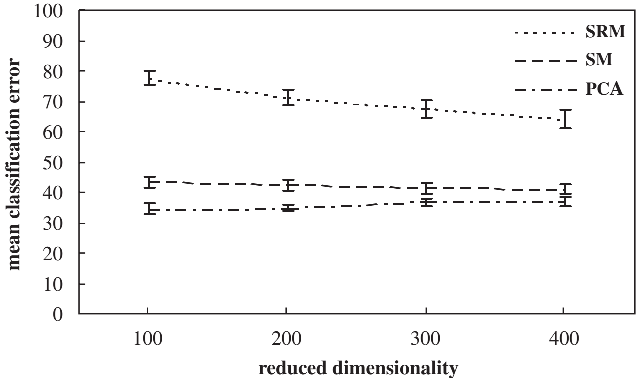

Fig. 3. Mean classification error as function of reduced dimension of document vectors. The bars denote one standard deviation over 150 experiments for SRM and SM (combination of the 30 matrices of projection with the five document maps) and five for PCA (combination of the matrix of projection with the five document maps).

Fig. 3 shows the averaged classification error in the test set generated by SRM, SM and PCA in function of the reduced dimensions. It was used 2 ones in each column of the matrices of projection for methods SRM and SM. The methods projecting *tfidf* document vectors did not generate classification errors smaller than the methods projecting binary document vectors. However, the behaviors of the methods were basically the same in the two sets of experiments. The performance of SM had a small variation, but the classification error decreases with increasing of the dimension of projection. The SRM's classification error decreases significantly with increasing of the dimension of projection. The PCA performance had a small variation, the classification error increases with the increasing of the dimension of projection, and this is because the principal components after 100 incorporate the variability of the noise. PCA had the best performance, followed for SM that generated classification errors smaller than SRM for all the reduced dimensions. The difference

Table 3
Better results generated by method

Method	Dimension	Training set mean classification error	Training set standard deviation of classification error	Test set mean classification error	Test set standard deviation of classification error
PCA	100	28.19	1.75	34.39	1.79
SM	400	33.66	1.18	41.15	1.51
SRM	400	52.75	2.14	64.1	3.15

between the performances of PCA and SRM is strong significant but lesser than 10%; this fact makes SRM a good alternative to PCA when the computational cost of PCA is very high.

Table 3 shows the best results achieved by each method. The dimension of projection 400 minimizes the averaged classification error for SRM and SM; for PCA the dimension that minimizes the classification errors is 100. All the differences between the performance of the methods are strong significant.

7. Conclusions

Analytically and experimentally, the characteristics extracted through the method of semantic mapping showed to be more representative of the content of the documents and better interpretable than those obtained through sparse random mapping in projecting binary and *tfidf* document representations.

The behavior of SM, SRM and PCA is basically the same projecting binary or *tfidf* document vectors of the K1 collection. The behavior of SRM is different from the one reported in [10].

SM showed to be a viable alternative to PCA in the dimensionality reduction of high-dimensional data due to the close performance to PCA and the computational cost linear to the number of characteristics in the original space as SRM.

Future investigations should consider to test SM, SRM and PCA methods in other document collections; to research methods to reduce the computation complexity of SM; and to elaborate and evaluate methods of construction of document maps based on the respective semantic map of terms.

Acknowledgments

The authors would like to thank CNPq and CAPES (Brazilian research agencies) for their financial support.

References

[1] E. Bingham, J. Kuusisto, K. Lagus, ICA and SOM in text document analysis, Proceedings of the SIGIR'02, Tampere, Finland. August 11–15, 2002.

[2] D. Boley, M. Gini, R. Gross, E. Han, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, J. Moore, Partitioning-based clustering for web document categorization, Decision Support Syst. 27 (1999) 329–341.

[3] H. Chen, C. Schuffels, R. Orwig, Internet categorization and search: a machine learning approach, J. Visual Commun. Image Representation 7 (1) (1996) 88–102.

[4] R.F. Correa, T.B. Ludermir, Dimensionality reduction by semantic mapping, Proceedings of VIIIth Brazilian Symposium on Neural Networks, vol. 1, IEEE Computer Society Press, Silver Spring, MD, 2004.

[5] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, Indexing by latent semantic analysis, J. Amer. Soc. Inf. Sci. 41 (1990) 391–407.

