QUT Digital Repository:
http://eprints.qut.edu.au/

QUT

This is the accepted version of this conference paper:

De Vine, Lance & Bruza, Peter D. (2010) *Semantic oscillations : encoding context and structure in complex valued holographic vectors.* In: Quantum Informatics for Cognitive, Social, and Semantic Processes (QI 2010), 11-13 November 2010, Arlington, Virginia.

Semantic Oscillations: Encoding Context and Structure in Complex Valued Holographic Vectors

Lance De Vine[1], Peter Bruza[1]

[1] Faculty of Science and Technology, Queensland University of Technology
l.devine@qut.edu.au, p.bruza@qut.edu.au

Abstract

In computational linguistics, information retrieval and applied cognition, words and concepts are often represented as vectors in high dimensional spaces computed from a corpus of text. These high dimensional spaces are often referred to as Semantic Spaces. We describe a novel and efficient approach to computing these semantic spaces via the use of complex valued vector representations. We report on the practical implementation of the proposed method and some associated experiments. We also briefly discuss how the proposed system relates to previous theoretical work in Information Retrieval and Quantum Mechanics and how the notions of probability, logic and geometry are integrated within a single Hilbert space representation. In this sense the proposed system has more general application and gives rise to a variety of opportunities for future research.

Introduction

In a variety of studies from cognitive science there have been encouraging results indicating that vector representations automatically generated from a corpus of text are able to replicate human word association norms, for example, semantic association (Lund and Burgess 1996; Landauer 2002; Lowe 2001; Widdows 2004). Such models are often referred to as "semantic space" models. These studies provide evidence that vector representations within semantic space models capture the semantics of words in a pragmatic, every day sense. This opens the door to exploiting such models for developing information processing technologies, which are at least partially sensitive to cognitive semantics.

Models such as Latent Semantic Analysis (Landauer and Dumais 1997) and Hyperspace Analogue to Language (Lund 1996) rely on co-occurrence matrices to produce high dimensional semantic vectors. Such large matrices are often dimensionally reduced via Singular Value Decomposition (SVD), a computationally expensive process. An alternative to these full co-occurrence matrix based methods is Random Indexing (Kanerva, Kristofersson and Holst 2000; Sahlgren 2005) (RI) which performs online dimensional reduction and is much more computationally efficient while still providing comparable quality (Widdows 2008). The semantic vectors generated by RI are not fixed but are rather incrementally updated to reflect the similarity in meaning between the entities that they represent.

The basis of the claim that Semantic Space models are able to capture meaning is the distributional hypothesis (Harris 1954), which states that words with similar meanings tend to occur in similar contexts. The vector corresponding to word w encodes co-occurrence information of words co-occurring with w in context and therefore the vector can be viewed as a computational manifestation of Firth's famous quote, "You shall know a word by the company it keeps" (Firth 1957).

Vector spaces are attractive for modeling the contextual meaning of words as they are mathematically well defined, are well understood and provide a computationally tractable framework with which to compute the semantics of a given textual corpus (Sahlgren 2005). Vector spaces also fit well with human cognition and its grounding in spatial metaphors (Gardenfors 2004). A metric defined on a vector space provides a means for easily computing similarity between objects in that space.

One of the challenges for Semantic Space models has been that they often don't capture the structural relationships between words in texts. The models are based on a 'bag of words' approach. The structural relationships are not captured due to the perceived difficulty of not being able to effectively encode structural information within vectors. It is relevant and worth mentioning that this difficulty is the same as that found in the area of neural networks which are often charged with not being able to represent compositionality and systematicity (Plate 1994). These problems are grounded in the tension between localist (symbolic) and distributed forms of representation.

Within the Semantic Spaces research community recent years have seen the emergence of several methods that have been shown to be effective in encoding structural information in high dimensional vectors. Jones, Kintsch and Mewhort (Jones, Kintsch, and Mewhort 2006) introduced the BEAGLE model which uses Holographic Reduced Representations (HRR) to simultaneously encode word order and context information within semantic vectors. More recently (Sahlgren, Holst, and Kanerva 2008) have introduced a derived model for encoding word

order based on permutations. The permutation model was recently introduced to the quantum interaction community (Widdows and Cohen 2009). There is evidence that capturing word order does improve the quality of the semantic representation (Jones, Kintsch and Mewhort 2006; Widdows and Cohen 2009).

