Title	Random Indexing Explained with High Probability
Author(s)	QasemiZadeh, Behrang
Publication Date	2015
Publication Information	QasemiZadeh, Behrang; (2015) Random Indexing Explained with High Probability . In: Vaclav Matousek et al eds. 18th International Conference, TSD 2015 PLZEŇ, CZECH REPUBLIC,
Link to publisher's version	http://atmykitchen.info/publications/random-indexing_tsd-long
Item record	http://hdl.handle.net/10379/5087

Random Indexing Explained with High Probability

Behrang QasemiZadeh

National University of Ireland, Galway, Ireland

University of Passau, Germany

`{behrang.qasemizadeh}@insight-centre.org`

Abstract. Random indexing (RI) is an incremental method for constructing a vector space model (VSM) with a reduced dimensionality. Previously, the method has been justified using the mathematical framework of Kanerva's sparse distributed memory. This justification, although intuitively plausible, fails to provide the information that is required to set the parameters of the method. In order to suggest criteria for the method's parameters, the RI method is revisited and described using the principles of linear algebra and sparse random projections in Euclidean spaces. These simple mathematics are then employed to suggest criteria for setting the method's parameters and to explain their influence on the estimated distances in the RI-constructed VSMs. The empirical results observed in an evaluation are reported to support the suggested guidelines in the paper.[1]

1   Introduction

In order to model any aspect of language, data-driven methods of natural language processing exploit patterns of co-occurrences. For example, distributional semantic models collect patterns of co-occurrences and investigate similarities in these patterns in order to quantify meanings. Vector spaces are mathematically well-defined models that are often employed to serve this purpose [2].

---

[1] The first three pages previously appeared in [1].

In a vector space model (VSM), each element $\vec{s}_i$ of its standard basis—informally, each dimension of the VSM—represents a contextual element. Given n context elements, linguistic entities are expressed using vectors $\vec{v}$ as linear combinations of $\vec{s}_i$ and scalars $\alpha_i \in \mathbb{R}$ such that $\vec{v} = \alpha_1\vec{s}_1 + \cdots + \alpha_n\vec{s}_n$. The value of α_i is acquired from the frequency of the co-occurrences of the entity that $\vec{v}$ represents and the context element that $\vec{s}_i$ represents. Therefore, the values assigned to the coordinates of a vector—that is, α_i—exhibit the correlation of an entity and context elements in an n-dimensional real vector space $\mathbb{R}^n$. In this VSM, a distance function, therefore, is employed in order to discover similarities. Amongst several choices of distance metrics, the Euclidean distance is an innate choice. A VSM is endowed with the ℓ_2 norm to estimate distances between vectors, which is accordingly called a Euclidean VSM (denoted by $\mathbb{E}^n$). Salton et al.'s classic document-by-term model is, perhaps, the most familiar example of the methodology described above [3].

In distributional methods of text analysis, as the number of entities in a VSM increases, the number of context elements employed for capturing similarities between them surges. As a result, high-dimensional vectors, in which most elements are zero, represent entities. However, the proportional impact of context elements on similarities lessens when their number increases. It becomes difficult to distinguish similarities between vectors unless the values assigned to context elements are considerably different [4]. Moreover, the high dimensionality of vectors hinders the ability to compute distances with high performance. This results in setbacks known as the *curse of dimensionality*, often tackled using a *dimensionality reduction* technique.

Dimensionality reduction can be achieved using a number of methods as an auxiliary process followed by the construction of a VSM. This process improves the computational performance by reducing the number of context elements employed for the construction of a VSM. In its simple form, dimension reduction can be performed by choosing a subset of context elements using a heuristic-based *selection process*. That is, a number of context elements that account for the most discriminative information in VSM are chosen using a heuristic such as a statistical weight threshold. Alternatively, a *transformation* method can be employed. This process maps $\mathbb{R}^n$ onto $\mathbb{R}^m$, $m \ll n$, in which $\mathbb{R}^m$ is the best approximation of $\mathbb{R}^n$ in a *sense*. For example, the well-known latent semantic analysis

method employs singular value decomposition (SVD) truncation, in which $\mathbb{R}^m$ gives the best approximation of the Euclidean distances in $\mathbb{R}^n$ [5].

