Concreteness and Corpora: A Theoretical and Practical Analysis

Felix Hill
Computer Laboratory
University of Cambridge
fh295@cam.ac.uk

Douwe Kiela
Computer Laboratory
University of Cambridge
dlk427@cam.ac.uk

Anna Korhonen
Computer Laboratory
University of Cambridge
alk23@cam.ac.uk

Abstract

An increasing body of empirical evidence suggests that *concreteness* is a fundamental dimension of semantic representation. By implementing both a vector space model and a Latent Dirichlet Allocation (LDA) Model, we explore the extent to which concreteness is reflected in the distributional patterns in corpora. In one experiment, we show that that vector space models can be tailored to better model semantic domains of particular degrees of concreteness. In a second experiment, we show that the quality of the representations of abstract words in LDA models can be improved by supplementing the training data with information on the physical properties of concrete concepts. We conclude by discussing the implications for computational systems and also for how concrete and abstract concepts are represented in the mind

1 Introduction

A growing body of theoretical evidence emphasizes the importance of concreteness to semantic representations. This fact has not been widely exploited in NLP systems, despite its clear theoretical relevance to tasks such as word-sense induction and compositionality modeling. In this paper, we take a first step towards integrating concreteness into NLP by testing the extent to which it is reflected by the superficial (distributional) patterns in corpora. The motivation is both theoretical and practical: We consider the implications for the development of computational systems and also for how concrete and abstract concepts are represented in the human mind. Experimenting with two popular methods of extracting lexical representations from text, we show both that these approaches are sensitive to concreteness and that their performance can be improved by adapting their implementation to the concreteness of the domain of application. In

addition, our findings offer varying degrees of support to several recent proposals about conceptual representation.

In the following section we review recent theoretical and practical work. In Section 3 we explore the extent to which concreteness is reflected by Vector-Space Models of meaning (VSMs), and in Section 4 we conduct a similar analysis for (Bayesian) Latent Dirichlet Allocation (LDA) models. We conclude, in Section 5, by discussing practical and theoretical implications.

2 Related work

2.1 Concreteness

Empirical evidence indicates important cognitive differences between abstract concepts, such as *guilt* or *obesity,* and concrete concepts, such as *chocolate* or *cheeseburger*. It has been shown that concrete concepts are more easily learned and remembered than abstract concepts, and that language referring to concrete concepts is more easily processed (Schwanenflugel, 1991). There are cases of brain damage in which either abstract or concrete concepts appear to be specifically impaired (Warrington, 1975), and functional magnetic resonance imaging (fMRI) studies implicate overlapping but partly distinct neural systems in the processing of the two concept types (Binder et al., 2005). Further, there is increasing evidence that concrete concepts are represented via intrinsic properties whereas abstract representations encode extrinsic relations to other concepts (Hill et al., in press). However, while these studies together suggest that concreteness is fundamental to human conceptual representation, much remains to be understood about the precise cognitive basis of the abstract/concrete distinction. Indeed, the majority of theoretically motivated studies of conceptual representation focus on concrete domains, and

comparatively little has been established empirically about abstract concepts.

Despite this support for the cognitive importance of concreteness, its application to computational semantics has been limited to date. One possible reason for this is the difficulty in measuring lexical concreteness using corpora alone (Kwong, 2008). Turney et al. (2011) overcome this hurdle by applying a semi-supervised method to quantify noun concreteness. Using this data, they show that a disparity in the concreteness between elements of a construction can facilitate metaphor identification. For instance, in the expressions *kill the process* or *black comedy*, a verb or adjective that generally occurs with a concrete argument takes an abstract argument. Turney et al. show that a supervised classifier can exploit this effect to correctly identify 79% of adjective-noun and verb-object constructions as literal or metaphorical. Although these results are clearly promising, to our knowledge Turney et al.'s paper is unique in integrating corpus-based methods and concreteness in NLP systems.