A disadvantage of the BEAGLE model is that it relies on a compression of the outer product of two semantic vectors which is computationally costly to compute. In this article an alternative is proposed and evaluated which involves using complex valued vectors in which the complex valued elements can be interpreted as the frequency components of semantic waveforms or signals. This approach uses aspects of both BEAGLE and the Permutation approach but also introduces novel conceptualizations that have interesting relationships to some existing areas of research.

BEAGLE - Beyond Bag of Words

BEAGLE, or Bound Encoding of the Aggregate Language Environment, is one of the more recent examples of a computational model of word meaning. The major advance offered by BEAGLE is that word representations include *both* a consideration of order information (i.e., structure) in addition to word context information. The basis for encoding structure is the use of an outer product of two vectors resulting in a matrix, or rank 2 tensor. By way of illustration, assume the word *dog* is represented by a column vector *d* and the word *bite* to be represented by the column vector *b*. The association between these words, "dog bite", can be represented as the outer product of these two vectors: db^T The resulting matrix represents an ordered association between the words *dog* and *bite* . Such matrices have been used to model ordered word associations in human memory, e.g., (Humphreys, Bain, and Pike 1989) and more recently in BEAGLE (using the compressed outer product). Note db^T is not necessarily equal to the matrix bd^T, which is what one would expect as word order in language is by and large not commutative.

The above scheme using outer products can be generalized into representing arbitrarily long sequences of words by using the Kronecker (tensor) product[1], but the tensor representations explode rapidly in dimensionality. One approach to redress this problem is to constrain the dimensionality by compressing the outer product into a vector representation of the same dimensionality as the constituent vectors in the outer product. This is the approach used in a type of representation known as a Holographic Reduced Representation (HRR) introduced by Plate (Plate 1991). BEAGLE uses HRRs as they enable the binding of representations without increasing the dimensionality. This binding (compressed outer product), is accomplished via circular convolution.

---

[1] the ``outer'' or ``dyadic'' is a special case of the tensor product between two vectors of the same dimension

For example, given two vectors x= $(x_0, ... ,x_n)$ and y = $(y_0, ... ,y_n)$, the circular convolution x$\otimes$y results in an *n*-dimensional vector z = $(z_0, ..., z_n)$ whereby each component z_i of the vector representation is computed according to the following equation:

$$z_i = \sum_{j=0}^{n} x_{j(\bmod n+1)} y_{(i-j)(\bmod n+1)}$$

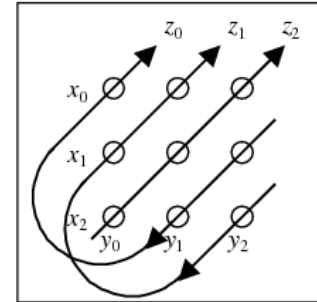The above equation can be visualized as depicted in Figure 1.

Figure 1. Circular Convolution

Circular Convolution has complexity $O(n^2)$ but this can be reduced to $O(n\ logn)$ by using a Fourier transform to transform both vectors into the frequency domain, applying an element-wise multiplication, and then transforming back.

The other operations required for BEAGLE are superposition, permutation and a similarity measure. Superposition equates to simple vector addition and allows one to represent items as sets (superpositions). Permutation corresponds to simply re-ordering the elements of the vector. The similarity measure that is generally used is the cosine of the angle between the normalized vectors that are being compared.

BEAGLE uses two different vectors for each word *w* in the model: a) an environmental vector, and b) a memory vector. The environmental vector is a word signature vector with elements of the vector sampled from a normal distribution with mean 0.0 and variance $1/n$, where *n* is the dimensionality of the vector. The information in the memory vector can also be stored in two separate vectors, one for context and one for structure.

The context vector is a standard co-occurrence vector for *w,* the components of which give a weighted representation of how words are co-occurring with word *w* using a sentence or term window as the unit of context.