A number of factors hamper the use of these dimension reduction methods. Firstly, a VSM at the original high dimension must be constructed. The VSM's dimension is then reduced in an independent process. Hence, the VSM at a reduced dimensionality is available for processing only after the whole sequence of these processes. Construction of the VSM at its original dimension is computationally expensive, and a delay in access to the VSM at the reduced dimension is not desirable. Secondly, reducing the dimension of vectors using the methods listed above is resource intensive. For instance, SVD truncation demands a process of the time complexity $O(n^2m)$ and space complexity $O(n^2)$.[2] Similarly, depending on the employed heuristic, a selection process can be resource intensive too—for example, frequencies often need to be sorted by some criteria. Last but not least, these methods are *data-sensitive*: if the structure of the data being analysed changes—that is, if either the entities or context elements are updated—the dimensionality reduction process is required to be repeated and reapplied to the whole VSM in order to reflect the updates. As a result, these methods may not be desirable in several applications, particularly when dealing with frequently updated big text-data. Random projections are mathematical tools that are employed to implement alternative dimensionality reduction techniques to alleviate the problems listed above.

In the remainder of this paper, Section 2 describes the use of random projections (RPs) in Euclidean spaces, which consequently arrives at the well-known random indexing (RI) technique. Section 3 articulates the outcome of this mathematical interpretation. To support the theoretical discussion, empirical results are reported in Section 4. Section 5 concludes this paper.

2   Random Projections in Euclidean Spaces

In Euclidean spaces, RPs are elucidated using the Johnson and Lindenstrauss lemma (JL lemma) [7]. Given an ϵ, $0 < \epsilon < 1$, the JL lemma states that for any set of p vectors in an

---

[2] However, the use of incremental techniques may relax these requirements to an extent; for example, see [6].

$\mathbb{E}^n$, there exists a mapping onto an $\mathbb{E}^m$, for $m \geq m_0 = \mathrm{O}(\frac{\log p}{\epsilon^2})$, that does not distort the distances between any pair of vectors, with high probability, by a factor more than $1 \pm \epsilon$. This mapping is given by:

$$\mathbf{M}'_{p \times m} = \mathbf{M}_{p \times n} \mathbf{R}_{n \times m}, \, m \ll p, n, \tag{1}$$

where $\mathbf{R}_{n \times m}$ is called the RP matrix, and $\mathbf{M}_{p \times n}$ and $\mathbf{M}'_{p \times m}$ denote the p vectors in $\mathbb{E}^n$ and $\mathbb{E}^m$, respectively. According to the JL lemma, if the distance between any pair of vectors $\vec{v}$ and $\vec{u}$ in $\mathbf{M}$ is given by the $d_{\mathrm{Euc}}(\vec{v}, \vec{u})$, and their distance in $\mathbf{M}'$ is given by $d'_{\mathrm{Euc}}(\mathbf{v}, \mathbf{u})$, then there exists an $\mathbf{R}$ such that $(1 - \epsilon)d'_{\mathrm{Euc}}(\mathbf{v}, \mathbf{u}) \leq d_{\mathrm{Euc}}(\mathbf{v}, \mathbf{u}) \leq (1 + \epsilon)d'_{\mathrm{Euc}}(\mathbf{v}, \mathbf{u}).[3] Accordingly, instead of the original high-dimensional $\mathbb{E}^n$ and at the expense of a negligible amount of error ϵ, the distance between $\vec{v}$ and $\vec{u}$ can be calculated in $\mathbb{E}^m$ to reduce the computational cost of processes.

