Similarity Metrics within a Point Of View

Sven Aerts[1], Kirsty Kitto[2], and Laurianne Sitbon[2]

[1] Centre for Interdisciplinary Studies CLEA,
Vrije Universiteit Brussel
saerts@vub.ac.be
[2] Faculty of Science and Technology,
Queensland University of Technology
{kirsty.kitto,laurianne.sitbon}@qut.edu.au

Abstract. In vector space based approaches to natural language processing, similarity is commonly measured by taking the angle between two vectors representing words or documents in a semantic space. This is natural from a mathematical point of view, as the angle between unit vectors is, up to constant scaling, the only unitarily invariant metric on the unit sphere. However, similarity judgement tasks reveal that human subjects fail to produce data which satisfies the symmetry and triangle inequality requirements for a metric space. A possible conclusion, reached in particular by Tversky et al., is that some of the most basic assumptions of geometric models are unwarranted in the case of psychological similarity, a result which would impose strong limits on the validity and applicability vector space based (and hence also quantum inspired) approaches to the modelling of cognitive processes. This paper proposes a resolution to this fundamental criticism of of the applicability of vector space models of cognition. We argue that pairs of words imply a context which in turn induces a point of view, allowing a subject to estimate semantic similarity. Context is here introduced as a point of view vector (POVV) and the expected similarity is derived as a measure over the POVV's. Different pairs of words will invoke different contexts and different POVV's. Hence the triangle inequality ceases to be a valid constraint on the angles. We test the proposal on a few triples of words and outline further research.

Keywords: Similarity, Semantic Space, Triangle Inequality, Metric, POVV, Context

1 Introduction

A mental lexicon refers to the words that comprise a language, and its structure is defined here by the associative links that bind this vocabulary together. Such links are acquired through experience and the vast and semi-random nature of this experience ensures that words within this vocabulary are highly interconnected, both directly and indirectly through other words. For example, the word *planet* can become associated with *earth*, *space*, *moon*, and so on, and within this set, *moon* can become linked to *earth* and *star* [7]. Words are so associatively interconnected with each other that it takes only a few associative steps to move from one word to any other in the lexicon [23]. The complexity of the mental lexicon makes it highly challenging to construct analytical and computational models of both its structure and behavior. Yet even relatively small steps towards achieving the automatic interpretation of human language have given us search engines capable of converting our human made queries into their mathematical equivalent, and identifying documents relevant to that query among the huge corpus of the internet. Thus, these small steps have transformed the way we use the internet today. It seems clear that having a better mathematical representation of human language will lead to an improved use of the information content of the internet, however, the question of how to best represent human language remains a theoretical challenge. In this paper we shall consider one particular challenge, that of metricity. While vector space based models of the human mental lexicon have proven successful in various respects, the manner in which they quantize similarity is different from human judgements of semantic similarity, which violate important properties op a metric[25]. We shall then propose a contextual resolution to this problem and conclude by suggesting a number of potential future avenues of investigation. We begin with a brief overview of current vector space models of the mental lexicon.

