Semantic Judgement of Medical Concepts: Combining Syntagmatic and Paradigmatic Information with the Tensor Encoding Model

Mike Symonds[1], **Guido Zuccon**[2], **Bevan Koopman**[1,2], **Peter Bruza**[1], **Anthony Nguyen**[2]

michael.symonds@qut.edu.au, guido.zuccon@csiro.au, bevan.koopman@csiro.au
p.bruza@qut.edu.au, anthony.nguyen@csiro.au

[1] School of Information Systems, Queensland University of Technology
[2] Australian e-Health Research Centre, CSIRO

Brisbane, Australia

Abstract

This paper outlines a novel approach for modelling semantic relationships within medical documents. Medical terminologies contain a rich source of semantic information critical to a number of techniques in medical informatics, including medical information retrieval. Recent research suggests that corpus-driven approaches are effective at automatically capturing semantic similarities between medical concepts, thus making them an attractive option for accessing semantic information.

Most previous corpus-driven methods only considered *syntagmatic* associations. In this paper, we adapt a recent approach that explicitly models *both syntagmatic and paradigmatic* associations. We show that the implicit similarity between certain medical concepts can only be modelled using paradigmatic associations. In addition, the inclusion of both types of associations overcomes the sensitivity to the training corpus experienced by previous approaches, making our method both more effective and more robust. This finding may have implications for researchers in the area of medical information retrieval.

1 Introduction

Semantic similarity measures are central to several techniques used in medical informatics, including: medical search (Voorhees and Tong, 2011; Cohen and Widdows, 2009), literature-based discovery (e.g., drug discovery (Agarwal and Searls, 2009)), clustering (e.g., gene clustering (Glenisson et al., 2003)), and ontology construction or maintenance (Cederberg and Widdows, 2003).

Automatically determining the similarity between medical concepts presents a number of specific challenges, including vocabulary mismatch. For example, the phrases *heart attack* and *myocardial infarction* are synonymous, referring to the same medical concept. Beyond vocabulary mismatch are situations where semantic similarity is based on implied relationships, for example the mention of an organism (e.g. *Varicella zoster virus*) suggests the presence of a disease (e.g. *chickenpox*).

Existing approaches for measuring medical semantic similarities fall into two major categories: (i) those that utilise path-based measures between concepts in medical thesauri/ontologies, and (ii) corpus-based approaches that derive similarity judgements from the occurrence and co-occurrence of concepts within text, e.g., using Latent Semantic Analysis (LSA).

Research comparing path-based methods with corpus-based methods highlighted the ability for corpus-based methods to provide superior performance on medical concept similarity judgements (Pedersen et al., 2007). However, research evaluating eight different corpus-based approaches found that the performance was sensitive to the choice of training corpus (Koopman et al., 2012). This finding means it is difficult to apply a corpus-based approach which is both robust and effective.

It is important to note that the corpus based approaches referred to in this paper do not rely on syntactic information found in extra-linguistic resources, and use solely the co-occurrence statistics of words found in natural language to model word associations. Therefore, research modelling semantic associations using part of speech (POS)

Michael Symonds, Guido Zuccon, Bevan Koopman, Peter Bruza and Anthony Nguyen. 2012. Semantic Judgement of Medical Concepts: Combining Syntagmatic and Paradigmatic Information with the Tensor Encoding Model. In *Proceedings of Australasian Language Technology Association Workshop*, pages 15−22.

taggers, parsers or hand-coded resources are not within the scope of this work.

Within this paper we adapt a novel corpus-based approach, known as the *tensor encoding* (TE) model (Symonds et al., 2011a), for use in judging the similarity of medical concepts. The TE approach explicitly models the two types of word associations argued to give words their meaning within structural linguistic theory. These are: (i) syntagmatic and (ii) paradigmatic associations.

A syntagmatic association exists between two words if they co-occur more frequently than expected from chance. Two medical concepts that likely have a strong syntagmatic association include *bone* and *x-ray*.

A paradigmatic association exists between two words if they can substitute for one another in a sentence without affecting the grammaticality or acceptability of the sentence. Medical concepts that display synonymy, like *heart attack* and *myocardial infarction* display a strong paradigmatic association.