1.2 Association / similarity

A proposed distinction between abstract and concrete concepts that is particularly important for the present work relates to the semantic relations *association* and (*semantic*) *similarity* (see e.g. Crutch et al. 2009; Resnik, 1995). The difference between these relations is exemplified by the concept pairs {*car, petrol*} and {*car, van*}. *Car* is said to be (semantically) similar to *van*, and associated with (but not similar to) *petrol*. Intuitively, the basis for the similarity of *car* and *bike* may be their common physical features (wheels) or the fact that they fall within a clearly definable category (modes of transport). In contrast, the basis for the association between *car* and *petrol* may be that they are often found together or the clear functional relationship between them. The two relations are neither mutually exclusive nor independent; *bike* and *car* are related to some degree by both association and similarity.

Based on fresults of behavioral experiments, Crutch et al. (2009) make the following proposal concerning how association and similarity interact with concreteness:

(C) *The conceptual organization of abstract concepts is governed by association, whereas the organization of concrete concepts is governed by similarity.*

Crutch et al.'s hypothesis derives from experiments in which participants selected the odd-one-out from lists of five words appearing on a screen. The lists comprised either concrete or abstract words (based on ratings of six informants) connected either by similarity (e.g. dog, wolf, fox etc.; theft, robbery, stealing etc.) or association (dog, bone, collar etc.; theft, law, victim etc.), with an unrelated odd-one-out item in each list. Controlling for frequency and position, subjects were both significantly faster and more accurate if the related words were either abstract and associated or concrete and similar. These results support (C) on the basis that decision times are faster when the related items form a more coherent group, rendering the odd-one out more salient. Hill et al. (in press) tested the same hypothesis on a larger scale, analyzing over 18,000 concept pairs scored by human annotators for concreteness as well as the strength of association between them. They found a moderate interaction between concreteness and the correlation between association strength and similarity (as measured using WordNet), but concluded that the strength of the effect was not sufficiently strong to either confirm or refute (C).

Against this backdrop, the present work examines how association, similarity and concreteness are reflected in LDA models and, first, VSMs. In both cases we test Hypothesis (C) and related theoretical proposals, and discuss whether these findings can lead to better performing semantic models.

3 Vector Space Models

Vector space models (VSMs) are perhaps the most common general method of extracting semantic representations from corpora (Sahlgren, 2006; Turney & Pantel, 2010). Words are represented in VSMs as points in a (geometric) vector space. The dimensions of the space correspond to the model features, which in the simplest case are high frequency words from the corpus. In such models, the position of a word representation along a given feature dimension depends on how often that word occurs within a specified proximity to tokens of the feature word in the corpus. The exact proximity required is an important parameter for model implementation, and is referred to as the *context window*. Finally, the degree to which two word representations are related can be calculated as some function of the distance between the corresponding points in the semantic space.

3.1 Motivation

VSMs are well established as a method of quantifying relations between word concepts and have achieved impressive performance in related NLP tasks (Sahlgren, 2006; Turney & Pantel, 2010). In these studies, however, it is not always clear exactly which semantic relation is best reflected by the implemented models. Indeed, research has shown that by changing certain parameter settings in the standard VSM architecture, models can be adapted to better reflect one relation type or another. Specifically, models with smaller context windows are reportedly better at reflecting similarity, whereas models with larger windows better reflect association. (Agirre et al., 2009; Peirsman et al., 2008)

Our experiments in this section aim first to corroborate these findings by testing how models of varying context window sizes perform on empirical data of both association and similarity. We then test if this effect differentially affects performance on concrete and abstract words.

3.2 Method

We employ a conventional VSM design, extracting representations from the (unlemmatised) British National Corpus (Leech et al., 1994) with stopwords removed. In the vector representation of each noun, our dimension features are the 50,000 most frequently occurring (nonstopword) words in the corpus. We experiment with window sizes of three, five and nine (one, two and four words either side of the noun, counting stopwords). Finally, we apply pointwise mutual information (PMI) weighting of our co-occurrence frequencies, and measure similarity between weighted noun vectors by the cosine of the angle between them in the vector space.