2 Vector Space Models of the Mental Lexicon

Computational representations of the mental lexicon have been investigated by researchers from a range of disciplines, including mathematics, logic, philosophy, artificial intelligence, computational linguistics, cognitive psychology, natural language processing and information retrieval [24]. The birth of vector space based models (VSBM) for the purpose of information retrieval can be traced back to the seminal paper of Salton et al. [20] who were searching for an appropriate mathematical space to represent documents. Starting from a few basic desiderata, they settled upon a vector in a high dimensional vector space as an appropriate representation of a document. Within this framework, a query is treated like a small (pseudo) document that is also converted to vector form. The documents in the corpus are then ranked according to their distance to the query; closer documents are considered more relevant than ones that are further away. The way was now open to include Boolean operators on the returned results, and thus the first search engines were born. One of the main drawbacks of this system was that it had trouble returning documents that would have been highly relevant if one of the words in the query was replaced by a synonym, and the next advance came from representing concepts latently in a so-called *semantic space* where they are not formally represented or labelled. Semantic spaces are instances of vector spaces, and represent words in a basis created from other words, concepts, documents, or topics. They are generally built from the observation of co-occurrences in large text corpora. In word spaces such as the Hyperspace Analogue to Language (HAL) [21] the basis consists of every word in the vocabulary. Thus, the vector for a given word W is calculated by summing the number of occurrences of word $W(i)$ in a given context window around each occurrence of W and writing that number at the position i in the vector that represents W. This number can be adjusted using the distance (defined in terms of the number of words) or mutual information measures such as Point-Wise Mutual Information, which allows for a weighting of the importance of the word at that position. It is also possible to take word order into account [12, 19] . The major evolution with respect to the original proposal of Salton et al., was to derive a more fundamental semantic value through a reduction of the initial word space using mathematical tools such as Singular Value Decomposition ([13]), Non Negative Matrix factorization ([14]), or random projection ([18]), all of which generate a new basis that is greatly reduced in the number of dimensions. This new basis can under certain conditions be naturally related to topics, objects and concepts [14]. Because of the dimensional reduction, words with similar meaning tend to cluster into single dimensions of the resulting reduced vector space, greatly reducing the problems the old VSBM had with synonyms.

Once a semantic space has been created, we need to rank the results returned by a query using a similarity measure. Several distance measures (such as cosine similarity, Euclidean distance, and the City Block metric [8]) have been applied to semantic analysis, all of which supposedly measure the similarity between words in a given space. The most popular of these used in semantic analysis is cosine similarity, which gives the angle between two vectors in semantic space. We will later explain why this is generally considered a good choice. A number of studies have shown that semantic spaces can be effective at performing tasks that are human like. For example they have shown success at synonymy detection, categorization tasks, information retrieval and query expansion [24]. They have also been shown to perform well at mimicking human word association norms [27]. As we will show later, this has led a number of researchers to propose semantic spaces as models of human cognition. In this paper we examine important issues related to such a move. Semantic spaces are metric spaces and this poses problems that must be resolved before they can become viable models of human cognition. We shall begin with a discussion of metric spaces and in particular of the properties that a set must satisfy before it can be identified as a metric space. We shall then proceed to a discussion of the way in which human behavior violates these conditions and propose a possible resolution to this problem in later sections.

2.1 Motivation for the use of the angle as a measure of similarity

It is notoriously difficult to formally describe the notion of *meaning*. Yet this precisely what Natural Language Processing aims for. VSBM solve this issue via the so-called *distributional hypothesis*, which claims that *words which occur in similar contexts tend to have similar meanings* [11, 10, 9]. In VSBM, the entries of the vectors are usually monotone functions of the frequency of co-occurrence.

Hence vectors that are "close" occur in similar contexts and, by the distributional hypothesis, ought to have similar meanings. Using the inner product or cosine measure as a representation of similarity then seems like a very plausible suggestion. There are good mathematical reasons as well. If the vectors that correspond to a word are represented by unit vectors, the state space of words becomes the unit sphere. The unit sphere is a simple example of a manifold and geodesics on this manifold are well known to correspond to great circles. On the unit circle, the length of a great circle between two points equals the angle expressed in radians. Indeed, we have that the angle between two points on the sphere is (up to constant scaling) the only unitarily invariant Riemann metric on the sphere [28]. But what precisely are the mathematical criteria for a function to be a bona fide distance function?

2.2 Requirements for a Metric Space

In this section we shall briefly sketch the requirements for a metric space before proceeding in the next section to a discussion of the manner in which semantic data obtained from humans tends to violate the requirements for metric spaces.

