BLOSSOM: Best path Length on a Semantic Self-Organizing Map

Robert V. Lindsey (lindsr@rpi.edu) **Michael J. Stipicevic** (stipim@rpi.edu) **Vladislav D. Veksler** (vekslv@rpi.edu) **Wayne D. Gray** (grayw@rpi.edu)

Rensselaer Polytechnic Institute
Cognitive Science Department, 110 8th Street
Troy, NY 12180 USA

Abstract

We describe Vector Generation from Explicitly-defined Multidimensional semantic Space (VGEM), a method for converting a measure of semantic relatedness (MSR) into vector form. We also describe Best path Length on a Semantic Self-Organizing Map (BLOSSOM), a semantic relatedness technique employing VGEM and a connectionist, nonlinear dimensionality reduction technique. The psychological validity of BLOSSOM is evaluated using test cases from a large free-association norms dataset; we find that BLOSSOM consistently shows improvement over VGEM. BLOSSOM matches the performance of its base-MSR using a 21 dimensional vector-space and shows promise to outperform its base-MSR with a more rigorous exploration of the parameter space. In addition, BLOSSOM provides benefits such as document relatedness, concept-path formation, intuitive visualizations, and unsupervised text clustering.

Keywords: Measures of Semantic Relatedness, Self-Organizing Maps, nonlinear dimensionality reduction, computational linguistics, natural language processing, VGEM, BLOSSOM, Dijkstra's algorithm, SOM traversal

Introduction

Measures of Semantic Relatedness (MSRs) are a class of computational methods that calculate association strengths between terms in order to quantify the meaning of text. MSRs have a wide variety of applications derived from their mathematical modeling of text meanings. These applications range from the computational modeling of human text comprehension (Lemaire, Denhière, Bellissens, Jhean-Larose, 2006) to automated essay-grading algorithms (Landauer & Dumais, 1997).

Most MSRs employ either statistical or vector-based techniques. Vector-based MSRs such as Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) and Generalized Latent Semantic Analysis (GLSA; Matveeva, Levow, Farahat, & Royer, 2005) generally model term-to-term association strengths well, but require significant computational resources to manipulate vector-space models of text. For example, LSA involves the difficult task of solving a "large, sparse symmetric eigenproblem" (Landauer, McNamara, Dennis, & Kintsch, 2007, p. 45).

Statistical MSRs such as Pointwise Mutual Information (PMI; Turney, 2001) and Normalised Google Distance (NGD; Cilibrasi & Vitanyi, 2007) are often based on search engine technology (e.g. Google search) and have virtually unlimited vocabularies. These measures are not as computationally intensive as vector-based MSRs, but suffer from a number of deficits. One issue is that statistical MSRs cannot measure the relatedness of large multi-word terms (e.g. documents). Another issue relates to second-order word co-occurrence (words that do not occur together, but are often found in similar contexts). There is evidence in the literature that vector-based measures provide better performance over statistical measures due to context-based word processing (Budiu, Royer, & Pirolli, 2007; Landauer, et al., 2007).

In this paper, we describe VGEM (Veksler, Govostes, & Gray, 2008), a technique combining the benefits of statistical MSRs with the power of vector-based MSRs. We then propose BLOSSOM, an MSR that utilizes a Self-Organizing Map to reduce noise in vector-space semantic models. The psychological validity of the proposed technique is then evaluated using human word-association norms. BLOSSOM is shown to exhibit consistent improvement over VGEM. Finally, we describe the unique benefits of using BLOSSOM over other MSRs, including intuitive visualizations and concept-path formations.

VGEM

To convert a statistical MSR M into vector form, we use Vector Generation from Explicitly-defined Multi-dimensional semantic space (VGEM). VGEM's semantic space is explicitly defined by a set of dimensions $d = \{d_1, d_2, ..., d_n\}$, where each word defines a single dimension. To compute the vector representation of a word in this semantic space, VGEM uses its base-MSR M to calculate the semantic relatedness between the target word w and each dimension in d as follows:

$$v(M,w,d) = [\ M(w,d_1), M(w,d_2), ..., M(w,d_n)\ ]$$

For example, if d = {"animal", "friend"} and the word in question is "dog", then the vector for "dog" would be [M("dog", "animal"), M("dog", "friend")]. If M("dog", "animal") is 0.81 and M("dog", "friend") is 0.84, then the vector is [0.81, 0.84] as demonstrated in Table 1 and Figure 1. This example uses only two dimensions, though typically many more are used.