To evaluate modeling of association, we use the University of South Florida (USF) Free-association Norms (Nelson & McEvoy, 2012). The USF data consist of over 5,000 words paired with their *free associates*. To elicit free associates, more than 6,000 participants were presented with cue words and asked to "*write the first word that comes to mind that is meaningfully related or strongly associated to the presented word*". For a cue word *c* and an associate *a,* the *forward association strength* (*association*) from *c* to *a* is the proportion of participants who produced *a* when presented with *c*. *association* is thus a measure of the strength of an associate relative to other associates of that cue. The USF data is well suited to our purpose because many cues and associates in the data have a concrete-

ness score, taken from either the norms of Paivio, Yuille and Madigan (1968) or Toglia and Battig (1978). In both cases contributors were asked to rate words based on a scale of 1 (very abstract) to 7 (very concrete).[1] We extracted the all 2,230 nouns from the USF data for which concreteness scores were known, yielding a total of 15,195 noun-noun pairs together with concreteness and *association* values.

Although some empirical word-similarity datasets are publically available, they contain few if any abstract words (Finkelstein et al., 2002; Rubenstein & Goodenough, 1965). Therefore to evaluate similarity modeling, we use Wu-Palmer Similarity (*similarity*) (Wu & Palmer, 1994), a word similarity metric based on the position of the senses of two words in the WordNet taxonomy (Felbaum, 1998). *similarity* can be applied to both abstract and concrete nouns and achieves a high correlation, with human similarity judgments (Wu & Palmer, 1994).[2]

3.3 Results

In line with previous studies, we observed that VSMs with smaller window sizes were better able to predict *similarity*. The model with window size 3 achieves a higher correlation with similarity (Spearman rank r_s = -0.29) than the model with window size 9 (r_s = -0.25). However, the converse effect for association was not observed: Model correlation with *association* was approximately constant over all window sizes. These effects are illustrated in Fig. 1.

---
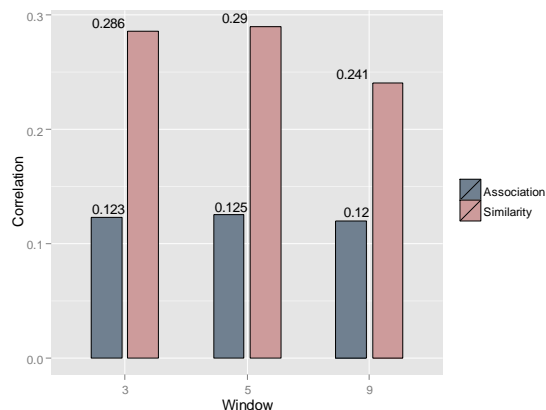
[1] Although concreteness is well understood intuitively, it lacks a universally accepted definition. It is often described in terms of reference to sensory experience (Paivio et al., 1968), but also connected to specificity; *rose* is often considered more concrete than *flora*. The present work does not address this ambiguity.

[2] similarity achieves a Pearson correlation of r = .80 on the 30 concrete word pairs in the Miller & Charles (1991) data.

Figure 1: Spearman correlations between VSM output and *association* and *similarity* for different window sizes.

In addressing the theoretical Hypothesis (C) we focused on the output of our VSM of window size five, although the same trends were observed over all three models. Over all 18,195 noun-noun pairs the correlation between the model output and *association* was significant (r_s = 0.13, $p < 0.001$) but notably lower than the correlation with *similarity* (r_s = -0.29, $p < 0.001$). To investigate the effect of concreteness, we ranked each pair in our sample by the total concreteness of both nouns, and restricted our analysis to the 1000 most concrete and 1000 most abstract pairs. The models captured *association* better over the abstract pairs than concrete concepts, but reflected *similarity* better over the concrete concepts. The strength of this effect is illustrated in Fig. 2.
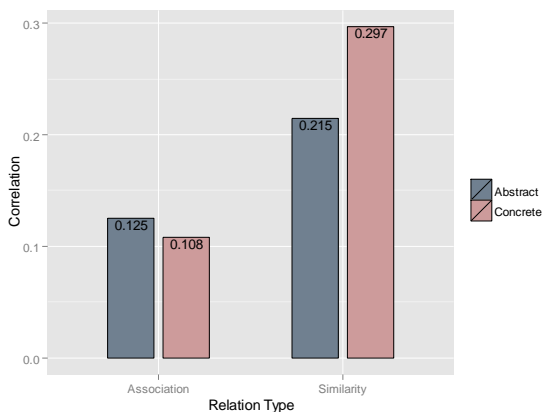
Figure 2: Spearman correlation values between VSM output and *similarity* and *association* over subsets of concrete and abstract pairs.