The TE model combines measures of syntagmatic and paradigmatic association within a single, formal framework. In this paper we demonstrate that not only does the TE model provide robust, superior performance across a wide variety of data sets when compared to past corpus-based approaches, but offers a flexible framework whose performance is not sensitive to the choice of training corpus. Our findings provide a robust and effective model for predicting semantic similarity between medical concepts, and also draws out useful statistical behavior relating to the modelling of syntagmatic and paradigmatic associations that exist within medical documents.

The remainder of this paper is set out as follows. Section 2 provides background on corpus-based approaches previously evaluated on medical concept similarity judgements. In Section 3 we describe the TE model and outline our novel variant for use in judging the similarity of medical concepts. Section 4 details the experiments to be used in evaluating the performance of the TE approach, with the results and their discussion following in Section 5. Concluding remarks and suggestions for future work are presented in Section 6.

2 Background

Corpus-based models that learn the relationships between words based on their distribution in natural language have a strong history in the field of natural language processing. Some of the most well-known include LSA (Latent Semantic Analysis (Landauer and Dumais, 1997)) and HAL (Hyperspace to Analogue of Language (Lund and Burgess, 1996)).

A rigorous evaluation of eight different corpus-based approaches on the task of judging medical concept similarity found that the best performance was achieved using a positive point-wise mutual information (PPMI) measure. This measure had an average correlation of ≈ 0.7 with judgements made by expert human assessors (Koopman et al., 2012).

PPMI is a variation of PMI where negative values are substituted by zero-values. The strength of PMI between word q and w within a stream of text can be expressed as:

$$S_{\text{ppmi}}(q,w) = \begin{cases} \log\left[\frac{p(q,w)}{p(q)p(w)}\right] & \text{if } \log\left[\frac{p(q,w)}{p(q)p(w)}\right] > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $p(q,w)$ is the joint probability of q and w, and $p(q)$, $p(w)$ are the expected probabilities of q and w respectively. In practice, these probabilities are computed as:

$$p(q,w) = \frac{|D_q \cap D_w|}{|D|},$$

$$p(q) = \frac{|D_q|}{|D|}, \quad p(w) = \frac{|D_w|}{|D|},$$

where D_q is the set of documents containing term q and D is the set of documents in the collection.

Although the PPMI measure achieved the best average performance across a number of test sets, it displayed a high degree of sensitivity when the training corpus was changed, thus reducing its overall utility.

Next, we introduce the tensor encoding model as a robust and effective alternative to previous proposed measures.

3 The Tensor Encoding Model

A recent corpus-based approach, known as the *tensor encoding* (TE) model, was originally presented as a model of word meaning (Symonds et al., 2011a), and later used to provide a flexible approach to semantic categorisation (Symonds et al., 2012). The TE model provides a formal framework for combining two measures that explicitly model syntagmatic and paradigmatic associations between words.

As the TE model has a strong theoretical basis in linguistics, it has potential applications in other areas that deal with natural language, including e.g. information retrieval. To demonstrate, the TE model was used to perform similarity judgements within the query expansion process of an ad-hoc information retrieval task (Symonds et al., 2011b). This method, known as *tensor query expansion* achieved robust and significant performance improvements over a strong benchmark model. This result was attributed to the inclusion of information about paradigmatic associations, which are not effectively modelled in existing information retrieval systems. Similarly, these paradigmatic associations are not explicitly modelled in previous corpus-based measures of medical concept similarity. We hypothesise that the inclusion of paradigmatic associations in semantic similarity measures would better capture similarities between medical concepts.

To support this insight, consider how the PPMI measure in Equation (1) is oblivious to paradigmatic associations that may exist between two words. If word q and word w do not co-occur in any documents (i.e., $|D_q \cap D_w| = 0$) then $S_{\text{ppmi}}(q, w) = 0$. This result suggests q and w are unrelated. However, consider a toy example using *heart attack*(q) and *myocardial infarction*(w). One clinician may use the first concept exclusively in a document, while another may use the second term exclusively. If the PPMI score between *heart attack* and *myocardial infarction* was calculated using these two example documents, the score would be zero and the two concepts considered unrelated.

From a structural linguistic viewpoint, one might say there are no syntagmatic associations between the two concepts, as they do not co-occur in the same context (i.e., medical report). Therefore, PPMI only captures syntagmatic associations, and hence fails to model any paradigmatic information that may exist.