The structure vector o_w is used to accumulate word order information formed by the superposition of vectors representing *n*-grams involving word *w*. For example, consider the sentence "A dog bit the mailman". The structure vector o_{dog} is built up as a sum of so-called

"bindings", each of which is defined in terms of a convolution. For example, for bi-grams ($n=2$),

$$\text{bind}_{dog,1} = e_A \otimes (\Pi \; \Phi),$$
$$\text{bind}_{dog,2} = \Phi \otimes (\Pi \; e_{bit})$$

Where e_A is the environmental vector for "A", e_{bit} is the environmental vector for "bit", Π is a predefined permutation, and $\otimes$ is the convolution product. The permutation is required to make convolution non-commutative. The position of the word being coded is represented by a constant random placeholder vector, Φ (sampled from the same element distribution from which the environment vectors e were constructed).

All n-gram vectors ($2 \leq n \leq 7$) are thus formed are then superposed into a single vector, normalised, and then added to the structure vector for the target word. For example,

$$o_{dog} = \sum_{j=1}^{7} \text{bind}_{dog,j}$$

Once again, when all words in the corpus have been processed, the structure vector for each word is normalised, and this normalised vector represents the structure signal for that word in the context of the corpus. Observe how multiple circular convolutions must be computed for the structure vector of each word in the vocabulary of interest. This is computationally intensive. For a full and detailed description of the construction of structure vectors and their use the reader is encouraged to see (Jones, Kintsch, and Mewhort 2006)

Random Indexing

Random Indexing (Kanerva, Kristofersson and Holst 2000) (RI) introduced an effective and scalable method for constructing semantic spaces from large volumes of text. The method is based on work by Kanerva on Sparse Distributed Representations (Kanerva 88, 2000). A recent introduction to computing with large distributed representations is given by Kanerva in (Kanerva 2009). The RI method is based on the observation that there are many more nearly orthogonal than truly orthogonal directions in high dimensional space (Hecht-Nielsen 1994) so that if we project points in a vector space into a randomly selected subspace of sufficiently high dimensionality, the distances between the points are approximately preserved (Johnson and Lindenstrauss 1984). The random projection matrix is often constructed from a Gaussian distribution but (Achlioptas 2001) has shown that simpler distributions can be used. RI generally uses sparse ternary random vectors with values (1, 0,-1).

Encoding Word Order with Permutations

Sahlgren, Holst, and Kanerva introduced the idea of permuting the coordinates of random vectors to encode information about word order (Sahlgren, Holst, and Kanerva 2008). When the coordinates of a vector are shuffled with a random permutation, the resulting vector is nearly orthogonal to the original one with very high probability. The original vector can be recovered by applying the reverse permutation, meaning that permutation is invertible. Since the elements of the random vectors are independent of each other, a simple rotation of the elements can be used for permutation. To encode the word order vector for "dog" using the phrase "A dog bit the mailman" and using a term context window of two, we have

$$<\text{dog}> = (\Pi^{-1}a) + (\Pi^{+1}\text{bit}) + (\Pi^{+2}\text{the})$$

where Π^n indicates rotation by n places. Using permutation to encode word order is very computationally efficient compared to using convolution to encode word order as implemented in BEAGLE. The primary difference is that ngrams encoded using convolution contribute very specific information to the focus word vector, while encoding using permutation contributes less specific but more generalizable information (Sahlgren, Holst, and Kanerva 2008). Permutation can also be used to simply encode whether a word is to the left or to the right of the focus word.

Circular Holographic Reduced Representations

We now describe a variation of standard HRRs that are described by Plate In his PhD thesis (Plate 1994; see also Plate 1991). Plate informally refers to this variation as "circular vectors" or "HRRs in the frequency domain". For the purpose of this paper we will refer to them as Circular HRRs or CHRRs. The idea is based on the observation that if circular convolution is equivalent to element-wise multiplication with complex numbers in the frequency domain, then maybe it is possible to work exclusively with complex numbers and avoid some of the computational cost of Fourier transforms. The representation therefore becomes a vector of complex elements where each element can be thought of as representing a specific frequency component of a signal, or vector, with associated phase angle and amplitude. The representation is normalized when the complex elements all lie on the unit circle in the complex plane.

Figure 2. A CHRR of dimension 5

Plate describes how operations on standard HRRs map to operations on CHRRs in the frequency domain. A primary advantage of the CHRR representation is that binding (convolution) can be computed in $O(n)$ time in

contrast to O(*n logn*) when using standard HRRs and Fourier transforms.

