Exploring the Use of Random Indexing for Retrieving Information

Maya Carrillo Ruiz, Chris Eliasmith, Aurelio López López

Technical Report No. CCC-08-006
July 31th, 2008

Luis Enrique Erro 1
Sta. Ma. Tonantzintla,
72840, Puebla, México.

Exploring the Use of Random Indexing for Retrieving Information

Maya Carrillo Ruiz [1], Chris Eliasmith[2], Aurelio López López[3]
[1, 3]Coordinación de Ciencias Computacionales,
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro 1, Sta.Ma. Tonantzintla
72840, Puebla, México

[2] Department of Philosophy, Department of Systems Design Engineering
Centre for Theoretical Neuroscience
University of Waterloo
200 University Avenue West
Waterloo, Canada

[1,3]{cmaya,allopez}@inaoep.mx
[2]{celiasmith}@uwaterloo.ca

Resumen. Este reporte evalúa representaciones para documentos con el propósito de realizar búsqueda en corpus. Se incluyen representaciones novedosas, que omiten e incorporan estructura textual. Para codificar relaciones entre palabras (p. ej. frases nominales, sujeto-verbo, objeto-verbo, y adverbio-verbo) se emplean representaciones holográficas reducidas (HRRs). El objetivo principal fue la utilización de la indexación aleatoria (RI) para construir índices de documentos. La RI hace uso de información de co-ocurrencia entre palabras para generar vectores de contexto. Las relaciones son extraídas del texto, codificadas mediante HRRs y sumadas para formar el vector del documento correspondiente. Mostramos que esta representación puede emplearse de manera exitosa para recuperar información y reducir la dimensión del modelo vectorial tradicional. Desafortunadamente no se comporta al nivel de los métodos estándares de bolsa de palabras (BoW). Sin embargo, una representación que utiliza sólo vectores índices se comporta de manera equivalente a la BoW.

Palabras clave: Recuperación de Información, Modelo de Recuperación, Modelo Vectorial, Relaciones Textuales, Frases Nominales, Indexación Aleatoria

Abstract. This report evaluates representations for documents for the purpose of searching textual corpora. These representations include novel proposals that do and do not incorporate textual structure through the use of holographic reduced representations (HRRs) which can encode relations between words (i.e. noun phrases, subject-verb, object-verb, and adverb-verb). A main focus is on documents that are indexed using random indexing, which uses co-occurrence information among words to generate semantic context vectors. Relations are then extracted from the text, encoded using HRRs, and added together to form the corresponding document vector. We show that this representation can be successfully used in information retrieval, and reduces the dimensionality of the traditional vector model. Unfortunately, it does not perform to the level of standard bag-of-words (BoW) methods. However, a simpler representation that uses only index vectors performs as well as BoW.

Key words: Information Retrieval, Retrieval Model, Vector Model, Text Relations, Noun Phrases, Random Indexing

1. Introduction

The huge amount of information available on the Web has resulted in information search becoming a tool of daily life. However, going through hundreds or thousands of irrelevant hits can be tedious. As a result, research in information retrieval (IR) in computer science is dedicated to the theory and practice of searching information in an attempt to improve this situation.

The classic IR techniques rest on the assumption that if a document and a query have a word in common, then the document is about the query. If the number of words in common increases, the relation is stronger. Under this assumption, the IR problem is reduced to determining to what extent the keywords in the user query matches those representing the documents. This approach presents two difficulties: 1) the vocabulary problem (different documents describe the same topic using different words); and 2) the need of having a good mechanism to rank the documents in order of relevance.

The vector space model (VSM) for document representation to support search is probably the most well-known IR model [1, 2]. In this model the terms (words) are represented by vectors in an *n*-dimensional space, where *n* is the number of indexing terms in the corpus to be searched, and each dimension corresponds to one term. The documents and the query are represented as linear combination of the terms that occur in them. Given a query, a vector system produces a ranked list of documents ordered by similarity to the query, where the similarity between a query and a document is computed using the cosine of the angle formed by their corresponding vectors.

The effectiveness of a search is typically quantified using two metrics: recall and precision. Precision is the ratio of relevant documents retrieved to the total number of retrieved documents. Recall is the ratio of the relevant documents retrieved to the total number of relevant documents existing in the corpus or collection.