The JL lemma does not specify $\mathbf{R}$. Establishing a random matrix $\mathbf{R}$ is therefore the most important design decision when using RPs. In [7], the lemma was proved using an orthogonal projection. Subsequent studies simplified the original proof that resulted in projection techniques with enhanced computational efficiency (see [8] for references). Recently, it has been shown that a sparse $\mathbf{R}$, whose elements r_{ij} are defined as:

$$r_{ij} = \sqrt{s} \begin{cases} -1 & \text{with probability } \frac{1}{2s} \\ 0 & \text{with probability } 1 - \frac{1}{s}, \\ 1 & \text{with probability } \frac{1}{2s} \end{cases} \tag{2}$$

for $s \in \{1, 3\}$, results in a mapping that also satisfies the JL lemma [9]. Subsequent research showed that $\mathbf{R}$ can be constructed from even sparser vectors than those suggested in [9]. In [10], it is proved that in a mapping of an n-dimensional real vector space by a sparse $\mathbf{R}$, the JL lemma holds as long as $s = \mathrm{O}(n)$, for example, $s = \sqrt{n}$ or even $s = \frac{n}{\log(n)}$. The sparseness of $\mathbf{R}$ consequently enhances the time and space complexity of the method by the factor $\frac{1}{s}$.

Another benefit when computing $\mathbf{M}'$ is obtained using the linearity of matrix multiplication. As stated earlier, each vector $\vec{v}_{e_i}$ in $\mathbb{E}^n$ (i.e., the ith row of $\mathbf{M}$) is given by a

---

[3] In addition, the lemma states that this mapping can be found in randomized polynomial time.

linear combination of the basis vectors $\vec{v}_{e_i} = w_{i1}\vec{s}_{c_1} + \cdots + w_{in}\vec{s}_{c_n}$ ($i \leq p$ and $j \leq n$). By the basic properties of the matrix multiplication, the projection of $\vec{v}_{e_i}$ in $\mathbf{M'}$ is given by $\vec{v}'_{e_i} = \vec{v}_{e_i}\mathbf{R} = w_{i1}\vec{s}_{c_1}\mathbf{R} + \cdots + w_{in}\vec{s}_{c_n}\mathbf{R}$. In turn, since by definition all the elements of $\vec{s}_{c_k}$ are zero except the kth element (i.e., 1), $\vec{v}'_{e_i}$ can be written as:

$$\vec{v}'_{e_i} = w_{i1}\vec{r}_1 + \cdots + w_{in}\vec{r}_n, \tag{3}$$

where $\vec{r}_j$ is the jth row of $\mathbf{R}$. Equation 3 means that row vectors $\mathbf{v}'_{e_i}$, thus $\mathbf{M'}$, can be computed directly without necessarily constructing the whole matrix $\mathbf{M}$. The jth row of $\mathbf{R}_{n \times m}$ represents a context element in the original VSM that is located at the jth column of $\mathbf{M}_{p \times n}$. Therefore, an entity at a reduced dimension can be computed directly by accumulating the row vectors of $\mathbf{R}$ that represent the context elements that co-occur with the entity.

The explanations above result in a two-step procedure similar to the one suggested earlier as the RI technique [11][12]: the construction of (a) *index vectors* and (b) *context vectors*. In the first step, each context element is assigned *exactly* to one *index vector*. Sahlgren [12] indicates that index vectors are high-dimensional, randomly generated vectors, in which most of the elements are set to 0 and only *a few* to 1 and -1. In the second step, the construction of *context vectors*, each target entity is assigned to a vector of which all elements are zero and that has the same dimension as the index vectors. For each occurrence of an entity (represented by $\vec{v}_{e_i}$) and a context element (represented by $\vec{r}_{c_k}$), the context vector is accumulated by the index vector (i.e., $\vec{v}_{e_i} = \vec{v}_{e_i} + \vec{r}_{c_k}$). The result is a vector space model constructed directly at reduced dimension. As can be understood, the first step of RI is equivalent to constructing the random projection matrix $\mathbf{R}$, whose elements are given by Equation 2. Each index vector is a row of the random projection matrix $\mathbf{R}$. The second step of RI deals with computing $\mathbf{M'}$. Each context vector is a row of $\mathbf{M'}$, which is computed by the iterative process justified in Equation 3.