However, consider the same example using the TE model's paradigmatic measure, and hence a pure paradigmatic perspective. Within the TE model, the strength of paradigmatic associations between two medical concepts, q and w can be defined as:

$$S_{\text{par}}(q, w) = \sum_{i=1}^{N} \frac{f_{\overline{iq}} . f_{\overline{iw}}}{\max(f_{\overline{iq}} , f_{\overline{iw}} , f_{\overline{wq}})^2}, \quad (2)$$

where $f_{\overline{iq}}$ is the unordered co-occurrence frequency of concepts i and q, $f_{\overline{iw}}$ is the unordered co-occurrence frequency of concepts i and w, and N is the number of concepts in the vocabulary.

Intuitively, Equation (2) enhances the score for concept w if q and w co-occur with the same concepts, independent of whether q and w occur within the same document. In fact this measure has a factor, $\frac{1}{f_{\overline{wq}}}$, that reduces the paradigmatic score if concepts q and w occur within the same document often. In our simple example, this would mean that if *heart attack* and *myocardial infarction* co-occurred with any of the same terms (e.g., CPR, chest pain, etc.) then they would have a paradigmatic score greater than 0.

The fact that pure paradigmatic information is not currently utilised within most corpus-based approaches leads us to hypothesise that more robust performance on medical concept similarity judgements can be achieved by adding paradigmatic information to the similarity estimate. In the remainder of this paper the measure in Equation (2) will be referred to as **PARA**.

The TE model uses a Markov random field to formalise the estimate of observing one concept w given a second concept q:

$$P(w|q) = \frac{1}{Z} \left[\gamma S_{\text{par}}(q, w) + (1 - \gamma) S_{\text{ppmi}}(q, w) \right], \quad (3)$$

where $\gamma \in [0, 1]$ mixes the paradigmatic $S_{\text{par}}()$ and syntagmatic $S_{\text{ppmi}}()$ measures, and Z normalises the resulting distribution. We refer the interested reader to Symonds et al. (2012) for details.

The estimate in Equation (3) can be reduced to the following rank equivalent measure of semantic similarity between q and w:

$$S_{\text{TE}}(q, w) \propto \gamma S_{\text{par}}(q, w) + (1 - \gamma) S_{\text{ppmi}}(q, w). \quad (4)$$

In the remainder of this paper the model defined in Equation (4) will be referred to as **TE**.

It is worth noting that the TE model formally supports the combining of any measure of syntagmatic and paradigmatic information. Therefore, if a more effective measure of syntagmatic or paradigmatic information is developed, it can be applied within the TE framework.

Corpus	# Docs	Avg. doc. len.	Vocab Size
TREC'11 MedTrack	17,198	5,010	54,546
OHSUMED	293,856	100	55,390

Table 1: Document collections (corpora) used.

Test: Corpus (data set)	Training: Corpus (data set)	γ	TE	PPMI
MedTrack (Ped)	OHSUMED (Ped)	0.5	$r = 0.6706$	$r = 0.4674$
MedTrack (Cav)	MedTrack (Ped)	0.5	$r = 0.6857$	$r = 0.6154$
OHSUMED (Ped)	OHSUMED (Cav)	0.2	$r = 0.7698$	$r = 0.7427$
OHSUMED (Cav)	MedTrack (Cav)	0.4	$r = 0.8297$	$r = 0.8242$

Table 2: Performance of TE using the γ produced by the specified train/test splits; performance of PPMI included for comparison.

4 Experimental Setup

In this section we outline the experimental setup used to evaluate the performance of the TE approach on two separate medical concept similarity judgement data sets. The first data set involves judging the similarity of 29[1] UMLS medical concept pairs. These were first developed by Pedersen et al. (Pedersen et al., 2007) and human assessments of semantic similarity were produced by 9 clinical terminologists (coders) and 3 physicians, with inter-coded relatedness equal to 0.85. Assessors scored each pair between 1 and 4, with 1 being unrelated and 4 being highly synonymous. This data set is indicated as **Ped** in the remainder of this paper.

The second data set is comprised of 45 UMLS concept pairs, developed by Caviedes and Cimino (2004), for which semantic similarity assessments were performed by three physicians. Similarities were scored between 1 and 10, with higher scores indicating a stronger similarity between concepts. This data set is indicated as **Cav** in the remainder of this paper.

Two separate corpora were used as data to prime all models; corpus statistics are shown in Table 1. The TREC MedTrack collection consists of documents created from concatenating clinical patient records for a single visit, while the OHSUMED collection is based on MEDLINE journal abstracts.