Like other vector-based measures (e.g. LSA, GLSA), VGEM defines similarity between two words to be the cosine of the angle between the vectors that represent those words. As the angle becomes smaller and the cosine approaches 1.0, the words are considered more related. A value of 1.0 means the two words are identical in meaning. For example, in Figure 1, θ, the angle between "dog" and "cat", is relatively small. Consequently, the cosine of θ is close to 1.0 (.994) and the two words are considered to be closely related.

Table 1: Sample VGEM Computations

Words	Dimensions	
	Animal	Friend
Dog	0.81	0.84
Cat	0.81	0.67
Tiger	0.79	0.13
Robot	0.02	0.60

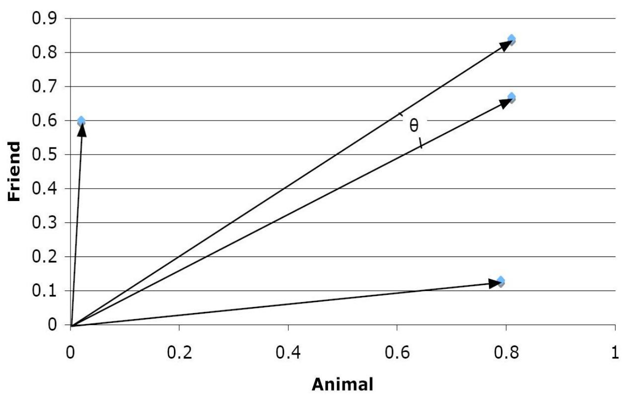

Figure 1: Sample VGEM semantic space.

This vector-based approach allows VGEM to represent a group of words as the vector sum of the words that make up the group. To compute the vector representation of a paragraph, VGEM creates a vector for each word in that paragraph and adds them component by component. This vector sum represents the meaning of the whole paragraph; its relatedness to other vectors may be measured as the cosine of the angle between those vectors. Continuing with the example in Table 1 and Figure 1, the vector representing the words "dog cat tiger" is the sum of first three vectors in Table 1, [2.41, 1.64].

Advantages of VGEM

One of the advantages VGEM has over statistical MSRs is that it can compute relatedness between large, multi-word terms. A statistical MSR cannot find the relatedness between two paragraphs because the probability of any two paragraphs co-occurring (word for word) in any context is virtually zero. VGEM, like other vector-based measures, can represent a paragraph or a document as a vector, and then compare that vector to other vectors within its semantic space.

The main advantage of VGEM over other vector-based MSRs is that VGEM does not require extensive preprocessing. This affords VGEM a larger dynamic lexicon. Other vector-based MSRs cannot handle corpora that are very large or dynamic.

Dimensions

Choosing a good set of dimensions is critical to VGEM's performance in tests of psychological validity. Although genetic algorithms may be used to derive a good set of dimensions (Veksler et al., 2008), this process is computationally intensive, especially for a large set of dimensions. BLOSSOM, as described below, achieves high performance even from suboptimal VGEM dimensions.

BLOSSOM

Best path Length on a Semantic Self-Organizing Map (BLOSSOM) is an MSR that calculates semantic relatedness as the inverse of the distance one would have to "travel" from one term to another on a Self-Organizing Map (SOM). This SOM must be trained on an input semantic vector-space. In this paper we concentrate exclusively on the VGEM vector-space; future work will explore other vector-spaces. BLOSSOM uses a SOM to reduce the dimensionality of its input vector-space. To find the relatedness between two terms, BLOSSOM calculates the cost associated with the shortest path between the terms in the SOM's internal representation of semantic space.

Self-Organizing Maps afford a variety of benefits. Critical to BLOSSOM is the ability to create a weighted graph representation of the input space. The graph enables the calculation of the shortest path between any two nodes. This shortest path represents the distance between two points in semantic space.

We are not the first to consider this type of path-analysis on a SOM. As Bui and Takatsuka (2007) describe, "we envisage that if an abstract knowledge space can be represented in the form of a map, transitions of knowledge can be defined as a path on this map." Because our abstract knowledge space is the English language (in vector form), "transitions of knowledge" are conceptual differences in words. Thus, BLOSSOM measures similarity as the shortest path through the knowledge space.