3   The Significance of the Proposed Mathematical Justification

In contrast to previous research in which the RI's parameters were left to be decided through experiments (e.g., see [13,14]), one can leverage the adopted mathematical frame-

work to provide a guideline for setting the parameters of RI. In an RI-constructed VSM at reduced dimension m (i.e., $\mathbb{E}^m$), the degree of the preservation of distances in $\mathbb{E}^n$ and $\mathbb{E}^m$ is determined by the number of vectors in the model and the value of m. If the number of vectors is fixed, then the larger m is, the better the Euclidean distances are preserved at the reduced dimension m. In other words, the probability of preserving the pairwise distances increases as m increases. Hence, m can be seen as the capacity of an RI-constructed VSM for accommodating new entities. Compared to $m = 4000$ suggested in [11] or $m = 1800$ in [12], depending on the number of entities that are modelled in an experiment, m can be set to a smaller value, such as 400.

Based on the proofs in [10], when embedding $\mathbb{E}^n$ into $\mathbb{E}^m$, the JL lemma holds as long as s in Equation 2 is $\mathrm{O}(n)$. In text processing applications, the number of context elements (i.e., n) is often very large. When using RI, therefore, even a careful choice such as $s = \sqrt{n}$ in Equation 2 results in highly sparse index vectors. Hence, by setting only two or four non-zero elements in index vectors, distances in the RI-constructed $\mathbb{E}^m$ resemble distances in $\mathbb{E}^n$. If the dimension of index vectors (i.e., m) is fixed, then increasing the number of non-zero elements in index vectors causes additional distortions in pairwise distances. For index vectors of fixed dimensionality m, if the number of non-zero elements increases, then the probability of the orthogonality between index vectors decreases; hence, it stimulates distortions in pairwise distances (see Fig. 1)—although in some applications, distortions in pairwise distances can be beneficial.

4   Experimental Results

In order to show the influence of the RI's parameters on the ability of the method to preserve pairwise Euclidean distances, instead of a task-specific evaluation, an intrinsic evaluation is suggested.

In the reported experiments, a subset of Wikipedia articles chosen randomly from WaCkypedia (a 2009 dump of the English Wikipedia [15]). A document-by-term VSM at its original high dimension is first constructed from a set of $10,000$ articles (shown by D). A pre-processing—that is, white-space tokenisation followed by removing non-alphabetic
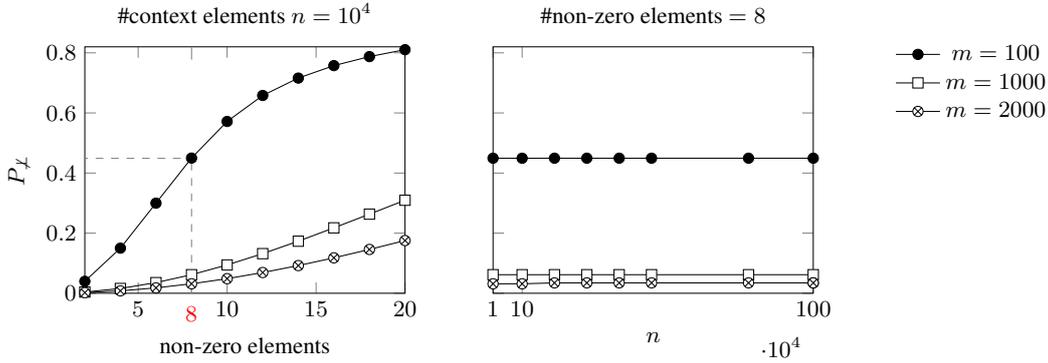
Fig. 1. Orthogonality of index vectors: the y-axis shows the proportion of non-orthogonal pairs of index vectors (denoted by $P_{\not\perp}$) for sets of index vectors of various dimensions obtained in a simulation. For index vectors of the fixed size $n = 10^4$, the left figure shows the changes of $P_{\not\perp}$ when the number of non-zero elements increases. The right figure shows $P_{\not\perp}$ when the number of non-zero elements is fixed to 8; however, the number of index vector n increases. As shown in the figure, $P_{\not\perp}$ remains constant independently of n.