Given that small window sizes are optimal for modeling similarity, and that WSMs appear to model similarity better over concrete concepts than over abstract concepts, we explored whether different window sizes were optimal for either abstract or concrete word pairs. When comparing the model output to *association*, no interaction between window size and concreteness was observed. However, there was a notable interaction when considering performance in modeling *similarity*. As illustrated in Fig. 3, performance on concrete word pairs is better for smaller window sizes, whereas with abstract word pairs a larger window size is preferable.
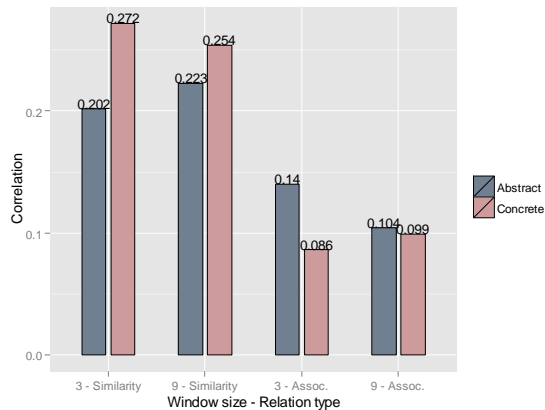
Figure 3: Spearman correlation values between VSM output and similarity and association for different window sizes over abstract and concrete word pair subsets

3.4 Conclusion

Our results corroborate the body of VSM research that reports better performance from small window sizes in modeling similarity. A likely explanation for this finding is that similarity is a paradigmatic relation: Two similar entities can be plausibly exchanged in most linguistic contexts. Small context windows emphasize proximity, which loosely reflects structural relationships such as verb-object, ensuring that paradigmatically related entities score highly. Models with larger context windows cannot discern paradigmatically and syntagmatically related entities in this way. The performance of our models on the association dataset did not support the converse conclusion that larger window sizes perform better. Overall, each of the three models was notably better at capturing similarity than association. This suggests that the core architecture of WSMs is not well suited to modeling association. Indeed, 'first order' models that directly measure word co-occurrences, rather than connecting them via features, seem to perform better at this task (Chaudhari et al., 2011). This fact is consistent with the view that association is a more basic or fundamental semantic relation from which other more structured relations are derived.

The fact that the USF association data reflects the instinctive first response of participants when presented with a cue word is important for interpreting the results with respect to Hypothesis (C). Our findings suggest that VSMs are better able to model this data for abstract word pairs than for concrete word pairs. This is consistent with the idea that language fundamentally determines which abstract concepts come to be associated or connected in the mind. Conversely, the

fact that the model reflects associations between concrete words less well suggests that the importance of extra-linguistic information is lower for connecting concrete concepts in this instinctive way. Indeed, it seems plausible that the process by which concrete concepts become associated involves visualization or some other form of perceptual reconstruction. Consistent with Hypothesis (C), this reconstruction, which is not possible for abstract concepts, would naturally reflect *similarity* to a greater extent than linguistic context alone.

Finally, when modeling similarity, the advantage of a small window increases as the words become more concrete. Similarity between concrete concepts is fundamental to cognitive theories involving the well studied notions of prototype and categorization (Rosch, 1975; Rogers & McClelland, 2003). In contrast, the computation of abstract similarity is intuitively a more complex cognitive operation. Although the accurate quantification of abstract similarity may be beyond existing corpus-based methods, our results suggest that a larger context window could in fact be marginally preferable should VSMs be applied to this task.

Overall, our findings show that the design of VSMs can be tailored to reflect particular semantic relations and that this in turn can affect their performance on different semantic domains, particularly with respect to concreteness. In the next section, we investigate whether the same conclusions should apply to a different class of distributional model.