The VSM assumes that term vectors are pair-wise orthogonal. This assumption is very restrictive because words are not independent, as different words are used to represent the same, or similar, concepts (e.g. 'cup' and 'mug'). Moreover, words are combined in phrases and larger structures and thus remain joined by relations such as structural dependencies, co-references, semantic roles, speech dependency, intentions, etc.

There have been various attempts to solve the vocabulary problem. This problem includes both the problem of synonymy (the same concept can be expressed using different words), and polysemy (the same word can express different concepts). It has been argued that the use of dependence can help solve these problems [3, 4]. Most of the methods use co-occurrence data to include groups of words, like phrases or expressions that denote meaningful entities or relations within the search domain. One method to deal with the vocabulary problem is to build representations for documents and queries that are semantically richer than only vectors based on the frequency of terms occurrence. One example is Latent Semantic Indexing (LSI), which assumes that there is some underlying latent semantic structure (concepts) that can be estimated by statistical techniques. LSI uses singular value decomposition (SVD) to analyze a large term-document matrix and construct a semantic space wherein terms and documents that are closely associated are placed near each other. The new space will have a lower dimensionality than the original matrix [5]. Another approach is Generalized Vector Space Model (GVSM)[6,7], which involves the derivation of new (fundamental) concepts from the terms used to index documents in a collection and, subsequently, the use of these concepts as a new orthogonal basis vectors to transform the original vector space. A third approach is Probabilistic Latent Semantic Indexing (PLSI) [8], which is based on a latent variable model for general co-occurrence data, which associates an unobserved latent class variable with each observation. The model maps the documents to a reduced vector space, the latent semantic space. The number of latent factors will be much smaller than the number of words and the factors act as prediction variables for words. The factors are obtained using a generalization of the Expectation Maximization algorithm.

The previous models are seen as a transformation of the original vector space to capture semantic relations among terms and deal with synonymy. They are known as word space models. The term word space model was introduced by Hinrich Schütze [9], to capture approaches that represent semantic information about words derived from co-occurrence data. These approaches rely on the distributional hypothesis formulated by the linguist Zellig Harris, which states that terms with similar distributional patterns tend to have the same meaning [10].

Often, the word space model is seen as a spatial representation of word meaning. The central idea is that semantic similarity can be represented as proximity, where semantically related words are close, and unrelated ones are distant in an *n*-dimensional space [11]. The traditional word space methodology produces the high dimensional vector space storing co-occurrence data in a matrix M known as co-occurrence matrix, where each row M_w represents a word and each column M_c a context (a document or other word). The cell M_{wc} records the co-occurrence of word w in the context c. The M_w rows are vectors, whose size depends on the number of contexts, and are known as 'context vectors' of the words because they represent the contexts in which words are present. Therefore, consistent with the distribution hypothesis, it is possible to compute semantic similarity between words by comparing their context vector using existing vector similarity measures.

We explored two different processes to generate context vectors: document occurrence representation, DOR, and term co-occurrence representation, TCOR. Both representations are based on the distributional hypothesis, and can be used to represent the meaning of a document as a bag of concepts (BoC), i.e. the sum of the meanings (context vectors) of its terms [23].

In DOR, term t_j is represented as a vector $\vec{t}_j = (w_{1j}, w_{2j,} ..., w_{mj})$ of context weights, where m is the cardinality of the document collection, and w_{kj} represents the contribution of context k to the specification of the semantics of term t_j. In this representation, the meaning of a term is considered as the bag of contexts in which it occurs. In this case, contexts are defined as entire documents.

In TCOR, on the other hand, term t_j is represented as a vector of term weights, $\vec{t}_j = (w_{1j}, w_{2j,} ..., w_{mj})$ where m is the cardinality of the vocabulary and w_{kj} represents the contribution of term k to the specification of the semantics of term t_j. The meaning of a term is viewed as the bag of terms with which it co-occurs in some context [12].

Any algorithm that implements a word space model has to handle the potentially high dimensionality of the context vectors, to avoid affecting its scalability and efficiency. It is crucial to get a balance between the amount of co-occurrence data used and the size of the co-occurrence matrix, which serves as a basis for generating the context vectors. Notably, the majority of the cells in the co-occurrence matrix will be zero given that most words occur in limited contexts.

