Can We Model Conceptual Combination using Distributional Information?

Dermot Lynott (dermot.lynott@ucd.ie)
Department of Computer Science, UCD, Belfield, Dublin 4.

Michael Ramscar (michael@cogsci.ed.ac.uk)
Institute for Communicating and Collaborative Systems, Division of Informatics, 2 Buccleuch Place, Edinburgh EH8 9LW.

Abstract

Recent years have seen the development of several competing theories of conceptual combination (including CARIN, Gagné, 2000; the constraint theory, Costello and Keane, 2000; Dual-Process Theory, Wisniewski, 1997), and computational implementations of these theories, most notable are the CARIN model (Gagné, 2000, in press), the C^3 model (Costello & Keane, 2000) and the Smith et al's Selective Modification model (1988). While most models of conceptual combination employ hand-coded, rule-based implementations we present an approach which utilizes a more distributed mechanism, latent semantic analysis (LSA), in order to model data from the conceptual combination literature. We concluded that LSA, without explicit encoding of features, relations etc. could be used to model a wide range of tasks including the broad classification of novel noun-noun combinations, the degree of relatedness of features to novel compounds and the complexity of novel compounds. However, while these results provide evidence for the use of co-occurrence techniques in modelling cognitive processes it is clear that LSA does not provide a complete story for the processes underlying phenomena such as conceptual combination.

Introduction

Recent research into conceptual combination has produced several competing theories (including CARIN, Gagné, 2000; constraint theory, Costello and Keane, 2000; Dual-Process Theory, Wisniewski, 1997), and models, most notable are the CARIN model (Gagné, 2000, in press), C^3 model (Costello & Keane, 2000) and Smith et al's Selective Modification model (1988). Implementing these theories as testable computational models requires the hand-coding of a large amount of knowledge, and making numerous assumptions concerning conceptual structure that are at best speculative, and often theoretically or empirically dubious (see Komatsu, 1992; Ramscar & Hahn, in submission).

Recent research suggests an alternative strategy for looking at aspects of conceptual processing. Numerous studies have indicated *distributional information* can play a powerful role in many aspects of human cognition. In particular, it has been proposed that people can exploit statistical regularities in language to accomplish a range of conceptual and perceptual learning tasks. Saffran, Aslin & Newport (1996) have demonstrated that infants and adults are sensitive to simple conditional probability statistics, suggesting one way in which the ability to segment the speech stream into words may be realized (see also Saffran, Newport & Aslin, 1996). Redington, Chater & Finch (1998) suggest that distributional information may contribute to the acquisition of syntactic knowledge by children.

Here, we examine whether a different metric of conceptual knowledge, based upon distributional measures, can provide an objective and parsimonious competitor to existing models of conceptual combination effects. The work presented here explores whether a distributional measure (in this case LSA, Landauer & Dumais, 1997) can be used to simulate existing data and predict human performance in conceptual combinations tasks.

What is LSA?

LSA (Latent Semantic Analysis) is both a theory and a model of representing word meaning. It uses the context in which the words appear to provide a measure of similarity between words, paragraphs and documents. It is a statistical technique for analysing (and predicting) the contextual usage of words, based on what words they have co-occurred with in a corpus. The distribution of a word is represented as a vector in high-dimensional space. We can gauge the similarity between two words by calculating the distance between the their vectors[1]. In this way LSA provides us with an objective measure of semantic similarity.

In three experiments below we outline how LSA reflects human performance in tasks related to conceptual combination including reflecting the complexity of novel compounds, discriminating between classes of novel compound and judging the relatedness of features to novel compounds.

Experiment 1: Complexity of Novel Compounds

Murphy (1990) analysed compounds in terms of their complexity (or ease of comprehension). In three experiments Murphy distinguished between combination types

[1] This is done by calculating the cosine of the angle between the two vectors. A cosine approaching 1.0 means the words are highly similar. Conversely, a cosine approaching 0.0 means the words are dissimilar.

of (i) typical adjective-noun e.g. *sharp saw,* atypical adjective-noun e.g. *mortal god,* and noun-noun e.g. *park dive*; (ii) typical adjective-noun, atypical adjective-noun and non-predicating adjective-noun e.g. *criminal lawyer* and (iii) typical (*pointy* and *rocket*) and atypical adjectives (*dull* and *rocket*) in relevant and irrelevant contexts. Each of the three experiments involved a rating and a lexical decision task. We examined whether LSA similarity scores between a compounds constituents reflected ease of comprehension of the compounds.