Following the procedure outlined by Koopman et al. (2012) the original textual documents for both corpora were translated into UMLS medical concept identifiers using MetaMap, a biomedical concept identification system (Aronson and Lang, 2010). After processing, the individual documents contained only UMLS concept ids. For example, the phrase *Congestive heart failure* in the original document will be replaced with `C0018802` in the new document. Both data sets (Ped and Cav) contained UMLS concept pairs (which may actually represent term phrases rather than single terms); converting the corpora to concepts thus allows direct comparison of the single concept pairs contained in the two data sets.

When modelling paradigmatic associations it is common to consider only those terms close to the target term, i.e. within a window of text centred around the target term. However, here we used the whole document as the context window. In this way we aim to capture in MedTrack the associations that exists within the context of a single patient record and in OHSUMED the associations that exists within the context of a single medical abstract.

As the TE model in Equation (4) is parameterized over γ, i.e. the mix between paradigmatic and syntagmatic information, this parameter was tuned to maximise the correlation with human similarity judgements (gold standard/ground truth labels). To fairly tune γ and also provide insight into the robustness of the TE model, a split train/test methodology was used. This was done by training on one data set and corpus to find the best value of γ and then testing on another data set/corpus, ensuring a cross corpus and cross data set combination was done for each.

Table 2 summarises the results obtained following this methodology and also reports the per-

[1]The pair *Lymphoid hyperplasia* was removed from the original set of 30 as neither concept existed in the test collections shown in table 1.

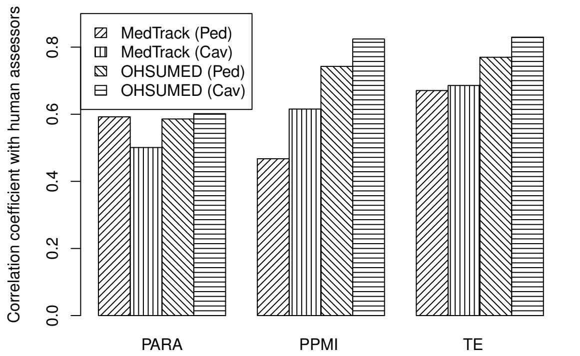

Figure 1: Correlation of medical concept judgements produced by PARA, PPMI and TE with those produced by human assessors.

formance of PPMI for comparison. Note that PPMI was shown to be the strongest performing measure when compared to thirteen other corpus based measures (Koopman et al., 2012).

5 Experimental Results

The effect of explicitly modelling both syntagmatic and paradigmatic information when estimating the similarity of medical concepts is shown in Figure 1. This graph shows that the TE model achieves a much higher correlation with human judged similarity scores (with an average correlation of 0.74 over all datasets and corpora) than both the paradigmatic (PARA: 0.57) and syntagmatic (PPMI: 0.66) approaches alone. To gain a broader understanding of how each of these measures compares to our TE variant, an updated graph showing the average performance of each across all data sets is provided in Figure 2. We refer the reader to Koopman et. al., (Koopman et al., 2012) for more details on the settings used for each of the other measures.

5.1 Sensitivity to the Mixing Parameter γ

The sensitivity to the mixing parameter γ of the TE approach is shown in Figure 3. This illustrates that for all datasets the best performance is achieved by some mix of both syntagmatic (PPMI) and paradigmatic (PARA) information.

The robustness of the PPMI and PAR measures across datasets and corpora can be inferred by comparing the distance between the end points of the TE lines drawn in Figure 3. The left hand side of the graph (where $\gamma = 0$) illustrates the performance of TE when only syntagmatic associations are considered, i.e. when TE uses only the PPMI

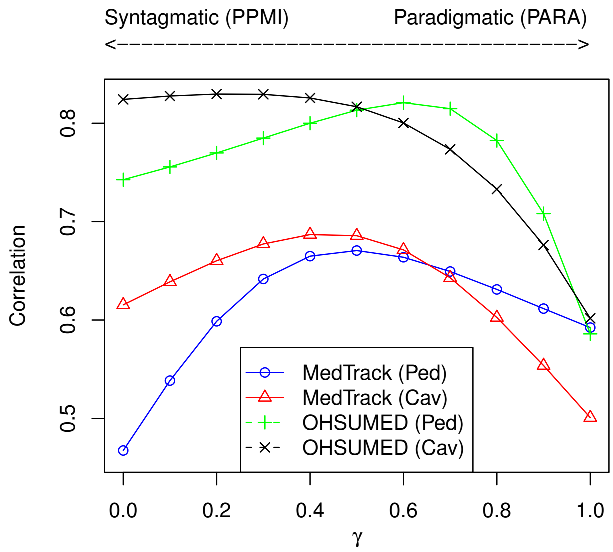

Figure 3: TE sensitivity to the mixing parameter γ. Results show that the TE model is robust across datasets and corpora.

measure (as Equation (4) reduces to only $S_{\text{ppmi}}()$ when $\gamma = 0$).