These problems of very high dimensionality and data sparseness have been approached using dimension reduction techniques such as SVD. However, these techniques are computationally expensive in terms of memory and processing time. Moreover, they require that first the huge co-occurrence matrix be built and then reduced. This process is repeated every time that new data is added and has to be completed before any processing can begin, which represents a serious deficiency given the computational cost of computing SVD on large matrices.

As an alternative to models that use SVD, there is a word space model called Random Indexing [13], which presents an efficient, scalable and incremental method for building context vectors. Here we explore the use of Random Indexing (RI) to represent documents and queries for IR.

In addition to this indexing, we explore the use of more complex linguistic structures (e.g., noun phrases) to index and retrieve documents [14, 15, 16, 17]. These have shown to be more effective than pure VSM in some circumstances; however, they have not achieved definitive success. These approaches extract noun phrases, and subsequently include them as new VSM terms. We explore a different representation of such structures, which uses a special kind of vector binding (called holographic reduced representations (HRRs) [18]) to reflect text structure and distribute syntactic information across the document representation. HRRs use circular convolution to associate items, which are represented by vectors. The use of circular convolution to represent text structure for retrieving information has been explored in [19] showing a slight precision improvement compared to VSM. A processing text task where HRRs have been used together with Random Indexing is text classification where they have shown improvement in certain circumstances [20, 21].

The remainder of this report is organized as follow. In Section 2 we review the Random Indexing methodology. Section 3 introduces the concept of Holographic Reduced Representations (HRRs) and presents how to use HRRs to add text structural information to document representations. Section 4 shows some experimental results that have been obtained in the CACM collection. Finally, Section 5 concludes the paper and gives some directions for future work.

2. Random Indexing

Random Indexing is a vector space methodology that accumulates context vectors for words based on co-occurrence data. The technique can be described as: a) First, a unique random representation known as *index vector* is assigned to each context (document or word). Index vectors are binary vectors with a small number of non-zero elements, which are either +1 or -1, with equal amounts of both. For example, if the index vectors have twenty non-zero elements in a 1024-dimensional vector space, they have ten +1s and ten -1s. Index vectors serve as indices or labels for words or documents; b) Index vectors are used to produce context vectors by scanning through the text and every time a given word occurs in a context, the context index vector is added to the word context vector. Therefore, a word is represented by a context vector that contains traces of every context, i.e., word or document, that the word has co-occurred with or in.

Traditional vector space methods represent context vectors in a co-occurrence matrix T of order w x c, where rows T_w represent words and columns T_c contexts. Here each row can be interpreted as a c-dimensional context vector $\vec{w}$ for the word w. In contrast to these methods, random indexing replaces the matrix T by a context matrix R of order w x k being $k << c$. Every row R_i is the k-dimensional context vector for word i.

Random Indexing can produce the standard co-occurrence matrix T of order w x c if unary c-dimensional index vectors are used. Unary vectors are orthogonal, but the random index vectors are only nearly orthogonal. This means that if the matrix R_{wxk} is formed using the context vectors produced by random indexing, the matrix will be an approximation of the standard co-occurrence matrix T_{wxc} because their corresponding rows are similar or dissimilar to the same degree, but with $k << c$. Therefore, since there, exists a much larger number of nearly orthogonal than truly orthogonal directions in a high-dimensional space (Hecht-Nielsen, 1994), choosing random directions result sufficiently close to orthogonality to provide an approximation of the unary vectors. The amount of noise introduced by choosing random directions is so small that it does not have any significant effect on the similarity relations between vectors, which means that the traditional frequency matrix and the random indexing matrix contain approximately the same information. Hence, defining a matrix U of order c x k, whose row U_i is the k-dimensional index vector for context i, the following relations holds:

$$R_{wxk} = T_{wxc}\, U_{cxk}$$

This means, as it was explained, that the random indexing context matrix R contains the same information that the standard co-occurrence matrix T multiplied with the random matrix U, where $U\, U^T$ is approximately I, the identity matrix.

3. Holographic Reduced Representation

Over the past two decades connectionist models have received attention as a means to represent higher-level cognitive activities such as language processing. A connectionist model is a network of processing units that communicate each other through weighted links. Units compute some simple function of the data they receive; the value obtained by the function is the state or activation of the unit and is the message that is passed to other units. There are two distinct locations for representing knowledge (data) in a connectionist model: the activation value of the units and the weights on the links among units. Links can be used to encode domain knowledge as constraints on solutions to a task or as transformations between input and output patterns. In contrast there are diverse representation schemes for data structures, which are used to store activations over the units of the network.

