Limits of Lexical Semantic Relatedness with Ontology-based Conceptual Vectors

Lim Lian Tze[1] and Didier Schwab[2]

[1] Computer-aided Translation Unit, School of Computer Sciences
Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

[2] GETALP-LIG Groupe d'Étude en Traduction/Traitement des Langues et de la Parole
Laboratoire d'Informatique de Grenoble, Université Pierre Mendès, France

Abstract. Conceptual vectors can be used to represent thematic aspects of text segments, which allow for the computation of semantic relatedness. We study the behavior of conceptual vectors based on an ontology by comparing the results to the Miller-Charles benchmark. We discuss the limits to such an approach due to explicit mapping, as well as the viability of the Miller-Charles dataset as a benchmark for assessing lexical semantic relatedness.

1 Introduction

Natural Language Processing (NLP) applications are often intended to be used via human-computer interactions. Example scenarios include receiving a useful proposal based on a goal and the available information (message planning); or to find the best word in a target language (interactive machine translation). To provide helpful responses to the user, it is important for such systems to be able to analyse or generate utterances adequately, relative to a user's intuitions about word meanings, or in particular the relatedness between the word meanings.

The Miller-Charles benchmark dataset [1] was compiled so that machine-computed semantic similarity measures of word pairs may be compared to human judgements. We use conceptual vectors to represent thematic aspects for text segments, with appropriate definitions of distances to computate semantic relatedness. We study the behavior of conceptual vectors based on an ontology by comparing the results to the Miller-Charles benchmark, and examine the limits to such an approach. We also discuss the viability of the Miller-Charles dataset as a benchmark for assessing lexical semantic relatedness.

2 Conceptual Vectors

2.1 Principle and Thematic Distance

Vectors have long been in use in NLP. The standard vector model (SVM) was first proposed for information retrieval [2], while latent semantic analysis (LSA) was developed for meaning representation [3]. They are inspired by distributional semantics [4] which

Lian Tze L. and Schwab D. (2008).
Limits of Lexical Semantic Relatedness with Ontology-based Conceptual Vectors.
In *Proceedings of the 5th International Workshop on Natural Language Processing and Cognitive Science*, pages 153-158
DOI: 10.5220/0001737301530158

hypothesises that a word meaning can be defined by its context. For example, the meaning of ‹*milk*› could be described by {‹*cow*›, ‹*cat*›, ‹*white*›,‹*cheese*›, ‹*mammal*›, ...}. Hence, distributional vector elements correspond directly (SVM) or indirectly (LSA) to lexical items from utterances.

The conceptual vector (CV) model is different as it is inspired by componential linguistics [5] which holds that the meaning of words can be described with semantic components. These can be considered as atoms of meaning [6], or as constituents of the meaning [7]. For example, the meaning of ‹*milk*› could be described by {*LIQUID*, *DAIRY PRODUCT*, *WHITE*, *FOOD*,...}. CVs model a formalism for the projection of this notion in a vectorial space. Hence, CV elements correspond to concepts indirectly, as we will see later. CVs can be associated to all levels of a text (word, phrase, sentence, paragraph, text, ...). As they represent ideas, they correspond to the notion of *semantic field*, the set of ideas conveyed by a term, at the lexical level; and also overall thematic aspects at the level of the entire text. CVs can also be applied to lexical meanings. They have been studied in word sense disambiguation using isotopic properties in a text, i.e. redundancy of ideas [7]. The basic idea is to maximise the overlap of shared ideas between word senses. This can be done by computing the angular distance between two CVs. For two CVs X and Y, the Sim function ($= \cos(\widehat{X,Y}) = \frac{X \cdot Y}{\|X\| \times \|Y\|}$) constitutes the *thematic proximity*, and $D_A = \arccos(Sim(A,B))$ measures the angle between the two vectors from a geometric point of view.

2.2 Operations on Vectors

Weak Contextualisation. When two terms are in presence of each other, some of the ideas in each term are accentuated by those in the other term. The contextualisation operation $\gamma(X,Y) = X \oplus (X \odot Y)$ emphasizes the features shared by both terms, where $\oplus$ is the normalised vectorial sum, averaging the two operand vectors; and $\odot$ is the vectorial term-to-term product, highlighting their common ideas [8].

Partial Synonymy. The synonymy function $Syn_R(X,Y,C) = D_A\left(\gamma(X,C), \gamma(Y,C)\right)$ tests the thematic closeness of two meanings (X and Y), each enhanced with what it has in common with a third (C) [9]. The Partial Synonymy function, $Syn_P(X,Y) = Syn_R(X,Y,\gamma(X, X \oplus Y) \oplus \gamma(Y, X \oplus Y))$, is simply Syn_R where the context is the sum of contextualisation of X and Y of their means (normalised sum) [10].