Self-Organizing Maps

A Self-Organizing Map is an artificial neural network employing unsupervised, competitive learning techniques that performs nonlinear dimensionality reduction. It provides a method to represent high-dimensional data in low-dimensional spaces, often in 2-d or 3-d visualizations. The mappings between high- and low-dimensional spaces preserve topological properties, meaning that high-dimensional structures are represented as independent clusters on the low-dimensional map (Ultsch, 1995). Topological preservation is crucial to BLOSSOM because its method of calculating semantic relatedness relies on an accurate low-dimensional mapping of semantic space.

The nodes of the SOM used by BLOSSOM are arranged in a lattice structure, though many other possibilities exist (Wu & Takatsuka, 2005). Associated with each node is a feature vector consisting of the same dimensionality as the VGEM input space. As with most neural networks, a SOM has two modes of operation, a learning mode and an operational mode.

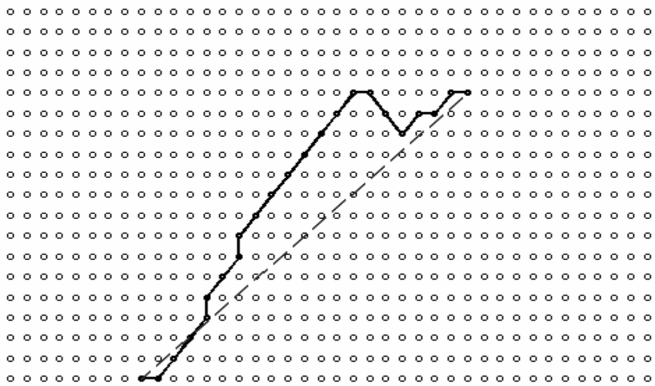

Figure 2a: Path Through SOM Nodes

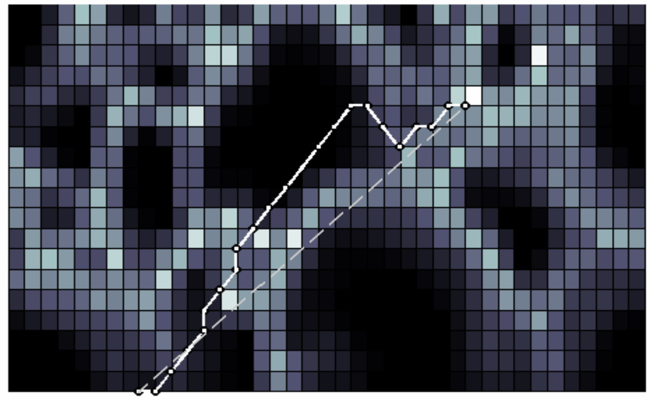

Figure 2b: Path Through SOM Nodes with Clustering Shown

Learning Mode When an input vector is given to a SOM, the Euclidian distance between the input vector and each node's feature vector is calculated. The node with the smallest Euclidian distance from the input vector, known as the Best-Matching Unit (BMU), has its feature vector adjusted according to the update formula:

$$\mathbf{w}_k(t+1) = \mathbf{w}_k(t) + \mu \mathrm{X}(i,k)(\mathbf{x} - \mathbf{w}_k(t))$$

where i denotes the BMU, k denotes another node, t represents a time index, μ is the learning rate, x is the input vector, and $\mathrm{X}(i,\ k)$ is a time-decaying neighborhood function. All nodes within the neighborhood of the BMU on the lattice also have their feature vectors adjusted. A node's adjustment is proportional to its distance from the BMU.

Operational Mode When a SOM is given an input vector, the BMU is found. The coordinates of the BMU on the node lattice are the low-dimensional representation of the input vector. Even when presented with a novel input vector, a SOM is able to reduce the vector to a point on the lattice.