4   Latent Dirichlet Allocation Models

LDA models are trained on corpora that are divided into sections (typically documents), exploiting the principle that words appearing in the same document are likely to have similar meanings. In an LDA model, the sections are viewed as having been generated by random sampling from unknown latent dimensions, which are represented as probability distributions (*Dirichlet* distributions) over words. Each document can then be represented by a probability distribution over these dimensions, and by considering the meaning of the dimensions, the meaning of the document can be effectively characterized. More importantly, because each latent dimension clusters words of a similar meaning, the output of such models can be exploited to provide high quality lexical representations (Griffiths et al., 2007). Such a word representation encodes the extent to which each of the latent dimensions influences the meaning of that word, and takes the form of a probability distribution over these dimensions. The degree to which two words are related can then be approximated by any function that measures the similarity or difference between distributions.

4.1   Motivation

In recent work, Andrews et al. (2009) explore ways in which LSA models can be modified to improve the quality of their lexical representations. They propose that concepts are acquired via two distinct information sources: *experiential data* – the perceptible properties of objects, and *distributional data* – the superficial patterns of language. To test this hypothesis, Andrews et al. construct three different LDA models, one trained on experiential data, one trained in the conventional manner on running text, and one trained on the same text but with the experiential data appended. They evaluate the quality of the lexical representations in the three models by calculating the Kulback-Leibler divergence between the representation distributions to measure how closely related two words are (Kullback & Leibler, 1951). When this data was compared with the USF association data, the combined model performed better than the corpus-based model, which in turn performed better than the features-only model. Andrews et al. concluded that both experiential and distributional data are necessary for the acquisition of good quality lexical representations.

As well as suggesting a way to improve the performance of LDA models on NLP tasks by supplementing the training data, the approach taken by Andrews et al. may be useful for better understanding the nature of the abstract/concrete distinction. In recent work, Hill et al. (in press) present empirical evidence that concrete concepts are represented in terms of intrinsic features or properties whereas abstract concepts are represented in terms of connections to other (concrete and abstract) concepts. For example, the features [legs], [tail], [fur], [barks] are all central aspects of the concrete representation of *dog*, whereas the representation of the abstract concept *love* encodes connections to other concepts such as *heart*, *rose*, *commitment* and *happiness* etc. If a *feature-based* representation is understood to be constructed from physical or perceptible properties (which themselves may be basic or fundamental concrete representations), Hill et al.'s characterization of concreteness can be summarized as follows:

(H) *Concreteness correlates with the degree to which conceptual representations are feature-based*

Because such differences in representation structure would in turn entail differences in the computation of similarity, (H) is closely related to a proposal of Markman and Stilwell (2001; see also Gentner & Markman, 2007):

(M) *Computing similarity among concrete concepts involves a feature-comparison operation, whereas similarity between abstract concepts is a structural, analogy-like, comparison.*

The findings of Andrews et al. do not address (H) or (M) directly, for two reasons. Firstly, they evaluate their model on a set that includes no abstract concepts. Secondly, they compare their model output to association data without testing how well it reflects similarity. In this section we therefore reconstruct the Andrews models and evaluate how well they reflect both association and similarity across a larger set of abstract and concrete concepts.

4.2   Method/materials

We reconstruct two of the three models developed by Andrews et al. (2009), excluding the features-only model because of the present focus on corpus-based approaches. However, while the experiential data applied in the Andrews et al. combined model was that collected by Vigliocco et al. (2004), we use the publicly available McRae feature production norms (McRae et al., 2005). The McRae data consist of 541 concrete noun concepts together with features for each elicited from 725 participants. In the data collection, *feature* was understood in a very loose sense, so that participants were asked to list both physical and functional properties of the nouns in addition to encyclopedic facts. However, for the present work, we filter out those features that were not perceptual properties using McRae et al.'s feature classes, leaving a total of 1,285 feature types, such as [has_claws] and [made_of_brass]. The importance of each feature to the representation of a given concept is reflected by the proportion of participants who named that feature in the elicitation experiment. For each noun concept we therefore extract a corresponding probability distribution over features.