Definition 1. *The ordered couple* (M, d) *with* M *a non emtpy set and* $d : M \times M \to \mathbb{R}$ *a function (called the distance or metric), is called a* metric space *if for any* $i, j, k \in M$*, the following hold:*

1. Non-negativity: the distance between two points must be greater than or equal to zero:

$$d(i, j) \geq 0. \tag{1}$$

2. Identity of indiscernibles: if the distance between two points is equal to zero then those two points are the same:

$$d(i, j) = 0 \Leftrightarrow i = j. \tag{2}$$

3. Symmetry: the distance between two points is equal, regardless of which direction it is measured in:

$$d(i, j) = d(j, i). \tag{3}$$

4. The Triangle Inequality: for three points in M, the distance from i to k is less than the distance which goes via j:

$$d(i, j) + d(j, k) \geq d(i, k), \tag{4}$$

Many authors prefer to list 1 and 2 in a single requirement. In fact, requirement 1 can be derived easily from 2,3 and 4. It is straightforward to verify that the angle α_{ij} between vectors u_i and u_j

$$\alpha_{ij} = \cos^{-1} \frac{\langle u_i, u_j \rangle}{|u_i||u_j|} \tag{5}$$

satisfies all four requirements. The angle between two vectors seems to be in accordance with the distributional hypothesis and satisfies all qualities of a mathematical metric. Moreover, its use has been tested in a wide variety of applications. As such we seem to have a very fundamental and valuable quantity. But the most important question is perhaps how we humans judge semantic similarity. This is a question that belongs to cognitive science so we shall now turn to an examination of similarity in this field, contrasting its results with those of VSBM.

3 Are Semantic Spaces Good Models of Human Cognition?

Vector spaces have been at the heart of many models in cognitive science. One of the more important examples for our purpose, is prototype theory. The basic idea of prototype theory is that some members of a category are more 'typical' than others [17]. For example, a cat is a more (prototypical) member of the category pet, whereas a donkey is clearly more peripheral. This idea is called 'graded categorization' and was formalized by representing concepts as vectors and categories as sets of vectors [15, 22]. However, these vectors are not based on co-occurrence, but on subjective numerical scores obtained by questioning human subjects. In this section we shall draw attention to a range of human derived data which violates a number of the properties that must be satisfied by a metric. We shall go through them in the order given in the previous section. The first requirement listed above is non-negativity. This is probably the least problematic of all requirements. Whether or not negative values of similarity occur, is decided by the questionnaire's scale on which human subjects are asked to judge similarity. Humans can quite naturally associate a concept of distance between two words as a measure of their similarity and this distance can be straight-forwardly assumed to be non-negative. However, in this section we shall show that every other requirement of a metric space can be quickly violated by spatial representations of similarity data.

3.1 Homographs and the non-identity of indiscernible

The identity of indiscernibles property implies that different words should be separated by some distance. However, many languages contain words with multiple meanings, multiple words for the same thing, and ambiguous structures, and these properties give us reason to question the general validity of this property, although some features of natural language support this property. For example, synonyms (different words for the same thing) appear to satisfy the identity of indescernibles property reasonably well; while they lie close together semantically synonyms generally have slightly different connotations. Thus, while 'student' and 'pupil' both mean essentially the same thing, there are slightly different senses to these two words, and hence they tend to appear close together, but with some distance separating them in most semantic spaces.

However, homographs create much more serious problems for attempts to generate a metric space. Homographs are words that have the same spelling and pronunciation but different meanings. For example, 'bat' is a homograph, as it has at least two senses: (1) as a small furry flying mammal; and (2) as a sporting implement.