Materials Materials were taken from Murphy (1990) Experiment 1. These comprised 26 head nouns, 26 typical adjectives, 26 atypical adjectives, 26 noun modifiers.

Procedure Semantic similarity scores were calculated in LSA[2] i.e. between each modifier (typical adjectives, atypical adjectives and nouns) and their head nouns.

Results and Discussion Using single factor ANOVA a difference between the groups was found at $p<0.001$ $F(2, 75)=16.31$. The difference between atypical adjectives and noun modifiers was significant at $p<0.001$ ($df=25$, $F=4.39$). The difference between typical adjectives and noun modifiers was significant at $p<0.001$ ($df=25$, $F=7.33$). The difference between typical and atypical adjectives was only significant at $p=0.1$ ($df=25$, $F=1.67$). The mean score for typical, atypical and noun modifiers were 0.273, 0.198 and 0.073 respectively. These averages reflect the results of Murphy's rating interpretation task that rated typical adjective-noun compounds easiest to interpret (6.61 average score on a 1-7 scale). Next easiest were atypical adjective-nouns (5.44) and finally, noun-noun compounds were rated most difficult to interpret (3.33). The LSA ranking matches those of Murphy’s response time experiment (1322 ms, 1574 ms and 1857 ms for typical, atypical and noun modifier compounds) and percentage of meaningful response results (99%, 88% and 65% in the same order).

Although this experiment indicates that LSA can separate combinations on the basis of complexity, it says nothing of the type of interpretation being arrived at. Below, we examine whether LSA can provide information on the type of interpretation subjects produce for novel compounds.

[2] Using General Reading up to 1st Year College semantic space, with term-to-term comparison, maximum factors.

Experiment 2: Distinguishing Compounds on the Basis of Similarity

Dual-process theory (Wisniewski, 1997) specifies two broad classes of combination: relation-based and property-based/hybrid. For example, possible interpretations of the compound *robin snake* might be "a snake that eats robins" (relation-based) or "a snake with a red underbelly" (property-based). Hybridisation occurs when the compound represents something which is both the head and the modifier e.g. *musician painter* (Wisniewski, 1996). The theory predicts that if the constituents of a compound are similar it is more likely that the interpretation produced will be property-based rather than relation-based. We examine whether LSA can distinguish between these two categories using data from Wisniewski & Love (1998) in 2a and Gagné (2000) in 2b.

Experiment 2a

Materials 30 noun-noun compounds were selected from Wisniewski & Love (1998) Experiment 1; 20 compounds were novel. 10 of these had constituents that were similar (e.g. car truck) and 10 had constituents that were dissimilar (e.g. mourner musician). The final 10 compounds were known/familiar combinations (e.g. jazz musician). Wisniewski & Love (1998) found that when constituents of a compound were similar, there was a much higher probability of subjects arriving at a property-based interpretation compared to when the constituents were dissimilar (65.7% compared to 16.4%). Conversely, when the constituents were dissimilar the number of relational interpretations produced was reliably higher (62.2% versus 22.3%).

Method The combinations were compared in LSA using the same parameters as per Experiment 1. LSA scores were then analysed and compared with the results from Wisniewski & Love.

Results Using single-factor ANOVA, there was a significant difference between the groups (novel similar, novel dissimilar and familiar) at $p<0.02$, $F(2, 27)=4.66$. Compounds with dissimilar constituents were differentiated from the two other categories at $p<0.001$, $F(1, 19)=0.1$. No difference was found between compounds with similar constituents and the other compounds ($p=0.4$, $F(1, 19)=0.82$). Similar constituent compounds were, however, differentiated from dissimilar constituent compounds at $p<0.003$, $F(1, 9)=8.3$ and dissimilar were differentiated from familiar at $p<0.001$, F(1, 9)=0.075. Tukey's HSD (post-hoc test) confirms these results. The mean LSA scores were 0.248, 0.08 and 0.286 for similar, dissimilar and familiar respectively.