Representations are important in connectionist models because they determine what the system can or cannot compute. A system cannot perform complex reasoning tasks if it cannot even represent the items involved in the task. Two types of representations exist in connectionist models: localist, which uses particular units to represent each concepts (objects, words, relationships, features), and distributed, in which a unit participates in the representation of many different concepts.

Distributed representations have the following advantages and disadvantages: efficient use of representational resources, analogical representation (i.e. similar objects will have similar representations), and continuity (i.e. representations are in a continuous vector space); but distributed representations have difficulty representing arbitrary associations, hierarchical structure, and identifying why a particular object has a specific representational pattern.

HRRs use distributed representation, having in addition the advantage that they permit representation of structure using a circular convolution operator to bind terms, without increasing vector dimensionality. Circular convolution operator ($\otimes$) binds two vectors $\vec{x} = (x_0, x_1, ..., x_{n-1})$ and $\vec{y} = (y_0, y_1, ..., y_{n-1})$ to give $\vec{z} = (z_0, z_1, ..., z_{n-1})$ where $\vec{z} = \vec{x} \otimes \vec{y}$ is defined as:

$$z_i = \sum_{k=0}^{n-1} x_k y_{i-k} \qquad i = 0 \text{ to } n-1 \text{ (subscripts are module-}n\text{)}$$

It can be thought as a multiplication operator for vectors. It has properties in common with scalar and matrix multiplication. It is commutative, associative, and bilinear. There is an identity vector, and a zero vector. A finite-dimensional vector space over the real numbers with circular convolution and the usual definition of scalar multiplication and vector addition form a commutative linear algebra system, so all the rules that apply to scalar algebra also apply to this algebra. As a result, it is not complicated to manipulate expressions containing additions, convolutions, and scalar multiplications [22].

We adopt HRRs to build a text representation scheme in which part of the document syntax can be captured and can help to improve retrieval performance. To define an HRR document representation, the following steps are performed:

1. All documents are indexed using random indexing
2. For each textual relation in a document, the index vectors of the words that participate in the relation are bound to their role identifier vectors (an HRR).
3. The tf × idf-weighted sum of the resulting vectors is taken to obtain a single HRR vector representing the textual relation.
4. HRRs of the textual relations are added to the document vector in order to obtain a single HRR vector representing the document, which is normalized.

In detail, suppose a relation *R* (r_1, r_2) exists between terms r_1 and r_2. Each term plays a different role in this structure (e.g. noun phrase right/noun phrase left, or subject/verb, or object/adjective, etc.). To encode these roles two special vectors are needed: $role_1, role_2$. Then, the relation vector is:

$$\vec{R} = (rol\vec{e}_1 \otimes r_1 + rol\vec{e}_2 \otimes \vec{r}_2).$$

Given a document *D*, with terms $t_1, t_2, \ldots, tx_1, ty_1, \ldots, tx_2, ty_2, \ldots, tx_n, ty_n, \ldots, t_n$, and different relations R_1, R_2 among terms $tx_1, ty_1; tx_2, ty_2$, respectively, its vector will be built as:

$$\vec{D} = \left\langle t_1 + t_2 + ... + t_n + (role_a \otimes t_{x1} + role_b \otimes t_{x2}) + (role_c \otimes t_{y1} + role_d \otimes t_{y2}) \right\rangle$$

where $\langle\ \rangle$ denotes a normalized vector. Queries will be represented in similar way.