The model design and inference are identical to those applied by Andrews et al. Our distributional model contains 250 latent dimensions and was trained using a Gibbs Sampling algorithm on approximately 7,500 sections of the BNC with stopwords removed.[3] The combined model contains 350 latent dimensions, and was trained on the same BNC data. However, for each instance of one of the 541 McRae concept words, a feature is drawn at random from the probability distribution corresponding to that word and appended to the training data. The latent dimensions in the combined model therefore correspond to probability distributions both over words and over features. This leads to an important difference between how words come to be related in the distributional model and in the combined model. Both models infer connections between words by virtue of their occurrence either in the same document or in pairs of documents for which the same latent dimensions are prominent. In the distributional model, it is the words in a document that determines which latent dimensions are ultimately prominent, whereas the in combined model it is both the words and the features in that document. Therefore, in the combined model, two words can come to be related because they occur not only in documents whose words are related, but also in documents whose features are related. For words in the McRae data, this has the effect of strengthening the relationship between words with common features. More interestingly, because it alters which latent dimensions are most prominent for each document, it should also influence the relationship between words not in the McRae data.

We evaluate the performance of our models in reflecting free association (*association*) and similarity (*similarity*). To obtain test items we rank the 18,195 noun-noun pairs from the USF data by the product of the two (BNC) word frequencies and select the 5,000 highest frequency pairs.

4.3   Results

As expected, the correlation of the combined model output with *association* was greater than the correlation of the distributional model output. Notably, however, as illustrated in Fig. 4, we observed far greater differences between the combined and the distributional models when comparing to *similarity*. Over all noun pairs, the addition of features in the combined model im-

---

[3] Code for model implementation was taken from Mark Andrews : http://www.mjandrews.net/code/index.html

proved the correlation with *similarity* from Spearman $r_s = 0.09$ to $r_s = 0.15$.
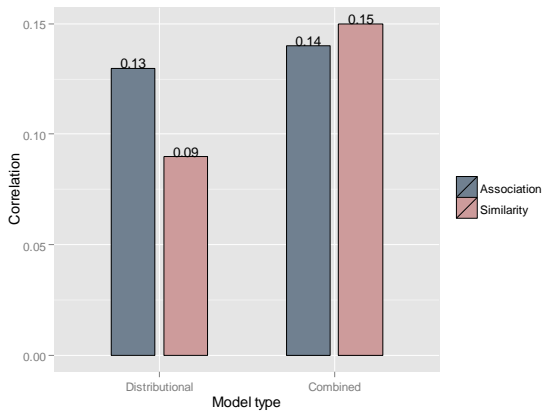

Figure 4: Spearman correlations between distributional and combined model outputs, *similarity* and *association*

In order to address Hypothesis (C) (Section 2.2), we analyzed the output of the combined model on subsets of the 1000 most abstract and concrete word pairs in our data as before. Perhaps surprisingly, as shown in Fig. 5, when comparing with *similarity*, the model performed better over abstract pairs, whereas when comparing with *association* the model performed better over concrete pairs. However, when these concrete pairs were restricted to those for which at least one of the two words was in the McRae data, and hence to which features had been appended in the corpus, the ability of the model to reflect *similarity* increased significantly.
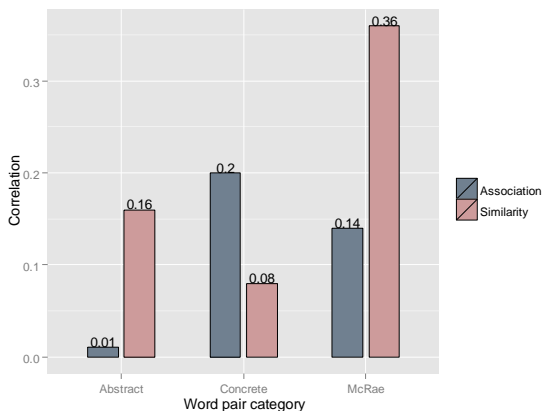

Figure 5: Spearman correlations between combined model output and *similarity* and *association* on different word pair subsets