Experiment 2b

Materials 64 noun-noun compounds were selected from Gagné's (2000) experiment 1. Gagné's four-group classification was used; property-based with similar constituents, property-based with dissimilar constituents, relation-based with similar constituents and relation-based with dissimilar constituents.

Method Each compound's constituents were compared in LSA using term-to-term comparison.

Results LSA scores reveal a significant difference between property-based and relation-based compounds at $p<0.001$, $F(1, 31)=9.74$. Tukey's HSD confirms this result ($p<0.001$). Average scores for the groups were 0.444 (property-based with similar constituents), 0.089 (relation-based with similar constituents), 0.105 (property-based with dissimilar constituents) and 0.085 (relation-based with dissimilar constituents).

Discussion

LSA clearly distinguishes between relation and property-based combinations. Interestingly, even where combinations have been pre-classified as having similar/dissimilar constituents, LSA still separates property-based from relation-based combinations. The question is whether people do actually use features/relations to distinguish between combinations or whether the feature/relation distinction arises from the distributional information encoded in concepts. We return to this in the General Discussion.

Experiment 3: Separating Related and Unrelated Features

In experiments 1 and 2 we have seen that LSA can distinguish between types of combination and reflect the ease of comprehension of compounds. The fact that metric distributional can provide this information is interesting but we need to examine whether distributional information can tell us what features are called to mind when a subject is presented with a novel combination. Connell & Ramscar (2001) have shown that LSA can be used to model typicality in category membership. Here we examine whether LSA can do the same for combined categories. This experiment differes from the previous experiments since it was run predictively i.e. using LSA to predict subject resposnes. Firstly, we will test whether LSA can distinguish between related and unrelated features of novel concepts and secondly, whether LSA can judge the relative order of importance of these features (as rated by human judges) to novel combinations.

Materials 18 noun-noun combinations were selected from Gagné (2000 and *in press*) plus three noun-noun combinations used as fillers. Between 7 and 9 related features (e.g. *job language* and *interview, jargon, boss*) were selected for each combination and one unrelated feature (e.g. *coal country* and *cinema.* All features were chosen by the authors based on intuition.

LSA Method Each combination was compared to a list of related and unrelated features using document to term comparison in LSA. The LSA scores were then scaled[3], ranked and then compared to the human relatedness ratings.

Procedure The experiment was a voluntary web-based experiment and all 16 candidates were native English speakers. Subjects were given relatedness-rating instructions (similar to typicality rating instructions as in Rosch & Mervis, 1975). They were asked to rate how related a feature was to a particular noun-noun combination (e.g. *urban complaint* and *noise*) using a scale of one to seven, with one being Very Related and

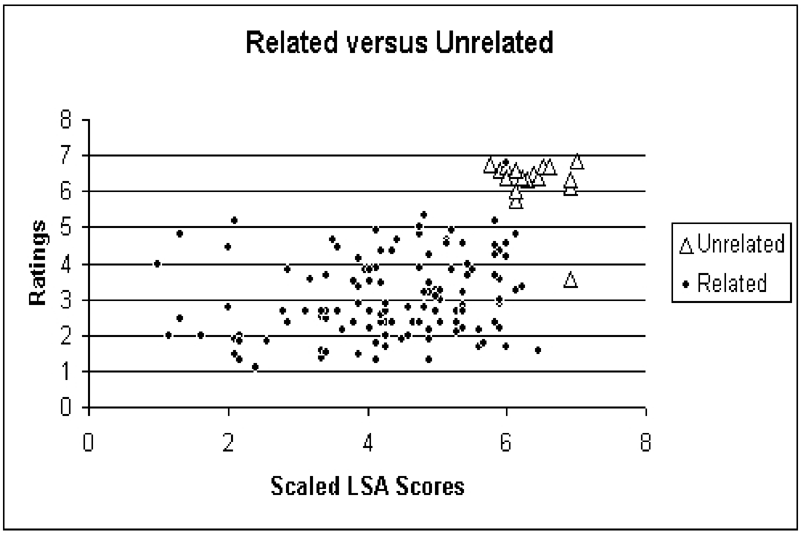

Figure 1. Separation of related and unrelated features of novel compounds.