Entity	Circular System	Standard System
Random vector	Elements iid as ∪(-π,π)	Elements iid as N(0,1/n)
Superposition	Angle of sum θ⊕φ	Addition x + y
Binding (Convolution)	Modulo-2π sum of angles.	Convolution x*y
Inverse	Negation(modulo-2π) -θ	Involution x* (approximate)
Similarity	Mean of the cosine of corresponding angle differences	Dot-product x·y
Normalization	Not needed when elements lie on unit circle	<x>
Permutation	Permute elements	Permute elements

Table 1. Comparison of operations for the Circular and Standard Systems of HRRs. Note: We also add the operation of permutation which has been found to be very useful for encoding information such as sequence position.

The operation of superposition is equal to the angle of the sum of the vectors being superposed. This operation raises some concerns when used in practical applications but how we handle it is described below. The similarity operation is equal to the mean of the cosines of the differences between the two vectors being compared.
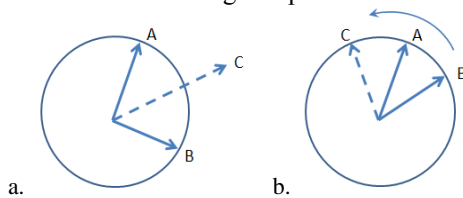

Figure 3. (a) Superposition,(A+B) (b) Convolution A⊗B

We also make use of the exact inverse operation, also called negation. The negation of one vector from another can be written as A not B. It is simply the superposition of A with the complement of B, in other words, the superposition of A with the vector that has components that are 180 degrees out of phase with the components of B. We use this operation in a similar way to the quantum negation operation introduced by Widdows and Peters in (Widdows and Peters 2003).
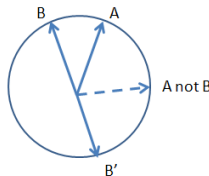

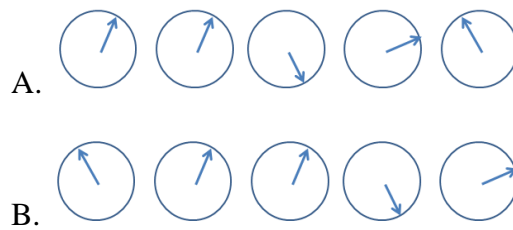Figure 4. Negation (A not B) via the superposition of the complement of B.

Figure 5. CHRR permutation – B is rotated one place to the right in relation to A.

We use CHRRs to construct a BEAGLE-like algorithm that is efficient and conceptually rich.

Application to Semantic Spaces

We now describe how we use CHRRs for encoding ngram structure vectors and how this method differs slightly from that used by BEAGLE. The easiest way to understand the method is to think of it as a hybrid BEAGLE-Permutation approach. NGrams are bound via convolution (in the frequency domain) as in BEAGLE, but the place-holder vector Φ is not used. Instead, ngrams that are to the left of the target word are bound and then permuted (rotated) by -1, and the ngrams that are to the right of the target word are bound and permuted by +1. We then bind these two results together so that the left and right portions of the term window remain associated when they are superposed into the memory vector for the target word. The motivation for constructing the structure vector in this way is that it eliminates the need to compute one of the convolutions. It is also possible, however, to not perform the final binding so that ngrams to the left and right of the target word are not associated with each other when they are superposed into the memory vector. This then results in a method that is very similar to pure Permutation with the difference being that it is not just single term vectors that are permuted and superposed, but also their combinations encoded as bound ngrams. As in BEAGLE, each time we bind two vectors, we permute one of them (the right hand vector) so that information relating to their relative order is preserved.

The proposed system can therefore be used to encode word order in a BEAGLE-like fashion, a pure Permutation fashion, or a combination of both. The system can also construct traditional context vectors which don't encode word order, either using a document context or term-window context. For the purposes of this paper we use the BEAGLE-like approach.

Implementation

Our implementation of the proposed method uses the Java programming language, primarily because it is cross-platform. We find Java to be adequately fast while providing reduced development time compared to other languages such as C and Fortran. We have also

implemented a Java version of BEAGLE which we use to compare with the newly proposed method. The BEAGLE implementation makes use of the Parallel Colt Java library for numerical routines and in particular Fourier transforms. We have also built our system to integrate as much as possible with the SemanticVectors package (Widdows and Ferraro 2008) maintained by Widdows and to which the first author of this paper is a contributor. A difference in data types in the underlying representation prevents complete integration.