Homographs pose a problem for the *if and only if* criterion in property 2. If we generate a set that represents each word in English, then 'bat' should appear only once in it ($i = j$); however, semantic spaces tend to correctly reveal the different meanings behind this word by using a mixture of the representation of both words. Thus, property 2 seems to pose a challenge for semantic space approaches a discernible words (such as 'bat' for sports and 'bat' the animal) are represented at exactly the same point in the space. We believe a finer resolution of homographs in semantic space is possible by examining the set of documents that contain the words. First a search in, for example, Wordnet will reveal if a word has several meanings and if so, how many. Say a word has n possible meanings. Then we ought to divide the set of all the words that substantially co-occur with the query word, into n sets of words such that each set shows a degree of cohesion in the words that co-occur with it. This may be implemented by an appropriate algorithm that reduces to n the dimension of the matrix that has as its rows the words that co-occur and as column the documents in which they occur. Interestingly, a very similar situation occurs in quantum mechanics in the case of degenerate energy levels. An energy level of a quantum system is called degenerate if different states correspond to the same energy level. If we think of the energy level of the system as 'the name' of the state that corresponds to that energy level, we have an analogy with homographs. Application of a well chosen perturbation to the Hamiltonian of the system allows us to separate the energy levels, so each energy level corresponds in a unique way to an energy level. We say that the perturbing field is 'lifting the degeneracy' and splits the energy level into finer energetic detail. If we see a separation of the two meanings of a single word in the semantic vector space, it seems we have provided enough context in the semantic space to lift the degeneracy of meanings corresponding to a single word. In an actual task of information retrieval, it is very valuable to be able to identify which meaning is more probable for a given word in a given context. For this we would have to judge to which of two statistical clusters a given vector (word) in a given context belongs. Language is extremely flexible and is perfectly able to shift perspective as we include more context, thereby changing the meaning. Take as an example, the word 'hits'. Without additional context, its meaning is degenerate; it could mean many things. We are then given a piece of context: 'Michael hits Billy'. Most probably 'hits' denotes a physical act of violence. We are then given an additional piece of context: 'Michael Jackson hits Billy Jean'. The meaning of 'hits' is now more likely to signify a musical hit. We are given a last piece of context: 'Michael Jackson number of Google hits for Billy Jean', the word 'hits' denotes the webpages Google relates to a query. In the example above every new level of context only adds words to the previous context; the previous context isn't changed in form, only in meaning. We feel the nature of language is simply too rich to allow for a strict separation, but VSBM do seem capable of at least statistically approaching the problem of homographs.

3.2 Human Similarity Judgements are not Symmetric

It was shown by Tversky that human similarity judgements are asymmetric, and so directly violate the symmetry requirement of metric spaces (i.e. $d(a,b) \neq d(b,a)$) [26]. A classic example was first provided by Rosch in her theory of prototypes [16], which shows that humans have a tendency to declare similarity with respect to an archetype. For example, when asked to give an example of the concept furniture, humans will much more frequently cite a “chair” than a “stool”, and this archetypical concept (“chair”) is the one that similarity judgements are preferentially, and asymmetrically, assigned by. Thus, the similarity of stool to chair is usually deemed to be greater than that of chair to stool, the similarity of North Korea to China is judged greater than the similarity of China to North Korea [25, 26], and pink is deemed more similar to red, than red is to pink. There seems to be a genuine linguistic phenomenon here that one eventually would like to model. Of course, these experiments are designed to test for asymmetry; experiments that do not show asymmetry are equally easy to design. Suppose we produce a deck of cards with on each card nothing but the two words “red” and “pink”. However, on half of the cards the word “red” is printed above the word “pink”, on the other half, “pink” is printed above “red”. Each test subject is given one card and asked to quantify the similarity of the two concepts printed on the card. The result will obviously be symmetrical, because there was no distinguished order of words on the deck of cards. For our present purpose, we will assume symmetrical data.

3.3 Human Similarity Judgements Violate the Triangle Inequality.

Finally, human similarity judgements do not appear to satisfy the triangle inequality, a result shown by Tversky & Gati [25]. Indeed, the contrast between human similarity judgements and distance notions in geometric models of cognition led them to conclude that ([25], p 153):

> *some basic properties of the geometric model (e.g., translation invariance, segmental additivity, and the triangle inequality), which enhance the interpretability and the appeal of spatial representations, cannot always be accepted as valid principles of psychological similarity.*

even before Semantic Space approaches to the mental lexicon were invented.