[3] Scaled in line with the seven-point scale the subjects would use. The scaling formula is as follows: X is the LSA score to be scaled and M is the maximum LSA score for this category set: Scaled LSA score=$M - (M - 1) / (M * X)$.

seven being Not Related at all (i.e. unrelated). Subjects were free to interpret what was meant by the term "Related" in this context. Combinations would appear in block capitals on the left-hand side of the screen, with features appearing, on the same level, on the right-hand side. Subjects had to enter a number between one and seven and press return to move to the next item. Items were presented in a random order.

Results In 95.6% of cases the subject judgments correlated with LSA predictions on what features were unrelated to each combination (see Figure 1.). Two-tailed t-tests were

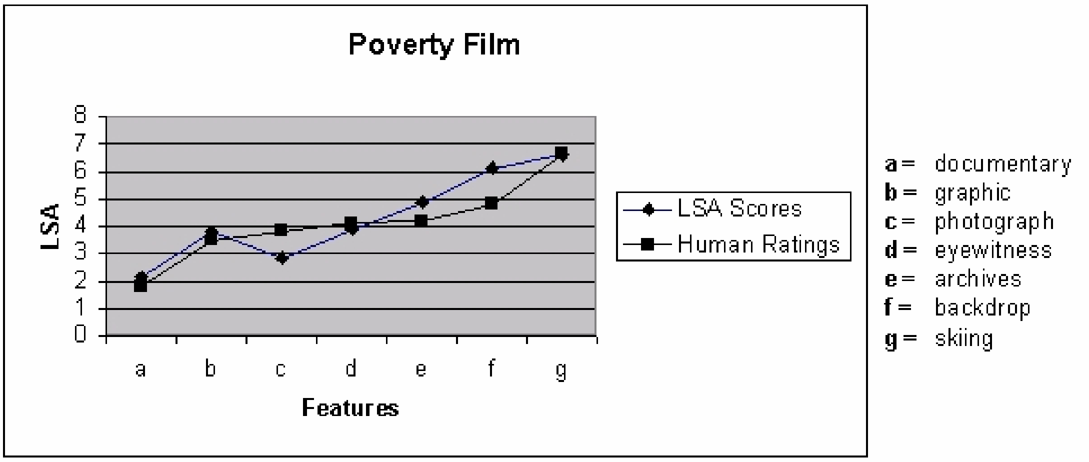

Figure 2: Illustrating the correlation between LSA scores and human ratings for the relatedness of features to novel compounds.

performed on each axis and were significant at p<0.001, t=6.92, df=141, for the x-axis (LSA scores) and p<0.001, t=12.01, df=141, for the y-axis (subject ratings). The overall correlation was significant at p <0.001, N=142, r=0.46667. One combination category had a significant correlation at p <0.01, r=0.964 (*poverty film* see Figure 2.). Three combination categories had a significant correlation at p <0.05, with r=0.833, 0.738 and 0.821 (*milk virus*, *marine gun* and *grain law* respectively). Two combinations had marginally significant correlations at p <0.1, with r=0.55 and 0.69 (*library antique* and *coal country* respectively). The other 13 combinations did not have significance at p <0.1.

Discussion

These results suggest that LSA can predict the separation of related and unrelated features in subject judgements. Despite the overall correlation being highly significant only a small number of the individual categories were themselves significant. Inconsistencies in the results may be explained in part by aspects of this particular implementation (e.g. small corpus, corpus contains many typographical errors) of a distributional model of language rather than indicating flaws in the underlying methodology. Despite this, the suggestion remains that distributional information can aid us in modelling tasks involving conceptual processing.