BLOSSOM Setup

Words, as represented by their VGEM vectors, are clustered on the SOM. Similar words are grouped within the same cluster. Figure 2b shows a 2-d SOM trained on the VGEM semantic space. In this figure, the colors range on a gradient from black to white; dark areas represent clusters of similar text and light areas represent cluster boundaries. Figure 2c represents Figure 2b in three dimensions, where clusters are represented as "valleys" and cluster boundaries are represented as "mountains" in this semantic landscape.
This representation illustrates the intuitive visual appeal of computing semantic relatedness as the traversal of a semantic landscape. There are two example paths traced through the semantic landscapes of Figures 2a and 2b. The traversal cost associated with a path is defined as the sum of a series of "tolls" associated with the transition from one node to another. Figure 2c represents the inter-node tolls as height; a high inter-node toll signifies a large conceptual difference between the two nodes. The dotted path in Figures 2a and 2b is a straight line from one point to another in semantic space. This straight line crosses several major cluster boundaries and may not be a good representation of

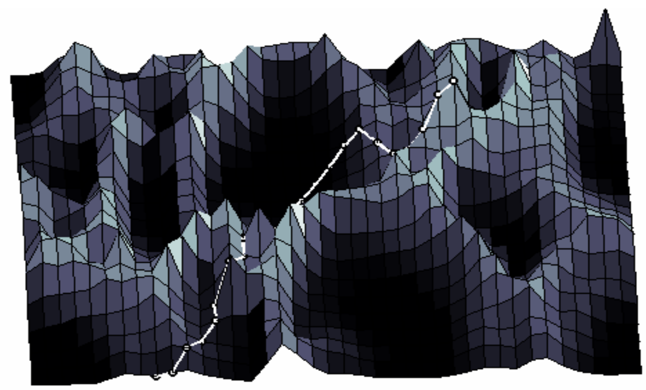

Figure 2c: Path Through SOM Nodes with Clustering Represented as Height

semantic distance (see Results). The solid path avoids directly crossing the semantic "mountains", and represents a less costly path through the space because no major conceptual boundaries are crossed. BLOSSOM defines semantic distance as the best path through the semantic space, which is the path that minimizes the total toll cost.

SOM Training The SOM employed by BLOSSOM must learn to represent the VGEM semantic space. This is accomplished by providing it with a large, representative sample of training data. For the purposes of this paper, we used VGEM vectors representing words randomly selected from the Factiva database (Factiva, 2004). The SOM training algorithm repeatedly selects a random vector from this training set. It then performs the steps of the Learning Mode by adjusting the BMU and its neighboring nodes to more accurately represent the training word. After a great many samplings, the nodes of the SOM will have sufficiently adapted to the VGEM representation of the English language.

Undirected Graph Representation Once the SOM is trained, an undirected graph is constructed to facilitate the calculation of shortest paths between nodes. The undirected graph represents the relations between feature vectors, but does not represent the feature vectors themselves. Edges in this graph are created only between neighboring nodes. Each edge corresponds to the "toll" of traveling from a node to one of its neighbors (as discussed above). More precisely, inter-node traversal cost is the angle between the two nodes' feature vectors. The angle represents the conceptual dissimilarity between the two nodes. Because of this, subsequent shortest-path calculations performed on this graph are forced to take into consideration the complex topological landscape encoded in the trained SOM (as demonstrated in Figure 2c). This ensures that sharp semantic jumps are avoided as much as possible.

SOM Traversal

Once the SOM has adapted to the VGEM semantic space and a corresponding undirected graph has been created, BLOSSOM is able to calculate the degree of association between two query words. This involves node selection and shortest path-cost calculation. For a pair of terms, BLOSSOM selects two nodes on the SOM best representing the terms. It then finds the shortest path between these two nodes on the undirected graph generated during setup.

Node Selection The two terms are converted into VGEM vectors based on the same explicitly-defined dimensions that the SOM was trained on. A Best Matching Unit is then found for each query word's VGEM vector.

Shortest Path Next, BLOSSOM finds the shortest path from one node to another on the graph representation of the SOM. We use Dijkstra's algorithm, though other shortest path algorithms (e.g. Floyd's algorithm) can be used. This shortest path represents semantic distance as measured by BLOSSOM. It is then converted to semantic relatedness by applying a monotonically decreasing function. For simplicity's sake, we chose this function to be f(x) = -x.

There are at least two alternatives to the above method of computing semantic relatedness: (1) use the Euclidean distance on the SOM between the BMUs of the two terms, as represented by the straight, dotted line in Figures 2a and 2b or (2) take the cosine between the feature vectors of the BMUs. These potential techniques are considered and analyzed in the Results section.