The right hand side of the graph ($\gamma = 1$) shows the performance of TE when considering only paradigmatic information, i.e. when only the PARA measure is used. With most lines converging to the same point on the right hand side, this demonstrates the increased robustness the PARA measure (and therefore paradigmatic associations) brings to the overall model.

5.2 Analysis of Paradigmatic and Syntagmatic Behaviour

To illustrate why the combination of both paradigmatic and syntagmatic measures can achieve such robust results across all datasets and corpora we compare the correlation of PPMI, PAR and TE against human assessments on a per concept-pair basis.

Figure 4 illustrates the normalised similarity scores (on log scale) of human assessors, PPMI, PARA and TE on the Caviedes and Cimino (Cav) dataset when using the OHSUMED corpora. The concept-pairs are placed in descending order of similarity as assessed by human judges, i.e. from the most similar human judged pairs to the least from left to right. The performance of a measure can be visualised by comparing its trend line with that of the descending human judged trend line. If the measure's trend line is parallel to that of the

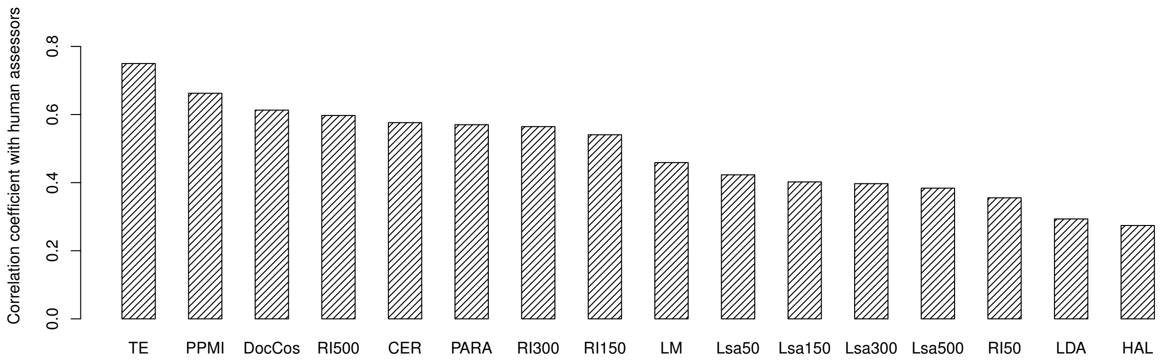

Figure 2: Comparison of all corpus-based measures (from Koopman et al., 2012), including TE and PARA; correlations averaged across datasets and corpora.

Pair #	Concept 1	Doc. Freq.	Concept 2	Doc. Freq.
11	Arrhythmia	2,298	Cardiomyopathy, Alcoholic	13
16	Angina Pectoris	1,725	Cardiomyopathy, Alcoholic	13
21	Abdominal pain	690	Respiratory System Abnormalities	1
34	Cardiomyopathy, Alcoholic	13	Respiratory System Abnormalities	1
36	Heart Diseases	1,872	Respiratory System Abnormalities	1
37	Heart Failure, Congestive	1,192	Respiratory System Abnormalities	1
38	Heartburn	104	Respiratory System Abnormalities	1

Table 3: Example concept pairs for which the PARA measure diverges from the human judgements on the OHSUMED corpus. Document frequencies showing the prevalence of the concepts in the corpus are reported. We conclude that the PARA measure is unable to estimate accurate semantic similarity when insufficient occurrence statistics are available for either concept.

human judges, then this indicates a strong correlation.

To better understand why the paradigmatic based measure differs from human assessors in Figure 4, the document frequencies of concept pairs 11, 16, 21, 34, 36, 37 and 38 from the Cav data set are reported in Table 3.

This table shows that for these concept pairs at least one concept occurs in a very small number of documents. This provides little evidence for the accurate estimation of paradigmatic associations between the concept pairs. We therefore conclude that the PARA measure requires that concepts occur in a sufficient number of documents for an effective semantic similarity estimation.