General Discussion

Without employing hand-tailored representations of conceptual structure such as features and relational links, LSA was able to simulate a variety of phenomena related to conceptual combination. We have shown that using a distributional metric (LSA) can reflect the relative complexity of novel combinations, distinguish between relation-based and property-based combinations, separate related and unrelated features of novel compounds and, to a lesser extent, rank the importance of features to novel compounds. The fact that a surface level statistical analysis of a corpus has managed to simulate such a wide range of phenomena related to conceptual processing raises interesting questions with regard to people's mental representations of concepts and the combining of these representations: Do people use distributional information to construct their representations of concepts and interpret combinations? Or are distributional properties of words (which models such as LSA extract) merely an epiphenomenon; a reflection of the fact that underlying concepts share certain semantic features? By the latter account, the distributional properties associated with words would arise *because* the concepts underlying the words possess certain features, and it is sensitivity to similarities between concepts that subjects actually manifest. However, work by MacDonald & Ramscar (in submission) seems to indicate the former. They have shown that manipulation of the distributional properties of the contexts in which nonce words are read can influence similarity judgements between existing words and nonce words. This result suggests that not *all* distributional responses can be explained in terms of existing conceptual structure: nonce words simply don't have existing conceptual structure. Equally, it seems highly unlikely that the structure of the linguistic environment is independent of the structure that people extract from their interactions with the world. What the results presented here (and other distributional research) seem to indicate is that any proper characterization of conceptual thought and conceptual combination will have to consider more than just the information that comes from physical experience and the physical environment. It is clear that in modelling concept use, one must also consider experience of language, and the structure of the linguistic environments in which speakers find themselves. While it may be the case that LSA cannot provide a complete account of conceptual combination, since it provides no explicit information on the relationships between heads and modifiers, it is clear that distributional information does have something to offer in the development of future cognitive models.

Acknowledgments

The authors would like to thank Louise Connell and Dan Yarlett for helpful comments on the work presented in this paper.

References

Connell, L. C. & Ramscar, M. J. A. (2001). Modelling Canonical and Contextual Typicality Using Distributional Measures. Proceedings of the 23rd Annual Conference of the Cognitive Science Society. Edinburgh, Scotland.

Costello, F. J. & Keane, M. T. (2000). Efficient creativity: Constraints on conceptual combination. Cognitive Science 24, 2, 299-349.

Gagné, C. L., (2000). Relation-based versus property based combinations: A test of the CARIN theory and dual-process theory of conceptual combination. Journal of Memory and Language, 42, 365-389.

Gagné, C. L., (in press).Relation-based and lexical-based priming. Journal of Experimental Psychology: Learning, Memory and Cognition.

Gerrig, R. J. & Murphy, G. L., (1992). Contextual influences on the comprehension of complex concepts. Language and Cognitive Processes, 7, 205-230.

Komatsu, L. K., (1992). Recent views of conceptual structure. Psychological Bulletin, 112-500-526.

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

MacDonald, S. & Ramscar, M. J. A. (2001) Testing the distributional hypothesis: the influence of context on judgements of semantic similarity. Proceedings of the 23rd Annual Conference of the Cognitive Science Society. Edinburgh, Scotland.

Murphy, G. L., (1990). Noun phrase interpretation and conceptual combination. Journal of Memory and Language, 29, 259-288.

Ramscar, M. J. A. & Hahn, U. (in submission). What family resemblances are not: Categorisation and the concept of ‘concept’. Manuscript in submission.

Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22,* 425-469.

Rosch, E. & Mervis, C. B., (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month old infants. Science, 274, 1926-1928.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996) Word segmentation: The role of distributional cues. Journal of Memory and Language, 35, 606-621.

Smith, E. E. & Osherson, D. N., Rips, L. J. & Keane, M., (1988). Combining prototypes: A selective modification model. Cognitive Science, 12, 485-527, 1988.

Wisniewski, E. J., (1996). Construal and Similarity in conceptual combination. Journal of Memory and Language, 35, 434-453.

Wisniewski, E. J., (1997). When concepts combine. Psychonomic Bulletin and Review, 4, 167-183.

Wisniewski, E. J. & Love, B. C., (1998). Relations versus properties in conceptual combination. Journal of Memory and Language, 38, 177-202.