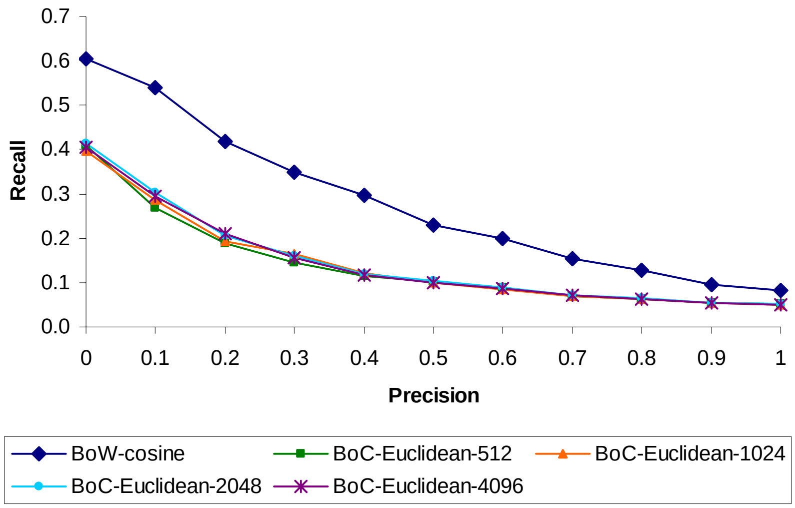

Figure 1. Bag of concepts using DOR, and varying vector dimension

4. Experiments

The proposed representation was applied to the CACM collection with 3204 documents and 64 queries. This collection is well-known and relatively small to initially test our representation. The traditional vector space model (VSM) was used as a baseline, implemented using tf.idf weighting scheme and a cosine measure to determine vector similarity. We compared this to our model, which used random indexing, Euclidean distance as similarity measure between vectors, and the same weighting scheme. Our first experiment was aimed to test the feasibility of the representation. We used only single terms (i.e. applied BoC) and carried out several investigations, including the effects of dimensionality, limited vocabulary, and context definition.

4.1 Dimensionality Investigation

These experiments looked into how the precision in retrieval is affected by BoC vector dimensionality. The exploration was made using dimensions: 512, 1024, 2048, and 4096. In these experiments, the context vectors were created using the DOR approach. Figure 1 shows the Precision/Recall curves for the results obtained as well as the BoW curve. Precision was calculated at standard recall values averaged for the number of queries. The precision obtained for all vector dimensions was very far from the BoW curve going from -50.37% in average for 512 vectors to -48.44% for 2048 down.

4.2 Limited Vocabulary Investigation

These experiments investigated how the precision is affected by limiting the vocabulary. The stop words, extremely common words as articles, prepositions, conjunctions, were removed from the previous representations. The results are shown in figure 2. Although in all cases the precision was higher, it continued below the BoW curve. The best results were obtained for the 1024 dimension, which were -30.98% in average below the BoW curve.

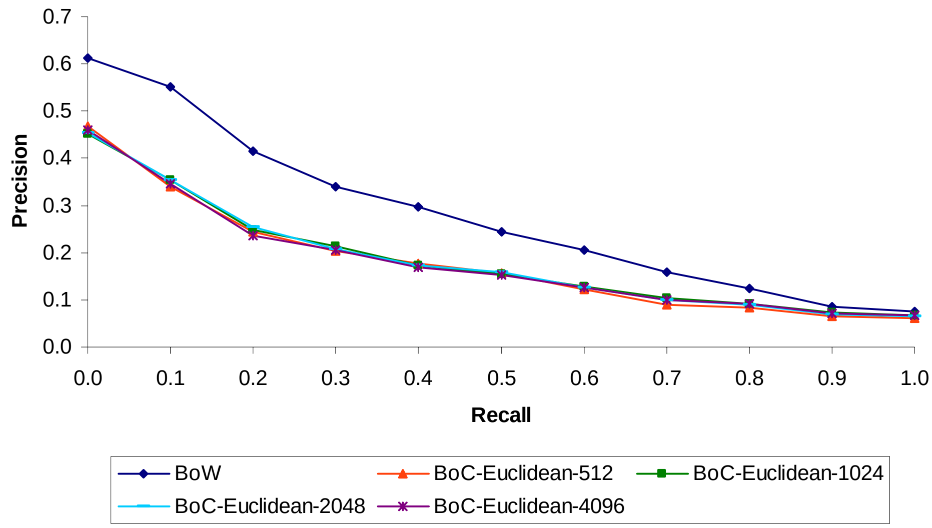

Figure 2. Bag of concepts using DOR different dimensions without stop words

4.3 Context Definition Investigation

These experiments investigated how the precision is affected by the selected methodology for creating context vectors. We explored the representations created with DOR, TCOR, and the index vectors to produce document representations using BoC. The outcomes were also compared to those obtained with BoW representation. The dimensionality of the vector space was 1024. Figure 3 shows that using index vectors as context vectors, the precision rose to be close to the BoW precision just -9.36% in average below. It is important to point out that the dimension used in this method was 1024 while the BoW dimension was 6867.