Evaluating BLOSSOM: Methodology

BLOSSOM has a large number of free parameters affecting performance (the ability to model human free-association data). In the Results section, we examine the impact of several of these parameters. One of the major variables that we kept static for testing purposes is the base-MSR used in VGEM vectors (denoted by *M* in the VGEM section). We chose to use an MSR known as Normalised Similarity Score (NSS) and utilized the Factiva text database. In the future, we will examine other MSRs (e.g. Pointwise Mutual Information), other corpora (e.g. Google, Wikipedia, Project Gutenberg), and other performance evaluation methods (e.g. Cilibrasi & Vitanyi, 2007).

NSS is based on Normalised Google Distance (NGD, Cilibrasi & Vitanyi, 2007). To be more precise,

$$NGD(w_1, w_2) = \frac{\max(\log f(w_1), \log f(w_2)) - \log f(w_1, w_2)}{\log N - \min(\log f(w_1), \log f(w_2))}$$

$$NSS(w_1, w_2) = 1 - NGD(w_1, w_2)$$

where $f(w)$ is the number of documents w occurs in, $f(w_1,w_2)$ is the number of documents w_1 and w_2 appear together in, and N is the number of documents in the corpus.

We used a subset (≈2.3 million paragraphs) of the Factiva database (Factiva, 2004) as a training corpus for NSS. This subset should be sufficient to reveal the power of BLOSSOM because it represents a sizeable sampling of the English language.

Evaluation To test the performance of the proposed technique, we used an evaluation function based on human word-association norms, previously used by Lindsey et al. (2007). Specifically, we used the Nelson-McEvoy free-association norms (Nelson, McEvoy, & Schreiber, 1998) to find out which *target* words were free-associated from *cue* words (e.g. when the cue is 'old', targets might be 'new', 'young', 'ancient', 'man', 'wrinkle', 'age'). For each set of *cue-targets* we picked *n* random distracters, where *n* is the number of targets. Distracter words were chosen randomly from the Factiva corpus, such that a third of the words were rare (e.g. 'manatees', 'videocassette'), a third were common (e.g. 'days', 'to'), and the rest were in between (e.g. 'specifically', 'external'). We took a random sample of 100 *cue-targets-distracters* test cases to evaluate a given MSR, *M*.

To evaluate each *cue-targets-distracters* test case, each of the *targets-distracters* lists is sorted in descending order of *M*-derived relatedness values, *M(cue, word)*, where $word \in \{targets, distracters\}$. The score for each *cue-targets-distracters* test case for *M* is calculated as follows:

$$Score_{\text{case}} = \frac{\textit{Number of targets in top n words}}{n}$$

where "*top n words*" is the top half of the ordered *targets-distracters* list. If, according to *M*, all target words are more related to *cue* than any of the distracter words, the score for that test case is 100%. If none of the target words are picked by *M* to be more related to the *cue* than any of the distracter words, the score is 0%. The overall score for *M* is the average of all test case scores.

Results

To evaluate BLOSSOM, we examined $\text{BLOSSOM}_{\text{NSS-Factiva}}$ using different SOM sizes and training set sizes. The dimensions we picked varied from random sets of 10, 100, 200, and 300 words from the Factiva corpus to a set of 21 words from a general ontology. For the purposes of this paper, we tried a relatively small variation of these settings until we could match the base-measure (NSS-Factiva) performance. Future work will involve a more rigorous exploration of these settings.

Our best results for BLOSSOM came from the set of 21 ontology-based dimensions, on a 5x15 SOM using 320,000 training words. In this case, BLOSSOM showed no significant difference from the base-MSR, $t(99)_{two\text{-}tail}=.405$, $p=.68$. BLOSSOM showed significant improvements over VGEM by 6.27%, $t(99)_{two\text{-}tail}=3.96$, $p<.01$.

BLOSSOM consistently outperformed VGEM, BMU-Euclidean-distance, and BMU-vector-cosine measures (the latter two measures are defined in the Shortest Path section, points (1) and (2) respectively). Using random dimensions, all tested SOM sizes improved on average compared to VGEM, BMU-Euclidean-distance, and BMU-vector-cosine measures by 11.67%, 7.50% and 3.07%, respectively.

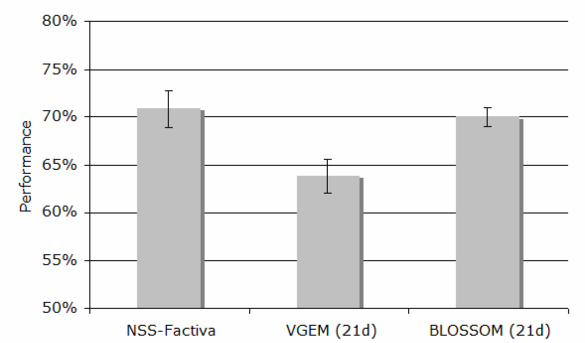

Figure 3. NSS-Factiva vs. VGEM vs. BLOSSOM using 21 dimensions. Error bars signify standard error.