2.3 Properties and Construction

The construction of CVs assumes that ideas should be considered relative to each other. It seems more relevant to compare the proportion of the different ideas conveyed by terms or meanings. Following this idea, all conceptual vectors are normalised, i.e. they have the same magnitude. Geometrically speaking, objects represented by a conceptual vector are projected onto a hypersphere[3].

CVs can be constructed based on definitions from different sources, including dictionaries, synonym lists, manually crafted indices, etc. Definitions are parsed and the

[3] A hypersphere is constituted with all points at the same distance d from an origin point in any dimensional space.

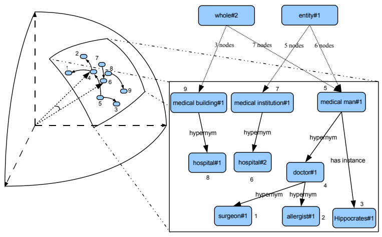

Fig. 1. Projecting WordNet concepts onto a hyperspace.

corresponding CV is computed. This approach fabricates, from existing CVs and definitions, new CVs. It requires a bootstrap with a kernel composed of pre-computed CVs.

2.4 Ontology-based Conceptual Vectors and Randomised Vectors by Emergence

An approach to induce CVs by emergence from randomised vectors was described in [11, 8], as opposed to ontology- or taxonomic hierarchy-based vectors in [9]. The emergence approach is attractive as the dimension of the vector space can be chosen freely. Also, the lexical density in a vector space computed by emergence is more constant than in a space with predefined concepts. However, as the iterative cycles of this approach requires more time and computing resources, we decided to construct the vectors by bootstrapping from an ontology.

3 Constructing Conceptual Vectors for WordNet Senses

WordNet (WN) [12] is a lexical semantic network comprising rich, explicit relations between English word senses. We construct a CV for each word sense in WN so that they can be explored via the usual WN relations *and also* the neighbourhood that they occupy in the hyperspace thus created. The right half of Fig. 1 shows several healthcare-related senses in WN's IS-A hierarchy. Visiting other related senses from one sense in the IS-A hierarchy often involves a "long" journey. For example, the path from ‹*doctor#n#1*› to ‹*hospital#n#2*› is 14 nodes in length. However, once CVs (all of unit length) are constructed for these senses, we see (on the left of Fig. 1) that they are projected onto a similar region in a hyperspace. In other words, these healthcare-related senses are now semantically closer to each other in the hyperspace.

3.1 Vector Construction Process

Niles and Pease [13] annotated all WN senses with class labels from SUMO, an upper ontology, and its related ontologies (including MILO) to aid the automatic processing of free texts. For example, ‹*doctor#n#1*› is subsumed by the class MedicalDoctor. We describe the construction of CVs for WN senses from these sources below.

Vector Construction for Ontology Classes. We used 1992 classes from SUMO and MILO for bootstrapping purposes: in other words, each CV will have 1992 vector components. For each ontology class $\mathcal{C}$, we initialise its vector $V^0(\mathcal{C})$ to be a boolean vector, such that each component $v_i^0 = 1$ if v_i corresponds to $\mathcal{C}$, $v_i^0 = 0$ otherwise. The final vector for each ontology class $\mathcal{C}$, $V(\mathcal{C})$, is then computed with each component $v_i(\mathcal{C}) = v_i^0(\mathcal{C}) + \sum_{j=1}^{dim(V)} \frac{v_j^0(\mathcal{C})}{2^{dist(\mathcal{C},\mathcal{C}_j)}}$. $dim(V)$ is the dimension of V (= 1992), $\mathcal{C}_j$ is the j-th ontological class corresponding to vector component v_j, and $dist(\mathcal{M}, \mathcal{N})$ is the path length connecting classes $\mathcal{M}$ and $\mathcal{N}$ in SUMO/MILO.

Kernel Vectors for WordNet Synsets. We set the kernel vector $V^0(s)$ for each WordNet synset s to be $V(\mathcal{C})$, where $\mathcal{C}$ is the SUMO/MILO class assigned to s in the data made available by [13]. For example, V^0(‹*doctor#n#1*›) $= V^0$(MEDICALDOCTOR).