If Tversky & Gati are correct then their criticism poses some very serious problems for both semantic space models, and hence quantum inspired models of the human mental lexicon. To put things in perspective, semantic spaces were developed and successfully put to use in spite of this problem, so perhaps we need not worry too much. On the other hand we would like to be able to model subjective similarity as it seems an important component of natural language processing. What makes the triangle inequality problem more severe than the three previous requirements we discussed, it that we cannot make it go away by devising another experiment, at least not straightforwardly. If we want symmetric or non-negative data we can always make sure that the experiment will give us only positive values. For non-negativity we need only to constrain the range of the possible answers; for the symmetry condition, we need only to make sure every couple's similarity is symmetric. Indeed, if $d(a,b) = d(b,a)$ and $d(b,c) = d(c,b)$, then obviously $d(a,c) = d(c,a)$. Can we design an experiment in such a way that it always satisfies the triangle inequality? We could give concepts in triples to subjects and ask them to draw a triangle with the three words on the vertices of the triangle and express the relative similarities by the relative lengths of the sides of the triangle. The triangle inequality would be trivially satisfied for this triple. However, if we have several triples that satisfy the triangle inequality, then there is no guarantee whatsoever, that from these triples we cannot pick words to form new triples that will violate the triangle inequality. Another proposal would be to abandon metric spaces, or geometric models for the representation of cognitive entities such as concepts and sentences. If we take into consideration the huge success this class of models has enjoyed then this seems like a rather radical step to take. An alternative answer to Tversky & Gati might be found through an adoption of the notion of context, and in what follows we shall start to develop an approach within a metric space that can recover the non-metric behavior of human judgements of similarity.

4 The point of view model

In vector space based accounts of cognition (such as quantum theory inspired approaches [1–3, 6, 5]) concepts are very often represented by unit vectors in a Hilbert space. Take three unit vectors u_1, u_2 and u_3 that represent three concepts. Call θ_{ij} the angle between u_i and u_j:

$$\cos\theta_{ij} = \langle u_i, u_j \rangle. \tag{6}$$

Because Hilbert space is a metric space, this has consequences for the possible range of values the angles between the vectors can assume:

$$|\theta_{ij} - \theta_{jk}| \leq \theta_{ik} \leq |\theta_{ij} + \theta_{jk}|. \tag{7}$$

The new model assumes that each time a subject is asked to quantify the similarity between two concepts they must take a stance, or a *point of view*, from which to judge their similarity. On an absolute scale we may argue that all concepts are very similar (they are, after all, just concepts) or we may argue no two concepts are alike. But if we are asked what the similarity is between *Moon* and *Ball,* we will not easily judge their similarity on an absolute scale. We rather inadvertently look for a proper context to judge their similarity. If our perspective is "*Shape*" then we will think of *Moon* and *Ball* as being somewhat similar. If the perspective would have been "*Play*", the two concepts would be judged rather dissimilar. So it is the two words, together with the state of the subject, that determine the point of view from where similarity will be judged. We model a point of view by assuming that for each pair of vectors u_i and u_j and a given subject S that is asked to judge their similarity, there is a *point of view vector* (POVV) u^S_{ij}. The cosine of the angle this observer sees between u_i and u_j, is:

$$\begin{aligned}\cos\alpha_{ij} &= \frac{\langle u_i - u^S_{ij}, u_j - u^S_{ijj}\rangle}{|\langle u_i - u^S_{ij}\rangle||\langle u_j - u^S_{ij}\rangle|} \\ &= \frac{\cos\theta_{ij} - \langle u^S_{ij}, u_j\rangle - \langle u_i, u^S_{ij}\rangle + |u^S_{ij}|}{|\langle u_i - u^S_{ij}\rangle||\langle u_j - u^S_{ij}\rangle|}\end{aligned} \tag{8}$$

In psychological experiments, the similarity is an average over many trials. The expected similarity is then derived as a measure over the POVV's. In what follows, we may assume that u^S_{ij} is already an averaged point of view in the sense that α_{ij} coincides with the average subjective similarity.

To determine which regions for u^S_{ij} lead to increased values of θ_{ij} and which lead to decreased values, we first look at the set of u^S_{ij} that leaves α_{ij} invariant.