Finally, to address hypotheses (H) and (M) we compared the previous analysis of the combined model output to the equivalent output from the distributional model. Surprisingly, as shown in Fig. 6, the ability of the model to reflect *association* over abstract pairs seemed to reduce with the

addition of features to the training data. Nevertheless, in all other cases the combined model outperformed the distributional model. Interestingly, the combined model advantage when comparing with *similarity* was roughly the same over both abstract and concrete pairs. However, when these pairs contained at least one word from the McRae data, the combined model was indeed significantly better at modeling *similarity*, consistent with Hypotheses (M) and (H).
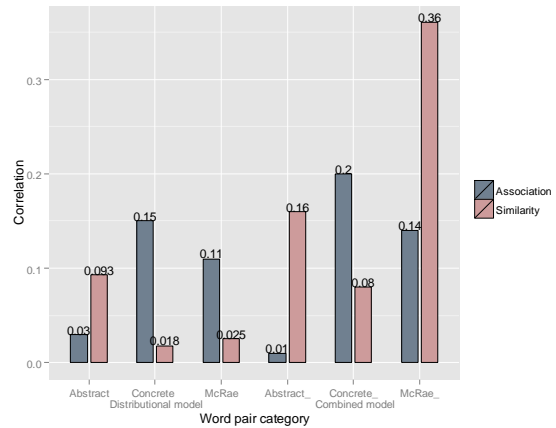

Figure 6: Comparison between distributional model and combined model output correlations with *similarity* and *association* over different word pair subsets

4.4   Conclusion

Our findings corroborate the main conclusion of Andrews et al., that the addition of experiential data improves the performance of the LDA model in reflecting association. However, they also indicate that the advantage of feature-based LDA models is far more significant when the objective is to model similarity.

The findings are also consistent with, if not suggestive of, the theoretical hypotheses (H) and (M). Clearly, the property features in the combined model training data enable it to better model both similarity and association between those concepts to which the features correspond. However, this benefit is greater when modeling similarity than when modeling association. This suggests that the similarity operation is indeed based on features to a greater extent than association. Moreover, this effect is far greater for the concrete words for which the features were added than over the other words pairs we tested. Whilst this is not a sound test of hypothesis (H) (no attempt was made to add 'features' of abstract concepts to the model), it is certainly consistent with the idea that features or properties are a more important aspect of concrete representations than of abstract representations.

Perhaps the most interesting aspect of the combined model is how the addition of feature information in the training data for certain words influences performance on words for which features were not added. In this case, our findings suggest that the benefit when modeling similarity is marginally greater than when modeling association, an observation consistent with Hypothesis (M). A less expected observation is that, between words for which features were not added, the advantage of the combined model over the distributional model in modeling similarity was equal if not greater for abstract than for concrete concepts. We hypothesize that this is because abstract representations naturally inherit any reliance on feature information from the concrete concepts with which they participate. In contrast, highly concrete representations do not encode relations to other concepts and therefore cannot inherit relevant feature information in the same way. Under this interpretation, the concrete information from the McRae words would propagate more naturally to abstract concepts than to other concrete concepts. As a result, the highest quality representations in the combined model would be those of the McRae words, followed by those of the abstract concepts to which they closely relate.

5   Discussion

This study has investigated how concreteness is reflected in the distributional patterns found in running text corpora. Our results add to the body of evidence that abstract and concrete concepts are represented differently in the mind. The fact that VSMs with small windows are particularly adept at modeling relations between concrete concepts supports the view that similarity governs the conceptual organization of concrete concepts to a greater extent than for abstract concepts. Further, the performance of our LSA models on different tasks and across different word pairs is consistent with the idea that concrete representations are built around features, whereas abstract concepts are not.

More practically, we have demonstrated that vector space models can be tailored to reflect either similarity or association by adjusting the size of the context window. This in turn indicates a way in which VSMs might be optimized to either abstract or concrete domains. Our experiments with Latent Dirichlet Allocation corroborate a recent proposal that appending training data with perceptible feature or property information for a subset of concrete nouns can signif-icantly improve the quality of the model's lexical representations. As expected, this effect was particularly salient for representations of words for which features were appended to the training data. However, the results show that this information can propagate to words for which features were not appended, in particular to abstract words.