At the heart of the implementation is a CHRR object which represents a CHRR and the operations defined on it. It stores complex vector components in two different formats. When the CHRR is normalized, ie. all components lie on the unit circle, the phase angle is stored as a 16 bit Java Unicode char (format 1). This means that the phase angle is discretised using 16 bit precision ie. from 0 to 65535. The Java char data type is also naturally circular so that when we add phase angles they are effectively added modulo 2π. During the operation of superposition the vector components stored as phase angles are converted to complex values where the real and imaginary components are stored as floats (format 2). When superposition is complete (all text is processed) and the representation is normalized the complex values are converted back to phase angles represented as Java chars. The conversion is executed using lookup tables to improve efficiency.

When performing similarity operations a lookup table is also used to save us computing the cosine of angles.

Experiments

For our experiments we used the King James Bible and also the TASA corpus. The TASA corpus (compiled by Touchstone Applied Science Associates) was made available to us courtesy of Professor Thomas Landauer, University of Colorado (Landauer, Foltz and Laham 1998). The TASA corpus contains a collection of English text that is approximately equivalent to what the average college-level student has read in his or her lifetime.

When computing context vectors we use the corpora with stop words removed. When computing structure vectors we include the stop words as they are generally very important for syntactic structure.

We generated models (semantic spaces) using a wide range of parameter values for dimension and ngrams, as well as using both document-term contexts and term-term contexts. The vector dimension varied from 500 to 2048. Ngrams (the maximum number of terms in generated ngrams) varied from 3 to 5. We did not use 7 as used by Jones and Mewhort, as from previous experience this was not found to improve results.

We compared structure queries on the KJB corpus using CHRRs with results obtained by (Widdows 2009) using permutations to encode word order. As can be seen in table 2, results are very similar. Results show similarity to the given query term in the generated semantic space.

"king of ?" (CHRRs)		"king of ?" (Permutation)	
0.712	Syria	0.728	Assyria
0.709	Assyria	0.699	Babylon
0.687	Zobah	0.662	Syria
0.616	Jarmuth	0.647	Zobah
0.610	Persia	0.604	Persia
0.496	Babylon	0.532	Judah

Table 2. Structure query, KJB corpus, ngrams = 3 (Permutation result is taken from (Widdows 2009)

We also performed structure queries using CHRRs on the TASA corpus. As can be seen in table 3, structure queries are quite good at identifying names, object properties and parts of speech

king ?		across the ?	
0.550	midas	0.200	pontoon
0.541	myron	0.144	flatlands
0.539	pellinore	0.135	rooftops
0.534	aegeus	0.135	lawn
0.519	farouk	0.129	quadrangle
0.472	lear	0.124	tfr
0.400	jr	0.119	campground
0.346	tuts	0.115	pleat
Blue ?		He ?	
0.484	jeans	0.688	shouted
0.410	jays	0.571	wrote
0.406	serge	0.561	remembered
0.399	elk	0.551	paused
0.301	litmus	0.550	asked
0.284	eyed	0.542	hated
0.266	ribbon	0.537	felt
0.258	mitten	0.529	spoke

Table 3. Structure query, TASA corpus, dim = 1024, ngrams = 5

Bank		bank NOT savings	
0.464	banks	0.455	bank
0.389	savings	0.257	yellowstone
0.377	checking	0.235	cochino
0.348	deposit	0.219	mississippi
0.326	invest	0.211	banks
0.322	customers	0.211	cuyahoga
0.319	borrow	0.210	nueces
0.315	cochino	0.208	ohio
0.311	escambia	0.207	hudson
0.310	depositors	0.205	anthem
0.310	balances	0.203	park
0.307	lend	0.202	tiber
0.306	depositor	0.202	rhine

Table 4. Similarity query with negation, term-term context, TASA corpus, dimension = 1000

We also successfully used the negation operation to identify the various senses of an ambiguous word, in this case the word "bank".