tokens—of documents in D results in a vocabulary of 192,117 terms. Each document in D is represented by a high-dimensional vector; each dimension represents an entry from the obtained vocabulary. Therefore, the constructed VSM using this *one-dimension-per-context-element* method has a dimensionality of $n = 192{,}117$.[4]

To keep the experiments a manageable size, each document d in D is randomly grouped by another nine documents from D, which consequently gives 10,000 sets of a set of ten documents. Using the constructed n-dimensional ($n = 192{,}117$) VSM, for each set of documents, the Euclidean distances between d and the remaining nine documents in the set are computed. Subsequently, these nine documents are sorted by their distance from d to obtain an ordered set of documents. This procedure thus results in 10,000 ordered sets of nine documents (the same steps are repeated for computing the cosine similarities).

The procedure described above is repeated, however, by calculating distances in VSMs that are constructed using the RI method. Each term in the vocabulary is assigned to an m-dimensional index vector and each document to a context vector. Context vectors are updated by accumulating index vectors to reflect the co-occurrences of documents and terms. Subsequently, the obtained context vectors are used to estimate the Euclidean dis-

---

[4] The frequencies of terms in documents are used as weights in corresponding vectors.
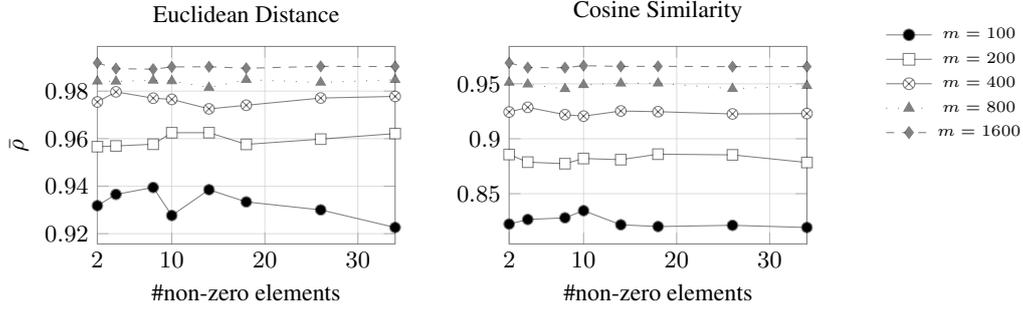
Fig. 2. Correlation between the ℓ_2-normed measures in the original high-dimensional VSM and RI-constructed VSMs: $\bar{\rho}$ shows the average of the Spearman's rank correlation between the ordered sets of documents that are obtained by computing in the original high-dimensional VSM and the RI-constructed VSMs. Results are shown for both Euclidean distances and the cosine similarities when parameters of the RI method are set to different values. The random baseline obtained in experiments is -0.002 (i.e., as expected, almost zero).

tances and the cosine similarities between documents. The estimated distances are then used to create the ordered sets of documents, exactly as explained above. This process is repeated several times when the parameters of RI—that is, the dimension m and the number of non-zero elements in index vectors—are set differently.

It is expected that the relative Euclidean distances as well as the cosine similarities between documents in the RI-constructed VSMs are the same as in the original high-dimensional VSM.[5] Hence, the ordered sets of documents obtained from the estimated distances in the RI-constructed VSMs must be identical to the corresponding sets that are derived using the computed distances in the original high-dimensional VSM. For each RI-constructed VSM, therefore, the resulting ordered sets are compared with the obtained ordered sets from the original high-dimensional VSM using the Spearman's rank correlation coefficient measure (ρ). The average of ρ over the obtained sets of ordered sets of documents ($\bar{\rho}$) is reported to quantify the performance of RI with respect to its ability to preserve ℓ_2-normed distances when its parameters are set to different values: the closer $\bar{\rho}$ is to 1, the more similar the order of documents in an RI-constructed VSM and the original high-dimensional VSM.