Learning CVs for WN Senses. The CV for each WN sense $V(s)$ is computed iteratively following [8, sect. 4], but where they used a randomly-generated vector as the "seed", we use the kernel vectors V^0 instead. This process was deployed on a grid cluster environment to maximise the use of computing resources.

4 Proximity Measures and Comparison with Miller–Charles Set

Miller and Charles [1] asked 38 native English speakers to rate the similarity between a chosen set of word pairs, on a scale of 0 to 4. The resultant dataset (hereafter M&C) has been used as the benchmark in many semantic similarity measure studies.

To evaluate the CVs, we define two thematic proximity measures between lexical objects A and B: $\text{prox}_{cv}(A, B) = 1 - \left(D_A(A, B) \div \frac{\pi}{2}\right)$; $\text{prox}_{syn}(A, B) = 1 - \left(Syn_P(A, B) \div \frac{\pi}{2}\right)$. For each word pair in the M&C dataset, we take the highest prox_{cv} and prox_{syn} values of all possible combinations of their noun senses in WN. The results are shown in Table 1, comparing prox_{cv} and prox_{syn} values with the human judgement scores from M&C. The rows are sorted by decreasing "accumulated correlation coefficient" between our proximity values and the M&C ratings (re-scaled to $[0, 1]$).

5 Discussion

The correlation coefficients of prox_{cv} and prox_{syn} to the M&C set are 0.644 and 0.634 respectively. While this does not seem impressive, a closer scrutiny reveals that they are due to a few word pairs at the end of Table 1. We discuss some possible causes below.

Suitability of WN–SUMO Mappings for CVs. Our proximity values for ‹*brother*›-‹*monk*› was too low due to the very different WN–SUMO mapping: they are mapped to the HUMAN and RELIGIOUSORGANISATION classes respectively. We also noticed that many adjectives (and their morphologically-related nouns) are mapped to SUBJECTIVE-ASSESSMENTATTRIBUTE, thus "overcrowding" that class, possibly skewing the proximity value for many other word pairs.

This may indicate that such explicit mapping data between a general domain lexical resource (WN) and an ontology, particularly an upper ontology (SUMO), is not suitable for the purpose of constructing CVs: such mapping efforts often have their own strict guidelines and philosophies to adhere to, and do not always reflect topical relatedness. On the other hand, while the word pairs ‹*coast*›-‹*hill*›, ‹*forest*›-‹*graveyard*› and ‹*coast*›-‹*forest*› are deemed by human judges as very dissimilar, their prox_{cv} and prox_{syn} values

Table 1. Comparing similarity scores for the Miller–Charles dataset.

Word Pair		M&C	prox$_{cv}$	prox$_{syn}$	Corr. with M&C	
					prox$_{cv}$	prox$_{syn}$
automobile	car	0.98	1.00	1.00	1.000	1.000
cord	smile	0.03	0.48	0.57	1.000	1.000
glass	magician	0.03	0.47	0.57	1.000	1.000
gem	jewel	0.96	1.00	1.00	1.000	1.000
rooster	voyage	0.02	0.44	0.53	0.999	0.998
magician	wizard	0.88	0.90	0.92	0.997	0.997
bird	crane	0.74	0.82	0.86	0.996	0.996
crane	implement	0.42	0.60	0.68	0.989	0.991
noon	string	0.02	0.38	0.47	0.988	0.988
bird	cock	0.76	0.78	0.83	0.983	0.985
coast	shore	0.93	0.86	0.89	0.980	0.982
journey	voyage	0.96	0.85	0.88	0.974	0.976
midday	noon	0.86	1.00	1.00	0.965	0.968
furnace	stove	0.78	0.70	0.77	0.950	0.956
implement	tool	0.74	0.67	0.74	0.936	0.944
brother	lad	0.42	0.47	0.55	0.925	0.930
food	rooster	0.22	0.34	0.43	0.920	0.921
lad	wizard	0.11	0.20	0.30	0.915	0.911
asylum	madhouse	0.90	0.70	0.76	0.901	0.897
food	fruit	0.77	0.60	0.69	0.885	0.885
boy	lad	0.94	0.62	0.70	0.856	0.860
monk	slave	0.14	0.62	0.69	0.837	0.839
car	journey	0.29	0.71	0.77	0.818	0.818
monk	oracle	0.28	0.71	0.77	0.800	0.798
coast	hill	0.22	0.71	0.77	0.777	0.774
forest	graveyard	0.21	0.81	0.85	0.735	0.731
coast	forest	0.11	0.70	0.77	0.711	0.704
brother	monk	0.71	0.34	0.43	0.644	0.634
Corr. with M&C			**0.644**	**0.634**		

are high as the LANDAREA class is very prominent in the conceptual vectors of these lexical items. This is, again, due to the WN–SUMO mapping data.