Discussion

The presented results are bounded by the small scope of the parameter space we explored. It is premature to conclude that BLOSSOM is limited by the performance of its base-MSR, NSS-Factiva. Evidence exists in the literature that dimensionality reduction techniques can be used to improve the performance of a base-MSR (Budiu, Royer, & Pirolli, 2007).

The conversion of NSS into vector form with randomly picked dimensions favors the BLOSSOM method over VGEM. Moreover, BLOSSOM's performance was better when using ontology-based dimension words rather than much larger sets of randomly picked dimension words. This may be due to the fact that Self-Organizing Maps cluster the semantic space; perhaps clusters are better formed when input vectors are based on categorical knowledge, rather than randomly sampled words.

One of the reasons that BLOSSOM, BMU-Euclidean-distance, and BMU-vector-cosine measures all perform better than VGEM given the same random dimensions is that using a SOM affords the ability to handle noisy, high-dimensional data (Pascual-Montano, Taylor, Winkler, Pascual-Marqui, & Carazo, 2001). A vector-space representation of human semantics gleaned from any text corpus is subject to a certain amount of error. The corpus may not lend itself toward a proper representation of human lexical knowledge, or the MSR used to generate the vector-space may simply be lacking. A SOM appears to compensate for these deficiencies.

However, there are some problems associated with using SOMs. One major problem we encountered is known as the border effect (Kohonen, 2001). Nodes near the border of a 2-d SOM are less likely to be updated during the learning process because they do not have many neighbors. This results in unreliable clustering in the outlying regions of the SOM, which may prevent BLOSSOM from fully realizing its potential. In the future, we plan to implement a Geodesic Self-Organizing Map to remove the Border Effect problem. Geodesic SOMs use a spherical lattice to eliminate borders entirely (Wu & Takatsuka, 2005).

Linear vs. Nonlinear Dimensionality Reduction

As previously discussed, Self-Organizing Maps are a technique for nonlinear dimensionality reduction. LSA, a well-established MSR (Landauer et al., 2007), employs a linear dimensionality reduction technique known as Singular Value Decomposition (SVD). Nonlinear reduction techniques have significant advantages over linear techniques for feature extraction (Backer, Naud, & Scheunders, 1998). If the VGEM semantic space contains nonlinear manifolds, some points in the low-dimensional mapping will be incorrectly mapped when using a linear technique. Self-Organizing Maps are able to properly cluster complex topological structures that linear statistical methods cannot (Ultsch, 1995). We believe that with further development, BLOSSOM, due to its nonlinear nature, may exceed the capabilities of SVD-based MSRs.

Concept-Path Formation

The shortest path between two words in BLOSSOM represents a gentle transition from one point to another in semantic space. Because the feature vectors of the SOM represent points in semantic space, each node traversed can be assigned a term whose VGEM representation lies nearby. The nodes on a path from one term to another represent the intermediate connecting concepts. For example, the path between “foot” and “shirt” may pass through nodes matching “shoe” and “clothes”, respectively. BLOSSOM is fully capable of creating conceptual paths through books, web pages, and other documents. In the future, we will be exploring possible applications of this, the foremost being the modeling of human free-association processes.

Plasticity

When used with VGEM, one of BLOSSOM's advantages is its ability to incorporate new vocabulary and word relationships without significant processing costs. When presented with a new vocabulary word, BLOSSOM only needs to calculate the word’s VGEM vector and proceed with the SOM Traversal steps.

Future Work

In the future, we plan to test BLOSSOM’s performance using different base-MSRs (e.g. WordNet). It may even be the case that using a combination of base MSRs will result in higher performance. BLOSSOM may be able to reduce the noise in the high-dimensional data produced by using multiple base MSRs in VGEM vectors.

We also plan to use BLOSSOM for measuring relatedness between paragraphs and documents. VGEM word vectors can be summed up to produce paragraph and document vectors that can be represented in the SOM. This sort of document clustering can be used to improve information-retrieval systems.