[6] G.E. Forsythe, M.A. Malcolm, C.B. Moler, Computer Methods for Mathematical Computations, Prentice Hall, Englewood Cliffs, NJ, 1977.

[7] H. Hotteling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (1933) 417–441 (498–520).

[8] S. Kaski, Dimensionality reduction by random mapping: fast similarity computation for clustering, Proceedings of the IJCNN'98 International Joint Conference on Neural Networks, vol. 1, 1998, pp. 413–418.

[9] T. Kohonen, Self-Organizing Maps. 2nd ed., Springer, Berlin, 1997.

[10] T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, A. Saarela, Self organization of a massive document collection, IEEE Trans. Neural Networks 11 (3) (2000) 574–585.

[11] X. Lin, D. Soergel, G. Marchionini, A self-organizing semantic map for information retrieval, Proceedings of the 14th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, 1991, Chicago, IL, pp. 262–269.

[12] D. Merkl, A.M. Tjoa, The representation of semantic similarity between documents by using maps: application of an artificial neural network to organize software libraries, in: Proceedings of the FID'94, General Assembly Conference Congress International Federation Information Documentation, 1994.

[13] E. Oja, A simplified neuron model as a principal component analyser, J. Math. Biol. 15 (1982) 267–273.

[14] K. Pearson, On lines and planes of closest fit to systems of points in space, Philos. Mag. 2 (1901) 559–572.

[15] D.G. Roussinov, H. Chen, A scalable self-organizing map algorithm for textual classification: a neural network approach to thesaurus generation, Commun. Cognit. Artif. Intell. J. (CC-AI) 15 (1–2) (1998) 81–111.

[16] G. Salton, M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1983.

[17] J.C. Scholtes, Unsupervised learning and the information retrieval problem, in: Proceedings of the IJCNN'91, International Joint Conference on Neural Networks, Singapore, vol. 1, 1991, pp. 95–100.

[18] F. Sebastiani, Machine learning in automated text categorization, Proc. ACM Comput. Surveys 34 (1) (2002) 1–47.

[19] G. Siolas, F. d'Alché-Buc, Mixtures of probabilistic PCAs and Fisher kernels for word and document modeling, Proceedings of International Conference on Artificial Neural Networks (ICANN 2002), Madrid, Spain, August 28–30, 2002, Lecture Notes in Computer Science, vol. 2415, Springer, Berlin, 2002, pp. 769–776. ISBN 3-540-44074-7.

[20] M.R. Spiegel, Schaum's Outline of Theory and Problems of Statistics, McGraw-Hill, USA, 1961.

[21] E. Wiener, J.O. Pedersen, A.S. Weigend, A neural network approach to topic spotting, in: Symposium on Document Analysis and Information Retrieval (SDAIR'95), University of Nevada, Las Vegas, Nevada, Las Vegas, 1995, pp. 317–332.

[22] Y. Yang, X. Liu, A re-examination of text categorization methods, Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99), 1999, pp. 42–49.

[23] Y. Yang, J.P. Pedersen, A comparative study on feature selection in text categorization, Proceedings of the 14th International Conference on Machine Learning (ICML'97), 1997, pp. 412–420.

[24] B. Yates, R. Neto, Modern Information Retrieval, Addison Wesley, Reading, MA, 1999.

Renato FernandesCorrêa received the B.S. degree in Computer Science in 2000 from Federal University of Viçosa, Brazil. He received the M.S. degree in Computer Science in 2002 from Federal University of Pernambuco, Brazil. Currently, he is a Doctoral student of the Center of Informatics at Federal University of Pernambuco, and Visiting Professor at Polytechnic School at Pernambuco University, Brazil. His research interests include neural networks and information retrieval systems.

Teresa Bernarda Ludermir received the Ph.D. degree in Artificial Neural Networks in 1990 from Imperial College, University of London, UK. From 1991 to 1992, she was a lecturer at Kings College London. She joined the Center of Informatics at Federal University of Pernambuco, Brazil, in September 1992, where she is currently a Professor and the head of the Computational Intelligence Group. She has published over a 150 articles in scientific journals and conferences, two books in NN and organized two of the Brazilian Symposium on Neural Networks. She is one of the editors-in-chief of the International Journal of Computation Intelligence and Applications. Her research interests include weightless NN, hybrid neural systems and applications of NNs.