Similar observations are valid across datasets and corpora. For example, consider the correlation of PARA with human judgements for the Pedersen et al. (Ped) data set and the MedTrack corpus, as shown in Figure 5. The document frequencies for a number of concept pairs that show divergence from the Ped data set are shown in Table 4.

For these concept pairs where PARA is inconsistent with human judges, the PPMI measure effectively estimates semantic similarity. Thus the TE model, which mixes the two form of associations, is still effective even when the PARA measure is unreliable. This further supports the inclusion of both paradigmatic and syntagmatic associations for assessing semantic similarity between medical concepts.

Figure 4 also illustrates a large number of discontinuities in the PPMI graph. A discontinuity, i.e. the absence of the data-point within the plot, is due to a PPMI score of zero for the concept pair[2].

In practice, these discontinuities represent instances where the concept pair never co-occurs within any document. The same situation applies across other datasets and corpora, for example the

[2]As the graph is in log scale, $\log(0) = -\infty$ cannot be plotted.

Pair #	Concept 1	Doc. Freq.	Concept 2	Doc. Freq.
9	Diarrhea	6,184	Stomach cramps	14
23	Rectal polyp	26	Aorta	3,555

Table 4: Example concept pairs for which the PARA measure diverges from the human judgements on the MedTrack corpus. Document frequencies showing the prevalence of the concepts in the corpus are reported. We conclude that the PARA measure is unable to estimate accurate semantic similarity when insufficient occurrence statistics are available for either concept.

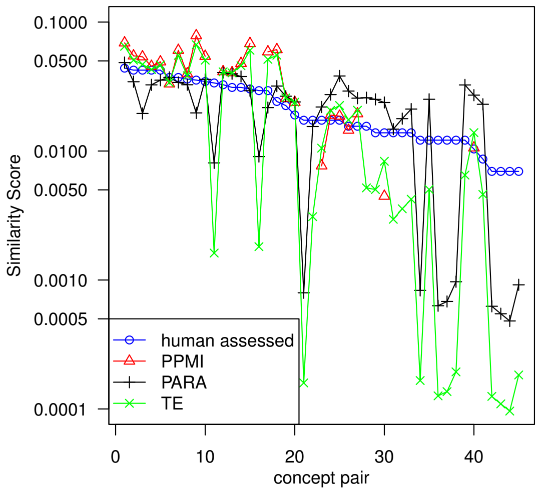

Figure 4: Normalised similarity scores (on log scale) of human assessors, PPMI, PARA and TE on Caviedes and Cimino dataset (Cav) when using the OHSUMED corpus for priming.

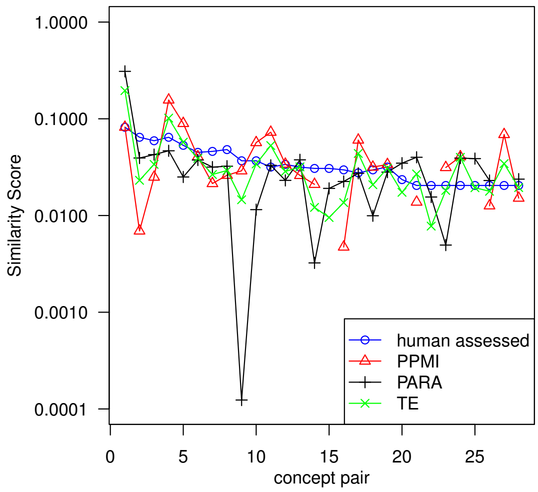

Figure 5: Normalised similarity scores (on log scale) of human assessors, PPMI, PARA and TE on Pedersen et al dataset (Ped) when using the Medtrack corpus for priming.

Ped data set on MedTrack corpus shown in Figure 5.

While PPMI discontinuities for concept pairs judged as unrelated by human assessors are correct estimates (as PPMI= 0 implies unrelatedness), discontinuities for concept pairs judged similar (e.g. pairs 11, 16, etc. in Figure 4) indicate a failure of the PPMI measure. These situations may provide the reason why the performance of PPMI, and indeed of many existing corpus-based approaches (Koopman et al., 2012), are sensitive to the choice of priming corpus. This may indicate that the ability to successfully model syntagmatic associations is sensitive to the corpus used for priming. However, because of the results obtained by the TE model we can conclude that appropriately mixing both syntagmatic and paradigmatic associations overcomes corpus sensitivity issues.