Faced with evidence that using index vectors as context vectors caused the precision to rise, we further explored the dimensionality impact in this circumstance. Table 1 shows that the highest precision was obtained with vectors dimension of 5000, only -0.45% below BoW, followed by that obtained with 4096 dimension, down -1.09%. With vector dimension higher than 5600 the precision decreased. Since HRR dimensions must always be a power of two, 4096 vectors were considered suitable for subsequent experiments.

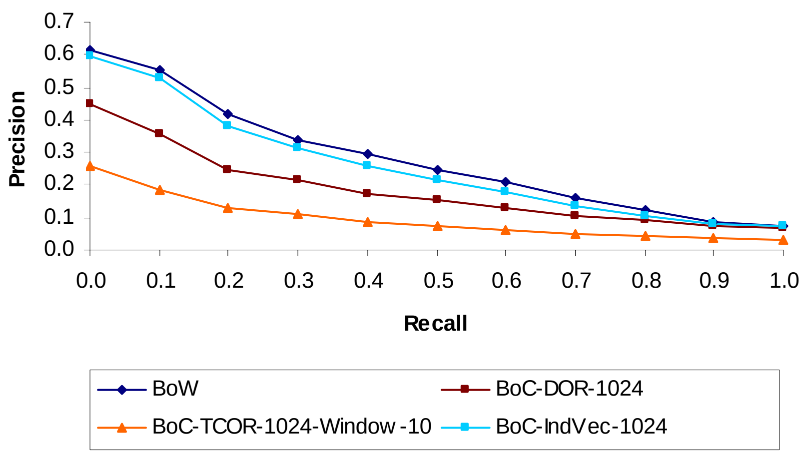

Figure 3. Bag of concepts with DOR, TCOR, and Index Vectors

Recall	BoW	BoC-InVec-1024	% Diff.	BoC-InVec-2048	% Diff.	BoC-InVec-4096	% Diff.	BoC-InVec-5000	% Diff.	BoC-InVec-5600	% Diff.
0	0.613	0.5962	-2.72%	0.5967	-2.64%	0.6228	1.62%	0.6145	0.26%	0.6203	1.21%
0.1	0.552	0.5286	-4.27%	0.5398	-2.25%	0.5561	0.71%	0.5596	1.34%	0.5550	0.51%
0.2	0.415	0.3782	-8.85%	0.3940	-5.04%	0.4016	-3.21%	0.4122	-0.65%	0.4196	1.13%
0.3	0.341	0.3141	-7.81%	0.3249	-4.64%	0.3378	-0.85%	0.3388	-0.56%	0.3402	-0.15%
0.4	0.297	0.2601	-12.37%	0.2730	-8.02%	0.2822	-4.92%	0.2792	-5.93%	0.2837	-4.41%
0.5	0.243	0.2153	-11.47%	0.2134	-12.25%	0.2315	-4.81%	0.2291	-5.80%	0.2376	-2.30%
0.6	0.206	0.1758	-14.58%	0.1830	-11.08%	0.2024	-1.65%	0.2030	-1.36%	0.2083	1.21%
0.7	0.158	0.1321	-16.39%	0.1344	-14.94%	0.1476	-6.58%	0.1526	-3.42%	0.1547	-2.09%
0.8	0.125	0.1074	-14.08%	0.1087	-13.04%	0.1218	-2.56%	0.122o	-2.40%	0.1202	-3.84%
0.9	0.086	0.0781	-9.40%	0.0786	-8.82%	0.0873	1.28%	0.0895	3.83%	0.0855	-0.81%
1.0	0.075	0.0738	-1.07%	0.0735	-1.47%	0.0813	8.98%	0.0819	9.79%	0.0775	3.89%
Average			-9.36%		-7.65%		-1.09%		-0.45%		-0.51%

Table 1. Precision representing documents using index vectors as context vectors

5. Conclusion and Future work

In this report, we have presented a proposal for representing documents and queries using random indexing. The presented experiments show that this approach is feasible, and able to support the retrieval of information, while reducing the vector dimensionality when compared to the classical vector model. However, performance was generally below that of the classical approach (except when index vectors were used directly). CACM is a collection whose queries ask for very specific information. We observed that the BoC representation performs better on queries that have a high number of relevant documents. So for, queries that are very specific (i.e., have few relevant documents) BoW works better (perhaps relying on keywords). We will explore this hypothesis through the use of another collection with more documents per query. This work will also focus on extracting relations among terms, and using these relations to create HRRs that further enrich the document representations. The relations to be extracted include: noun phrases, subject-verb, object-verb, and adverb-verb. An appropriate mechanism to incorporate the relations (HHRs) without affecting the precision and a suitable weighting scheme for them has yet to be defined.