Lemma 1. *Let $0, u_i$ and u_j be three non-collinear vectors and let C_{ij} be the circle that contains $0, u_i$ and u_j. Then for any $u^S_{ij} \in C_{ij}$ with $u^S_{ij} \neq u_i$ and $u^S_{ij} \neq u_j$ we have $\cos\alpha_{ij}{=}\cos\theta_{ij}$.*

Proof. The span u_i and u_j defines a two dimensional linear subspace containing the null vector. Let C_{ij} be the unique circle within this linear subspace that contains 0, u_i and u_j. By the inscribed angle theorem, the angle θ_{ij} inscribed in this circle does not change as its apex u^S_{ij} is moved to different positions on C_{ij}, hence $\alpha_{ij} = \theta_{ij}$. □

Let us call D_{ij} the open disk that is the interior of C_{ij}. It is easy to see a POVV inside D_{ij} yields an observed angle α_{ij} that is greater than θ_{ij}. The disk D_{ij} is an open convex set, so any open convex combination of 0, u_i and u_j is an element of D_{ij}. The maximal angle is reached for $u^S_{ij} = \frac{1}{2}(u_i + u_j)$, which clearly lies inside D_{ij}. The observed angle in this case, is:

$$\alpha_{ij} = \cos^{-1}\frac{\langle u_i - \frac{1}{2}(u_i+u_j), u_j - \frac{1}{2}(u_i+u_j)\rangle}{|\langle u_i - \frac{1}{2}(u_i+u_j)\rangle||\langle u_j - \frac{1}{2}(u_i+u_j)\rangle|} \tag{9}$$

$$= \cos^{-1}(-1) = \pi \tag{10}$$

So it is always possible to pick a POVV in D_{ij} that yields minimal similarity. This result makes sense geometrically: if your point of view is in the middle of the two concepts, then, to you, they couldn't be further apart from each other. For an intermediate situation, there are many possibilities. A particularly nice choice is to consider the $d-$parameter POVV that lies precisely between u_i and u_j and has length $d:\ u^S_{ij}(d) = \frac{d}{|u_i+u_j|}(u_i + u_j)$. If we consider the triangle which has as vertices $u_{ij}^S(d), o$ and u_i, the sine rule learns us that $\sin(\pi - \alpha_{ij}/2)/1 = \sin((\alpha_{ij} - \theta_{ij})/2)/d$, hence the relation between α_{ij}, θ_{ij} and d is given by:

$$d = \frac{\sin((\alpha_{ij} - \theta_{ij})/2)}{\sin(\alpha_{ij}/2)}. \tag{11}$$

We can get minimal similarity and intermediate values. It turns out the POVV constrains the maximum similarities. To see this, assume that

$$u_{ij}^S = -\frac{u_i + u_j}{|u_i + u_j|}. \tag{12}$$

This unit vector points in the direction opposite of $\frac{1}{2}(u_i + u_j)$. By the inscribed circle theorem, the observed angle α_{ij} is exactly $\theta_{ij}/2$. It turns out this is the minimal value for α_{ij} that the point of view model can attain; this minimum is reached if u_{ij}^S lies on the great arc between u_i and u_j of the unit circle. If u_{ij}^S lies on the unit circle, but on the short arc between u_i and u_j, then the quadruple of vectors $0, u_i, u_{ij}^S, u_j$ form a cyclic quadrilateral, from which we immediately get: $\alpha_{ij} = \pi - \theta_{ij}$.

4.1 The evocation data set

Let us provide a brief illustration of the model using data from the Evocation data set [4], collected by crowd sourcing using Amazon Mechanical Turk which allows for the quick collection of large amounts of data. The data was cleaned to the highest level of correlation with a smaller data set collected under controlled conditions. Users were asked how much a sense of a word brings the sense of another to mind (on a scale of 0 to 100), using the word themselves as well as a definition for disambiguation. The data for a pair of words are usually not symmetric, however for the purposes of this paper we have averaged the two similarities so that the resulting data is symmetric . In essence then, this data set contains human judgements of symmetrized semantic relatedness between pairs of words. For example, '*key*' and '*car*' were judged at 73% of similarity, '*car*' and '*light*' at 79,4% of semantic similarity, while '*key*' and '*light*' only at 14.3%. Other examples of triples that violate the triangle inequality from this data set include:

1. *night /day*: 86.3%, *day/year*: 62.8%, *night/ year*: 11.6%;
2. *school/university*: 83.7%, *university/court*: 73.2%, *school/court*: 7.6%;
3. *food/oil*: 81.5%, *oil/gold*: 62.8%, *food/gold*: 2.7%.