Suitability of the M&C Set. Resnik [14] commented on the difference between the notions of semantic *relatedness* and *similarity*: ‹*cars*› and ‹*gasoline*› are more closely related than ‹*cars*› and ‹*bicycles*›, but the latter pair is more similar. This is the reason for the low human ratings of ‹*car*›-‹*journey*› but high prox$_{cv}$ and prox$_{syn}$ values: while the meanings are very dissimilar, they are highly related.

This suggests that our thematic proximity measures do indeed indicate lexical semantic relatedness, as opposed to the M&C experiments which are concerned with lexical semantic similarity: Budanitsky and Hirst [15] commented that semantic similarity is a special case of semantic relatedness, and that human judges in the M&C experiment were instructed to assess similarity instead of generic relatedness. We are also of the opinion that the M&C word pairs are more suitable for assessing similarity, but are less helpful for assessing relatedness. Therefore, we propose that a future experiment be conducted to collect human judgements of lexical semantic relatedness, with a more suitable set of word pairs.

6 Conclusions and Future Work

We have shown how CVs can model the ideas conveyed by lexical meanings, how they can be constructed based on ontological sources, and how they can be used to measure lexical semantic relatedness. Although encouraging, our results confirm the weaknesses of a hierarchy-based vector construction approach as identified in [10, 11], such as the non-standard density of the hierarchy, and different philosophies in mapping lexical senses to ontology classes. We therefore plan to explore the effect of using hierarchy-free CVs, i.e. construction by emergence. Realising that the M&C study is more concerned with lexical similarity, we will also collect human judgement ratings pertaining to lexical semantic relatedness as a more suitable benchmark.

Acknowledgements

This work was funded by a ScienceFund grant (No. 01-01-05-SF0011) from the Malaysian Ministry of Science, Technology and Innovation, and a Research University Grant from Universiti Sains Malaysia. We thank Dr Chan Huah Yong, Michael Cheng and Aloysius Indrayanto for use of their grid computing facilities, and the four anonymous reviewers for their comments.

References

1. Miller, G.A., Charles, W.G.: Contextual correlates of semantic similarity. **6** (1991)
2. Salton, G.: The smart document retrieval project. In: Proc. of the 14th Annual Int'l ACM/SIGIR Conf. on Research and Development in Information Retrieval, Chicago (1991)
3. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. Journal of the American Society of Information Science **41** (1990)
4. Harris, Z.S., Gottfried, M., Ryckman, T., Mattick Jr., P., Daladier, A., Harris, T., Harris, S.: The form of Information in Science, Analysis of Immunology Sublanguage. Volume 104 of Boston Studies in the Philosophy of Science. Kluwer Academic Publisher, Dordrecht (1989)
5. Hjelmlev, L.: Prolégolème à une théorie du langage. éditions de minuit (1968)
6. Wierzbicka, A.: Semantics: Primes and Universals. Oxford University Press (1996)
7. Greimas, A.J.: Structural Semantics: An Attempt at a Method. University of Nebraska Press (1984)
8. Schwab, D., Lim, L.T., Lafourcade, M.: Conceptual vectors, a complementary tool to lexical networks. In: Proc. of the 4th Int'l Workshop on Natural Language Processing and Cognitive Science (NLPCS 2007), Funchal, Madeira, Portugal (2007)
9. Lafourcade, M., Prince, V.: Synonymy and conceptual vectors. In: NLPRS'2001, Tokyo, Japon (2001)
10. Schwab, D.: Approche hybride - lexicale et thématique - pour la modélisation, la détection et l'exploitation des fonctions lexicales en vue de l'analyse sémantique de texte. PhD thesis, Université Montpellier 2 (2005)
11. Lafourcade, M.: Conceptual vector learning - comparing bootstrapping from a thesaurus or induction by emergence. In: LREC'2006, Genoa, Italia (2006)
12. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: An on-line lexical database. International Journal of Lexicography (special issue) **3** (1990)
13. Niles, I., Pease, A.: Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In: Proc. of the IEEE Int'l Conf. on Information and Knowledge Engineering. (2003)
14. Resnik, P.: Using information content to evaluate semantic similarity. In: Proc. of the 15th Int'l Joint Conf. on Artificial Intelligence, Montreal, Canada (1995)
15. Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of lexical semantic relatedness. Computational Linguistics **32** (2006)