---

[5] The preservation of the cosine similarities can be verified mathematically by expressing it as the Euclidean distance when the length of vectors is normalised to unity.
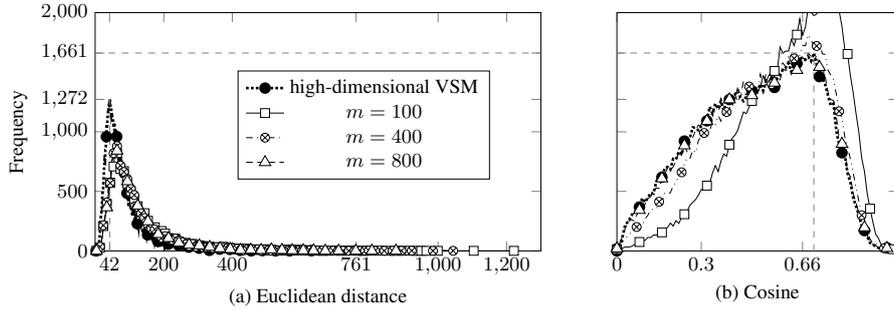
Fig. 3. A histogram of the distribution of (a) Euclidean distances and (b) cosine similarities between pairs of vectors in the original VSM of dimension 192,117 compared to the RI-constructed VSMs. For all values of m, the number of non-zero elements in index vectors is set to 2.

Figure 2 shows the obtained results. Since the original VSM is high dimensional and sparse, even for $m = 1600$, two non-zero elements per index vector are sufficient to construct a VSM that resembles relative distances between vectors in the original high-dimensional vector space. In addition, because only a small number of documents are modelled (i.e., $p = 10,000$), even for $m = 100$, the estimated distances in the RI-constructed VSM show a high correlation to the distances in the original vector space (i.e., $\bar{\rho} > 0.92$ for pairwise Euclidean distances and $\bar{\rho} > 0.82$ for the cosine similarity). As expected, the generated random baseline for $\bar{\rho}$ in Figure 2 is -0.002, that is, approximately 0. For $m = 1600$, the observed pairwise distances in the RI-constructed vector space are almost identical to the original vector space, that is, $\bar{\rho} > 0.99$ for Euclidean distances and $\bar{\rho} > 0.96$ for the cosine. Figure 3 compares the distribution of distances in the original high-dimensional VSM and the RI-constructed VSMs. As expected, when m increases, these distributions become more similar to each other.

5   Discussion

Random indexing—a well-known method for the incremental construction of VSMs—is revisited and justified using the theorems proved in [10]—that is, sparse random projections in Euclidean spaces. The results from an empirical experiment are shown to explain the method's behaviour with respect to its ability to preserve pairwise Euclidean distances.

Although a new method is not suggested, I would like to emphasise on several important outcomes of the description given in this paper.

Firstly, whereas the original delineation of the method did not provide a concrete guideline for setting the method's parameters, this paper ameliorated the previous two-step procedure with criteria for choosing the dimensionality as well as the proportion of zero and non-zero elements of index vectors. Secondly, the proposed understanding of the RI method helps us to discern its application domain. It is shown that the employed random projections by the RI method do not preserve distances other than ℓ_2 (e.g., see [16]). Hence, it is important to note that RI-constructed VSMs can only be used for estimating similarity measures that are derived from the ℓ_2 norm—for example, the Euclidean distance and the cosine similarity. This being the case, the use of RI-constructed VSMs for estimating city block distances—such as suggested in [17]—is not justified, at least mathematically.

Thirdly, the given understanding of the method helps one to generalise the RI method to normed spaces other than ℓ_2. This generalisation can be achieved using α-stable random projections—for example, as suggested in [18,19]—and by altering Equation 2. Simply put, altering a random projection matrix $\mathbf{R}$—hence, index vectors—so that it has an α-stable distribution[6] results in new techniques similar to RI, however, for estimating distances in ℓ_α-normed spaces (e.g., see [20,21,22]).[7]

Last but not least, the rationale given in this paper enables one to justify several proposed variations of the RI technique mathematically. Although these methods are based on plausible intuition, similar to RI, they lack theoretical justifications. For example, based on the description given in this paper, one can identify the method proposed in [23] as a variation of RI that employs Laplacian smoothing. This idea can be generalised for coordinating other major processes that are often involved when using VSMs.