Let us take the first example and label three vectors with an index that refers to the concepts: u_n is the vector that corresponds to *night*, and likewise we denote u_d for the concept *day* and u_y for the concept *year*. We first convert the given similarities to angles using $\cos\theta_{ij} = \langle u_i, u_j \rangle$. Then $\theta_{nd} = 0.53; \theta_{dy} = 0.89$ and $\theta_{ny} = 1.45$. Clearly this triple violates the triangle inequality, e.g. $|\theta_{ny} - \theta_{dy}| = |1.45 - 0.89| = 0.56 \geq \theta_{nd} = 0.53$. Because the triangle inequality is violated, there do not exist three vectors with the prescribed angles. However, from the $d-$parametrized POVV for θ_{ny}, $u_{ij}^S(d) = \frac{d}{|u_n+u_y|}(u_n + u_y)$, we obtain: $|u_n + u_y| = 2\cos(\theta_{ny}/2) \approx 1.5$. The value of θ_{ny} was 1.45; if it would have been 1.42, no violation would have occurred. Hence we choose $d = \sin((1.42 - 1.45)/2)/\sin(1.42/2) = -.023$. So the POVV $u_{ij}^S(d) = \frac{-1}{60}(u_n + u_y)$ restores the triangle inequality for this triple. It is easy to see we could also have taken a triple of vectors that respect the inequality (e.g., the "restored" vectors above) and, when one of the angles is viewed upon from a suitably chosen POVV (e.g., the opposite vector of $u_{ij}^S(d)$ in the example above), the resulting angles will violate the inequality.

5 Concluding remarks

The question we addressed in this paper is whether it is possible for a semantic space to be a metric space and at the same time be able to capture the non-metric behavior of human similarity judgements. Another strongly related and perhaps even more interesting question is whether it is possible to derive a vector space using subjective similarity in stead of co-occurrence. We presented a model that gives an affirmative answer, in principle. Although the model we offered here was derived in an essentially *ad hoc* way, the model is falsifiable and we feel the case for this model could be made stronger if it can be shown a POVV can be derived from the semantic space itself. In order to sketch out a viable avenue for further work, we shall refer to one of our above examples. It is not peculiar that *day* and *year* are considered close, as they are both important measures of time. Neither is it strange that *day* and *night* are judged to be close, as they are in a certain sense opposite to one another. Note that someone who is being asked how close *day* and *night* are, will think of *day* in the sense of daytime, which is not the same meaning the word has when we compare *day* and *year*. The last couple in our triple is then *night* and *year*, which are not so obviously connected, hence the lower similarity rating. We see that when we are asked to weigh the words for similarity, we unconsciously look for a minimal context that contains the two concepts, and depending on the words, this will be a different context. This is what the POVV model attempts to capture. However, for the POVV model to be convincing, we need to show there is a connection between the POVV and the concepts we are dealing with. In particular, the vectors that correspond to the words and their semantically associated vectors should determine the POVV. In a sense, the POVV is a "centre of gravity of meaning": if all concepts contribute to the centre of gravity, then the POVV will approximately be the zero of the vector space and the triangle inequality will hold; if not, deviations will arise. An important observation is that the model as it is right now, does not specify a unique POVV, so how will we know an eventual linkage between pairs of words and POVV's is viable? A valid confirmation would require a statistically significant test that uses only a semantic network and no human similarity measures, and which can predict human violations of the triangle inequality for triples of words. Whether this avenue will prove fruitful is left for future research.