Results

The experiments we have reported are only a small part of the experiments that are required to make a proper comparison of different semantic space building approaches. This will be the topic of future research. What we can say is that the results that we have obtained seem to be comparable in quality to results obtained using BEAGLE and Permutation. The system that we have described and implemented is significantly faster than that of BEAGLE but not as fast as that of using permutation to encode word order. We believe, however, that the real benefit of using CHRRs lies in the conceptual richness of the representation that is still yet to be explored.

Related Research

We have found that the representational system which we have described above for representing context and structure within text has many links to other areas of research. Some of these links may be formed purely by similarities in mathematical formalisms, while others may indicate a much deeper structural connection. We now elaborate on a few of these connections.

Holography

Before matrix memories were used to construct distributed associative memories, optical holography was proposed as an analogy for human memory. A number of authors including (Pribram 1966, Gabor 1968; Willshaw, Buneman, and Longuet-Higgins 1969; Borsellino and Poggio 1973) considered associative memory models based on convolution and correlation, mathematical operations used in holography. To construct a holographic image a wave which has been reflected from an object or scene (object wave) falls upon the recording medium. Another wave known as a reference beam also illuminates the recording medium so that interference occurs between the two beams. A seemingly random pattern is produced (hologram) which when illuminated by the original reference beam produces the original light field which was reflected from the objects or scene. This effect can be used to implement a form of holographic storage and it is a form of hetero-associative memory. It is this that inspired Plate to call his representations Holographic (Plate 1994). The vector memories described above to encode word order can also be compared to a type of holographic storage. There may be other types of wave-like phenomena that we can map to CHRRs and this opens up many avenues for further research.

Quantum Interaction

Recent times have seen the application of quantum formalisms to non-physical complex systems that exhibit contextual effects and interactions. (Kitto 2008). Cognitive processes, and in particular, the processes of concept formation (Gabora and Aerts 2002; Aerts, Broekaert,and Gabora 1999; Aerts and Gabora 2005), and of decision making (Busemeyer & Wang 2007; Khrennikov 2009) have been modeled using quantum formalisms. Semantics and Information retrieval are other areas that have also been increasingly modeled using quantum formalisms (Van Rijsbergen 2004; Widdows 2004; Bruza & Cole 2005; Nelson & McEvoy 2007; Bruza et al. 2008).

Like CHRRs, the representation space in Quantum Theory is a Hilbert space on a complex field. Are there relationships between QT and the representational system outlined in this paper that may lead to interesting insights and be usefully applied?

Van Rijsbergen (2004) describes a framework for unifying the vector space, probabilistic and logical models of information retrieval such that the reasoning that occurs within each of the models is formulated algebraically and can be shown to depend on the geometry of the information space. The approach taken is to use the mathematics of Hilbert spaces in a way which is very similar to its use in quantum mechanics. CHHRs would seem to provide a practical implementation of some of the ideas expressed within this unifying framework. For example, the state of a word (or other linguistic unit) in memory may be represented as, for example, a 1000 dimensional complex valued vector and can be written as $|v> \in \mathbb{C}^{1000}$. It is constructed from the weighted superposition of context and structure vectors that are encountered during the processing of text. Probabilities can be associated with subspaces and a language of conditional logic may be able to be constructed (Rijsbergen 2004).

Bruza & Cole (2005) relates the different senses of a word to the different eigenstates of the semantic representation. The eigenvalues are related to the probabilities of collapse when a quantum measurement is performed. In a Semantic Space using CHRRs, an observable such as semantic similarity, structural similarity, or both, could be measured via the application of a self-adjoint linear operator.

Can we think of the evolution of semantic CHRRs in the same way as that of a quantum state? What sought of operators may be employed? (see Gabora and Aerts 2009)

Reference should also be made to recent connections identified between HRRs in the frequency domain and Geometric Algebra in the context of quantum computation (Aerts, Czachor and De Moord 2009). This should be a subject of future research.