---

[6] Note that RI uses a 2-stable projection; that is, $\mathbf{R}$ derived from Equation 2 has a standard asymptotic Gaussian distribution.

[7] In this case, new distance estimators are required.

Acknowledgement

References

1. QasemiZadeh, B.: Random indexing revisited. In: Natural Language Processing and Information Systems. Volume 9103 of Lecture Notes in Computer Science. Springer International Publishing (2015)

2. Turney, P.D., Pantel, P.: From frequency to meaning: vector space models of semantics. J. Artif. Int. Res. **37**(1) (January 2010) 141–188

3. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM **18**(11) (November 1975) 613–620

4. Beyer, K.S., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is "nearest neighbor" meaningful? In: Proceedings of the 7th International Conference on Database Theory. ICDT '99, London, UK, UK, Springer-Verlag (1999) 217–235

5. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science **41**(6) (1990) 391–407

6. Brand, M.: Fast low-rank modifications of the thin singular value decomposition. Linear Algebra and its Applications **415**(1) (2006) 20–30 Special Issue on Large Scale Linear and Nonlinear Eigenvalue Problems.

7. Johnson, W., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. In: Conference in modern analysis and probability (New Haven, Conn., 1982). Volume 26 of Contemporary Mathematics. American Mathematical Society (1984) 189–206

8. Dasgupta, S., Gupta, A.: An elementary proof of a theorem of Johnson and Lindenstrauss. Random Structures & Algorithms **22**(1) (2003) 60–65

9. Achlioptas, D.: Database-friendly random projections. In: Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. PODS '01, New York, NY, USA, ACM (2001) 274–281

10. Li, P., Hastie, T.J., Church, K.W.: Very sparse random projections. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '06, New York, NY, USA, ACM (2006) 287–296

11. Kanerva, P., Kristoferson, J., Holst, A.: Random indexing of text samples for Latent Semantic Analysis. In: Proceedings of the 22nd Annual Conference of the Cognitive Science Society, Erlbaum (2000) 103–6

12. 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, TKE 2005. (2005)

13. Lupu, M.: On the usability of random indexing in patent retrieval. In Hernandez, N., Jäschke, R., Croitoru, M., eds.: Graph-Based Representation and Reasoning. Volume 8577 of Lecture Notes in Computer Science. Springer International Publishing (2014) 202–216

14. Polajnar, T., Clark, S.: Improving distributional semantic vectors through context selection and normalisation. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014), Gothenburg, Sweden, ACL (2014)

15. Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta, E.: The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. Language Resources and Evaluation **43**(3) (2009) 209–226

16. Brinkman, B., Charikar, M.: On the impossibility of dimension reduction in L1. J. ACM **52**(5) (September 2005) 766–788

17. Lapesa, G., Evert, S.: Evaluating neighbor rank and distance measures as predictors of semantic priming. In: Proceedings of the Fourth Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL), Sofia, Bulgaria, Association for Computational Linguistics (August 2013) 66–74

18. Indyk, P.: Stable distributions, pseudorandom generators, embeddings, and data stream computation. J. ACM **53**(3) (May 2006) 307–323

19. Li, P., Samorodnitsk, G., Hopcroft, J.: Sign cauchy projections and chi-square kernel. In Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K., eds.: Advances in Neural Information Processing Systems 26. Curran Associates, Inc. (2013) 2571–2579

20. Geva, S., De Vries, C.M.: TOPSIG: Topology preserving document signatures. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. CIKM '11, New York, NY, USA, ACM (2011) 333–338

21. Zadeh, B.Q., Handschuh, S.: Random Manhattan integer indexing: Incremental L1 normed vector space construction. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics (2014) 1713–1723

22. Zadeh, B.Q., Handschuh, S.: Random Manhattan indexing. In: 25th International Workshop on Database and Expert Systems Applications. DEXA'14, IEEE (2014) 203–208

23. Baroni, M., Lenci, A., Onnis, L.: ISA meets Lara: An incremental word space model for cognitively plausible simulations of semantic learning. In: Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, Prague, Czech Republic, Association for Computational Linguistics (June 2007) 49–56