Another potential application of BLOSSOM is automatic generation of a lexical ontology from a document. After BLOSSOM clusters words into a SOM, each cluster could be labeled by a word that best describes the cluster. The cluster descriptions could be put into another SOM and this process could be repeated to recursively create a lexical ontology.

Yet another direction left for future exploration is augmentation of web-search queries. Searches could include additional terms found within the original search term's cluster. This would help search engines provide results with subtle, indirect relationships to search terms.

Summary

In this paper, we briefly outlined VGEM, an MSR that integrates the benefits of vector-based and statistical MSRs. BLOSSOM was then introduced as a new MSR that reduces noise in vector-space models like VGEM. Results indicate that BLOSSOM offers a significant improvement over VGEM and can match the performance of VGEM's base-MSR with just a small set of dimensions. BLOSSOM also provides additional benefits, such as the visual appeal of the semantic landscape model and automatic concept-path formation. In the near future, we will attempt to increase BLOSSOM's performance with a more rigorous exploration of the parameter space. We believe that the proposed technology has a wide range of applications in cognitive modeling and cognitive engineering.

Acknowledgements

This work was supported in part by grants from the Office of Naval Research (N000140710033) and the Air Force Office of Scientific Research (FA9550-06-1-0074).

References

Backer, S.D., Naud, A., & Scheunders, P. (1998). Nonlinear dimensionality reduction techniques for unsupervised feature extraction. Pattern Recognition Letters. 19:711-720.

Budiu, R., Royer, C., & Pirolli, P. (2007). *Modeling Informattion Scent: A Comparison of LSA, PMI-IR, and GLSA Similarity Measures on Common Test and Corpora.* Paper presented at the RIAO'07, Pittsburgh, PA.

Bui, Michael; Takatsuka, Masahiro (2007). Path finding on a spherical SOM using the distance transform and floodplain analysis. The 6th International Workshop on Self-Organizing Maps (WSOM), 2007, ISBN: 978-3-00-022473-7

Cilibrasi, R., Vitanyi, P. (2007). The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No. 3, p. 370-383.

Factiva. (2004). 2007, from http://www.factiva.com

Kohonen, T. (2001). Self-Organizing Maps, Third Edition, Springer.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104(2), 211-240.

Landauer, T., McNamara, D., Dennis, S., Kintsch, W. (2007). Handbook of Latent Semantic Analysis. Lawrence Erlbaum Associates, Publishers. Mahwah, New Jersey.

Lemaire, B., Denhière, G., Bellissens, C., Jhean-Larose, S. (2006). A computational model for simulating text comprehension. Behavioral Research Methods, Volume 38, Number 4, pp. 628-637 (10).

Lindsey, R., Veksler, V. D., Grintsvayg, A., & Gray, W. D. (2007). Effects of Corpus Selection on Semantic Relatedness. 8th International Conference of Cognitive Modeling, ICCM2007, Ann Arbor, MI.

Matveeva, I., Levow, G., Farahat, A., & Royer, C. (2005). Term representation with generalized latent semantic analysis. Paper presented at the 2005 Conference on Recent Advances in Natural Language Processing.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. : http://www.usf.edu/FreeAssociation/.

Pascual-Montano, A., Taylor, K. A., Winkler, H., Pascual-Marqui, R.D., Carazo, J.M. (2002). Quantitative Self-Organizing Maps for Clustering Electron Tomograms. Journal of Structural Biology. Volume 138, Issues 1-2, 114-122.

Turney, P. (2001). Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. In L. De Raedt & P. Flach (Eds.), Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001) (pp. 491-502). Freiburg, Germany.

Ultsch, A. (1995). Self Organizing Neural Networks perform different from statistical k-means clustering. *Gesellscahft fur Klassifikation*, Basel.

Veksler, V. D., Govostes, R., & Gray, W. D. (2008). Defining the Dimensions of the Human Semantic Space. Accepted to the 30th Annual Meeting of the Cognitive Science Society.

Wu, Y. and Takatsuka, M. (2005). The Geodesic Self-Organizing Map and its error analysis. In Proceedings of the Twenty-Eighth Australasian Conference on Computer Science - Volume 38 (Newcastle, Australia). V. Estivill-Castro, Ed. ACM International Conference Proceeding Series, vol. 102. Australian Computer Society, Darlinghurst, Australia, 343-351.