Acknowledgements This project was supported by the Marie Curie International Research Staff Exchange Scheme: Project 247590, "QONTEXT - Quantum Contextual Information Access and Retrieval". KK is supported by the Australian Research Council Discovery grant DP1094974.

References

1. Aerts, D., Gabora, L.: A theory of concepts and their combinations I: the structure of the sets of contexts and properties. Kybernetes 34, 151–175 (2005)
2. Aerts, D. and Gabora, L.: A theory of concepts and their combinations II: A Hilbert space representation. Kybernetes, 34, pp. 192-221 (2005)
3. Aerts, D.: Quantum structure in cognition. Journal of Mathematical Psychology, 53, 314-348 (2009)
4. Boyd-Graber, J., Fellbaum, C., Osherson, D., Schapire, R.: Adding dense, weighted connections to wordnet. In: Proceedings of the Third International WordNet Conference (2006)
5. Bruza, P.D., Kitto, K., Ramm, B., Sitbon, L.: The non-decomposability of concept combinations (2011), under review
6. Bruza, P.D., Kitto, K., Ramm, B., Sitbon, L., Blomberg, S., Song, D.: Quantum-like non-separability of concept combinations, emergent associates and abduction. Logic Journal of the IGPL (2010), in Press
7. Bruza, P., Kitto, K., Nelson, D., McEvoy, C.: Is there something quantum-like about the human mental lexicon? Journal of Mathematical Psychology 53, 362–377 (2009)
8. Bullinaria, J.A., Levy, J.P.: Extracting semantic representations from word co-occurrence statistics : a computational study. Behavior research methods 39(3), 510–526 (August 2007)
9. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science 41(16), 391–407 (1990)
10. Firth, J.R.: Papers in Linguistics 1934-1951. Oxford University Press, London (1957)
11. Harris, Z.: Distributional structure. Word 10(23), 146–162 (1954)
12. Jones, M.N., Mewhort, D.J.K.: Representing word meaning and order information in a composite holographic lexicon. Psychological review 114(1), 1–37 (2007)
13. Landauer, T., Dumais, S.T.: A solution to plato's problem : the latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological review 104(2), 211–240 (1997)
14. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401(6755), 788–791 (1999)
15. Nosofsky, R.M.: Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General 115(1), 39–57 (1986)
16. Rosch, E.: Cognitive Representation of Semantic Categories. Journal of Experimental Psychology 104, 192–233

17. Rosch, E., Lloyd, B.B. (eds.): Cognition and categorization. Erlbaum, Hillsdale, NJ. (1978)
18. Sahlgren, M.: An introduction to random indexing. In: Proceedings of Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering. Copenhagen, Denmark (2005)
19. Sahlgren, M., Holst, A., Kanerva, P.: Permutations as a means to encode order in word space. In: Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci'08). Washington D.C., USA (July 23-26 2008)
20. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of the ACM 18(11), 613–620 (1975)
21. Schütze, H.: Automatic word sense discrimination. Computational linguistics 24(1), 97–123 (1998)
22. Smith, E.E., Osherson, D.N., Rips, L.J., Keane, M.: Combining prototypes: A selective modification model. Cognitive Science 12(4), 485–527 (1988)
23. Steyvers, M., Tennenbaum, J.: The large scale structure of semantic networks: statistical analyses and a model of semantic growth. Cognitive Science 21, 41–78 (2005)
24. Turney, P.T., Pantel, P.: From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research 37, 141–188 (February 2010)
25. Tversky, A., Gati, I.: Similarity, separability, and the triangle inequality. Psychological Review 89(2), 123–54 (1982)
26. Tversky, A.: Features of similarity. Psychological Review 84(4), 327–352 (1977)
27. Veksler, V.D., Govostes, R.Z., Gray, W.D.: Defining the dimensions of the human semantic space. In: V. Sloutsky, B.L., McRae, K. (eds.) 30th Annual Meeting of the Cognitive Science Society. pp. 1282–1287. Cognitive Science Society (2008)
28. Wooters, W.K.: The Acquisition of Information from Quantum Measurements. Ph.D. thesis, University of Texas at Austin (1980)