References

1. Salton, G.: Automatic text processing: the transformation, analysis and retrieval of information by computer. Addison Wesley, 1989.
2. Salton, G., McGill, M.: Introduction to modern information retrieval. New York, McGraw Hill, 1983.
3. Bollmann-Sdorra, P., Raghavan, V.V.: On the necessity of term dependence in a query space for weighted retrieval. In: Journal of the American Society for Information Science, Vol.49, pp.1161–1168, 1998.
4. Raghavan, V.V., Wong, S.K.M.: A critical analysis of vector space model for information retrieval. In: Journal of the American Society for Information Science, Vol.37, pp.279–287, 1986.
5. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R.: Indexing by latent semantic analysis. In: Journal of the American Society for Information Science, Vol.41, pp.391–407, 1990.
6. Wong, S. K. M., Ziarko, W., Wong, P. C. N.: Generalized vector space model in information retrieval. In: Proceedings of the 8th Annual International ACM-SIGIR Conference, pp.18-25, 1985.
7. Wong, S.K.M., Ziarko, W., Raghavan, V.V., Wong, P.C.N.: On modeling of information retrieval concepts in vector spaces. In: ACM Transactions on Database Systems, Vol. 12, pp.299–321, 1987.
8. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of the 22st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 50–57, 1999.

9. Schütze, H.: Word space. In: Advances in Neural Information Processing Systems, Ed. S. Hanson, J. Cowan , C. Giles, Morgan Kaufmann Publishers Inc., pp. 895-902, 1993.
10. Harris Z.: Mathematical structures of language. John Wiley & Sons, New York, US, 1968.
11. Sahlgren M.: The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high dimensional vector spaces. PhD thesis, Stockholm University, Stockholm, Sweden, 2006
12. Lavelli A., Sebastiani F., Zanoli R.: Distributional term representations: an experimental comparison. In: CIKM '04: Proceedings of the thirteenth ACM conference on information and knowledge management, ACM Press, pp. 615–624, 2004.
13. Sahlgren M.: An Introduction to Random Indexing, In: Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, 2005.
14. Lewis D., Sparck K.: Natural Language Processing for Information Retrieval. In: Communications ACM 39, pp. 92-101, 1996.
15. Mitra M., Buckley C., Singhal A., Cardie C.: An Analysis of Statistical and Syntactic Phrases. In: Proceedings of RIAO-97, 5th International Conference, pp. 200-214.
16. Evans D., Zhai C.: Noun-phrase Analysis in Unrestricted Text for Information Retrieval. In: Proceedings of the 34th Annual Meeting on Association for Computational Linguistics, pp. 17-24, 1996.
17. Plate T.A.: Analogy Retrieval and Processing with Distributed Vector Representation, Victoria University of Wellington, Computer Science, Technical Report CS-TR-98-4, 16 p.
19. Carrillo M., López-López A.: Towards an Enhanced Vector Model to Encode Textual Relations: Experiments Retrieving Information. In: IFIP International Federation for Information Processing, Vol.276, Artificial Intelligence in Theory and Practice II, ed. M. Bramer, pp. 383-392, 2008.
20. Fishbein J.M. and Eliasmith C.: Integrating structure and meaning: A new method for encoding structure for text classification. In: Advances in Information Retrieval: Proceedings of the 30th European Conference on IR Research, Ed. C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, R. W. White, Vol. 4956 of Lecture Notes in Computer Science, pp. 514–521, 2008.
21. Fishbein J.M. and Eliasmith C.: Methods for augmenting semantic models with structural information for text classification. In: Advances in Information Retrieval: Proceedings of the 30th European Conference on IR Research, Ed. C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, R. W. White, Vol. 4956 of Lecture Notes in Computer Science, pp. 575–579, 2008.
22. Plate T.A. Holographic Reduced Representation: Distributed representation for cognitive structures. CSLI Publications, 2003.
23. Sahlgren M. and Cöster R.: Using Bag-of-Concepts to Improve the Performance of Support Vector Machines in Text Categorization. In: Proceedings of the 20th International Conference on Computational Linguistics, pages 487– 493, 2004.