Interference

A primary effect in wave-like phenomena is the production of interference when two or more waves combine. When

waves are in phase then there is constructive interference, while when they are out of phase there is destructive interference. When CHRRs are superposed the frequency components that are in phase (or close to each other in phase value) resonate and are amplified while those that are out of phase destroy or inhibit each other.
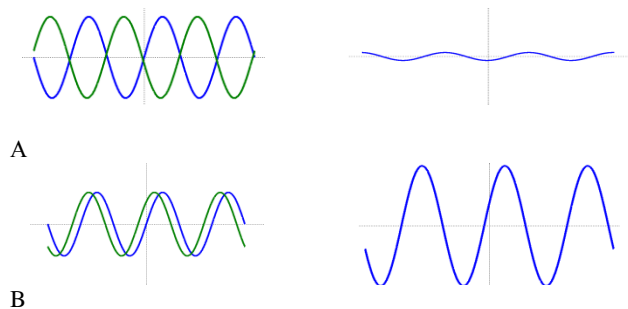

A

B

Figure 6. A – Destructive interference between nearly orthogonal waves. B – Constructive interference between waves with similar phase.

In the context of finding vectors that are nearest neighbors of a query vector, the negation operation described and exemplified above (bank not savings) can be thought of as using destructive interference to destroy or reduce the influence of the vector which is negated.

In a recent paper Zuccon, Azzopardi, and Rijsbergen introduced the Quantum Probability Ranking Principle (Zuccon, Azzopardi, and Rijsbergen, 2009) (QPRP) for information retrieval in which interference effects between documents are used to compute optimal ranking solutions for document retrieval. It is proposed that this principle be used instead of the Probability Ranking Principle (PRP) when documents cannot be assumed to be independent of each other. We propose that the representations described in this paper may provide a very natural and intuitive method for implementation of this principle.

Conclusion

We have described a novel and efficient approach to computing a semantic space via the use of complex valued vector representations. This approach builds upon previous work in Holographic Reduced Representations and their application to Semantic Spaces. We have shown that the proposed method can be practically implemented and have provided results from computational experiments. These results indicate that further research and experiments in this area are warranted and that rich conceptual relationships exist between the proposed method and other research areas such as optical holography and quantum interaction. A more exhaustive and rigorous set of investigations are required to compare the proposed method with existing methods.

References

Achlioptas, D.; 2001. Database-friendly random projections. *Symposium on principles of database system*s.

Aerts, D.; Broekaert, J.; and Gabora, L. 1999. Nonclassical contextuality in cognition: Borrowing from quantum mechanical approaches to indeterminism and observer dependence. In Campbell, R., ed., *Proceedings of 'Mind IV'*,Dublin, volume 10.

Aerts, D.; Aerts, S.; Broekaert, J.; and Gabora, L. 2000. The Violation of Bell Inequalities in the Macroworld. *Foundations of Physics* 30:1387–1414.

Aerts, D., Czachor, M., and De Moord, B. 2009. Geometric analogue of holographic reduced representation. *Journal of Mathematical Psychology*. 53(5), 389-398.

Aerts, D, and Gabora, L. 2005a. A theory of concepts and their combinations II: A Hilbert space representation. *Kybernetes*, 34(1-2), 192–221.

Borsellino, A., and Poggio, T. 1973. Convolution and correlation algebras. *Biological Cybernetics*, 13(2), 113–122.

Bruza, P., and Cole, R. 2005. Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning. In Artemov, S.; Barringer, H.; d'Avila Garcez, A.; Lamb, L.; and Woods, J., eds., *We Will Show Them: Essays in Honour of Dov Gabbay*, volume 1. College Publications. 339–361.

Bruza, P.; Kitto, K.; Nelson, D.; and McEvoy, C. 2008. Entangling words and meanings. In *QI2008 Proceedings*.

Busemeyer, J. R., and Wang, Z. 2007. Quantum Information Processing Explanation for Interactions between Inferences and Decisions. In Bruza et al. (2007).

Firth, J. R. 1957. A synopsis of linguistic theory 1930-1955. *Studies in linguistic analysis*, 51: 1-31.

Gabor, D. 1968. Holographic model for temporal recall. *Nature*, 217:1288--1289.

Gabora, L., & Aerts, D. 2009. A model of the emergence and evolution of integrated worldviews. Journal of Mathematical Psychology.

Gärdenfors, P. 2004. *Conceptual spaces: The geometry of thought*. The MIT Press.

Harris, Z. 1954. Distributional structure. *Word*, 10(23): 146-162.

Hecht-Nielsen, R. 1994. Context vectors; general purpose approximate meaning representations self-organized from raw data. In *Zurada, J. M.; R. J. Marks II; C. J. Robinson; Computational intelligence: imitating life*. IEEE Press.