In summary, the performance (both in terms of robustness and effectiveness) of the TE model is achieved by including both syntagmatic and paradigmatic associations between medical concepts; this is due to the diversification of the type of information used to underpin the semantic similarity estimation process.

6 Conclusion

This paper has presented a novel variant of a robust and effective corpus-based approach that estimates similarity between medical concepts that strongly correlates with human judges. This approach is based on the *tensor encoding* (TE) model. By explicitly modelling syntagmatic and paradigmatic associations the TE model is able to outperform state of the art corpus-based approaches. Furthermore, the TE model is robust across corpora and datasets, in particular overcoming corpus sensitivity issues experienced by previous approaches.

A significant contribution of this paper is to highlight the important role of paradigmatic associations. Our results suggest that paradigmatic information provides an alternative source of evidence from which semantic similarity judgements can be drawn. It is this diversity of both syntagmatic and paradigmatic information that allows the TE model to be robust and effective.

A possible area of future work is the development of an adaptive TE approach. An adaptive approach would determine the best mix of syntagmatic and paradigmatic information on a case-by-case basis, using corpus statistic features. Our analysis in fact has shown that paradigmatic associations require a minimum number of *occurrences of concepts* within documents. While, syntagmatic associations require a minimum number of *co-occurrences of concept pairs* within documents. These corpus statistics could represent features for a machine learning approach that predicts the optimal mix of syntagmatic and paradigmatic information for the TE model.

Finally, because of its effectiveness and robustness, the TE model has other potential applications beyond semantic similarity measures. One relevant application may include using the TE model within query expansion tasks in ad-hoc medical information retrieval (as this process already relies heavily on similarity judgements).

References

Pankaj Agarwal and David B Searls. 2009. Can literature analysis identify innovation drivers in drug discovery? *Nature reviews. Drug discovery*, 8(11):865–78, November.

Alan R Aronson and François-Michel Lang. 2010. An overview of MetaMap: historical perspective and recent advances. *JAMIA*, 17(3):229–236.

Jorge E Caviedes and James J Cimino. 2004. Towards the development of a conceptual distance metric for the UMLS. *Journal of biomedical informatics*, 37(2):77–85, April.

Scott Cederberg and Dominic Widdows. 2003. Using LSA and noun coordination information to improve the precision and recall of automatic hyponymy extraction. In *Proc of CoNLL'03*, pages 111–118, NJ, USA.

Trevor Cohen and Dominic Widdows. 2009. Empirical distributional semantics: Methods and biomedical applications. *Journal of Biomedical Informatics*, 42(2):390–405.

P Glenisson, P Antal, J Mathys, Y Moreau, and B De Moor. 2003. Evaluation Of The Vector Space Representation In Text-Based Gene Clustering. In *Proc Pacific Symposium of Biocomputing*, pages 391–402.

Bevan Koopman, Guido Zuccon, Peter Bruza, Laurianne Sitbon, and Michael Lawley. 2012. An Evaluation of Corpus-driven Measures of Medical Concept Similarity for Information Retrieval. In *The 21st ACM International Conference on Information Knowledge Management 2012*, pages 2439–2442.

Thomas K. Landauer and Susan T. Dumais. 1997. Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review*, 104(2):211–240.

Kevin Lund and Curt Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments and computers*, 28:203–208.

Ted Pedersen, Sarguei Pakhomov, Siddharth Patwardhan, and Christopher Chute. 2007. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40(3):288–299.

Michael Symonds, Peter Bruza, Laurianne Sitbon, and Ian Turner. 2011a. Modelling Word Meaning using Efficient Tensor Representations. In *Pacific Asia Conference on Language, Information and Computation 2011*, pages 313–322.

Michael Symonds, Peter Bruza, Laurianne Sitbon, and Ian Turner. 2011b. Tensor Query Expansion: a cognitive based relevance model. In *Australasian Document Computing Symposium 2011*, pages 87–94.

Michael Symonds, Peter Bruza, Laurianne Sitbon, and Ian Turner. 2012. A Tensor Encoding Model for Semantic Processing. In *The 21st ACM International Conference on Information Knowledge Management 2012*, pages 2267–2270.

Ellen Voorhees and Richard Tong. 2011. Overview of the TREC Medical Records Track. In *Proc of TREC'11*, MD, USA.