Humphreys, M. S.; Bain, J. D., and Pike, R. 1989. Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, 96(2), 208–233.

Johnson, W. B., and Lindenstrauss, J. 1984. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26(189-206), 1–1.

Jones, M. N.; Kintsch, W., and Mewhort, D. J. 2006. High-dimensional semantic space accounts of priming. *Journal of memory and language*, 55(4), 534–552.

Jones, M. N., and Mewhort, D. J. 2007. Representing word meaning and order information in a composite holographic lexicon. *Psychological review*, 114(1): 1-37.

Kanerva, P. 1988 *Sparse distributed memory*. The MIT Press.

Kanerva, P.; Kristofersson, J. and Holst, A.. 2000. Random Indexing of text samples for Latent Semantic Analysis. *Proceedings of the 22nd annual conference of the cognitive science society*. New Jersey: Erlbaum.

Kanerva, P. 2009. Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1(2): 139-159.

Khrennikov, A. (2009). Quantum-like model of cognitive decision making and information processing. *Biosystems*, 95(3), 179–187.

Kitto, K. 2008. High End Complexity. *International Journal of General Systems*.

Landauer, T., and Dumais, S. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition. *Psychological Review* 104(2), 211–240.

Landauer, T. K.; Foltz, P. W. and Laham, D. 1998. An introduction to latent semantic analysis. *Discourse Processes, 25,* 259–284.

Landauer, T. K. 2002. On the computational basis of learning and cognition: Arguments from LSA. *The psychology of learning and motivation: Advances in research and theory*, 41, 43–84.

Lowe, W. 2001. Towards a theory of semantic space. In *J. D. Moore and K. Stenning (Eds.) Proceedings of the Twenty-first Annual Meeting of the Cognitive Science Society* LEA. 576-581.

Lund, K., and Burgess, C. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods Instruments and Computers*, 28(2), 203–208.

Nelson, D., and McEvoy, C. L. 2007. Entangled associative structures and context. In *Proceedings of the AAAI Spring Symposium on Quantum Interaction*. AAAI Press.

Plate, T. A. 1991. Holographic reduced representations: Convolution algebra for compositional distributed representations. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence.* 30–35.

Plate, T. A. 1994. Distributed Representations and Nested Compositional Structure. Ph.D. diss., Graduate Dept. of Computer Science. University of Toronto.

Pribram, K.H. 1966. Some dimensions of remembering: Steps toward a neuropsychological model of memory. In *Gaito, J. (Ed.), Macromolecules and Behaviour*. Academic Press.

Recchia, G. L., Jones, M. N., Sahlgren, M., & Kanerva, P. 2010. Encoding sequential information in vector space models of semantics: Comparing holographic reduced representations and random permutations. In *S. Ohisson & R. Catrambone (Eds.), Proceedings of the 32nd Annual Cognitive Science Society*. Austin TX: CSS

Sahlgren, M. 2005. An Introduction to Random Indexing. *In Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering*.

Sahlgren, M., Holst, A., and Kanerva, P. 2008. Permutations as a means to encode order in word space. *In Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci'08)*, July: 23-26.

Van Rijsbergen, C. J. 2004. *The geometry of information retrieval*. Cambridge Univ Press.

Widdows, D. 2004. *Geometry and Meaning*. CSLI Publications.

Widdows, D., and Ferraro, K. 2008. Semantic vectors: a scalable open source package and online technology management application. *In Sixth International Conference on Language Resources and Evaluation* (LREC 2008).

Widdows, D., and Cohen, T. 2009. Semantic vector combinations and the synoptic gospels. *Quantum Interaction*, 251–265.

Widdows, D, and Peters, S. 2003. Word vectors and quantum logic: Experiments with negation and disjunction. *Mathematics of language*, 8, 141–54.

Willshaw, D. J.; Buneman, O. P., and Longuet-Higgins, H. C. 1969. Non-holographic associative memory. *Nature*, 222(7), 960–962.

Zuccon, G.; Azzopardi, L. A., and Rijsbergen, K. 2009. The Quantum Probability Ranking Principle for Information Retrieval. In *Proceedings of the 2nd International Conference on Theory of Information Retrieval: Advances in Information Retrieval Theory*, 240.