The fact that certain perceptible aspects of meaning are not exhaustively reflected in linguistic data is a potentially critical obstacle for corpus-based semantic models. Our findings suggest that existing machine learning techniques may be able to overcome this by adding the required information for words that refer to concrete entities and allowing this information to propagate to other elements of language. In future work we aim to investigate specifically whether this hypothesis holds for particular parts of speech. For example, we would hypothesize that verbs inherit a good degree of their meaning from their prototypical nominal arguments.

References

Agirre, E., Alfonseca, E., Hall, K., Kravalova, J. Pasca, K,. & Soroa,A. 2009. A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches. *In Proceedings of NAACL-HLT 2009.*

Andrews, M., Vigliocco, G. & Vinson, D. 2009. Integrating experiential and distributional data to learn semantic represenations. *Psychological Review,* 116(3), 463-498.

Barsalou, L. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.

Binder, J., Westbury, C., McKiernan, K., Possing, E., & Medler, D. 2005. Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience* 17(6), 905-917.

Chaudhari, D., Damani, O., & Laxman, S. 2011. Lexical Co-occurrence, Statistical Significance, and Word Association. *EMNLP 2011,* 1058-1068.

Crutch, S., Connell, S., & Warrington, E. 2009. The different representational frameworks underpinning abstract and concrete knowledge: evidence from odd-one-out judgments. *Quarterly Journal of Experimental Psychology*, 62(7), 1377-1388.

Felbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Finkelstein, L., Gabrilovich, Matias, Rivlin, Solan, Wolfman & Ruppin. 2002. Placing Search in Context: The Concept Revisited. *ACM Transactions on Information Systems, 20*(1):116-131.

Gentner, D., & Markman, A. 1997. Structure mapping in analogy and similarity. *American Psychologist*, 52. 45-56.

Griffiths, T., Steyvers, M., & Tenembaum, J. 2007. Topics in semantic representation. *Psychological Review*, 114 (2), 211-244.

Hill, F., Korhonen, A., & Bentz, C. A quantitative empricial analysis of the abstract/concrete distinction. *Cognitive Science*. In press.

Kullback, S., & Leibler, R.A. 1951. On Information and Sufficiency. *Annals of Mathematical Statistics* 22 (1): 79–86.

Kwong, O, Y. 2008. A Preliminary study on inducing lexical concreteness from dictionary definitions. *22nd Pacific Asia Conference on Language, Information and Computation*, 235–244.

Leech, G., Garside, R. & Bryant, R. 1994. Claws4: The tagging of the British National Corpus. *COL-ING 94*, Lancaster: UK.

Markman, A, & Stilwell, C. 2001. Role-governed categories. *Journal of Theoretical and Experimental Artificial Intelligence,* 13, 329-358.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37, 547-559

Miller, G., & Charles, W. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1).

Nelson, D., & McEvoy, C. 2012. The University of South Florida Word Association, Rhyme and Word Fragment Norms. *Retrieved online from: http://web.usf.edu/FreeAssociation/Intro.html*.

Paivio, A., Yuille, J., & Madigan, S. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement,* 76(1, Pt. 2).

Peirsman, y., Heylen, K. & Geeraerts, D. 2008. Size Matters. Tight and Loose Context Definitions in English Word Space Models. *In Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics, Hamburg, Germany*

Resnik, P. 1995. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. *Proceedings of IJCAI-95*.

Rogers, T., & McLelland, J. 2003. *Semantic Cognition*. Cambridge, Mass: MIT Press.

Rosch, E. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 104(3), (September 1975), pp. 192–233.

Rubenstein, H., & Goodenough, J. 1965. Contextual correlates of synonymy. *Communications of the ACM* 8(10), 627-633.

Sahlgren, M. 2006. *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Ph.D. dissertation, Department of Linguistics, Stockholm University.

Schwanenflugel, P. 1991. *Why are abstract concepts hard to understand?* In P. Schwanenflugel. The psychology of word meanings (pp. 223-250). Hillsdale, NJ: Erlbaum.

Toglia, M., & Battig, W. 1978. *Handbook of semantic word norms*. Hillsdale, N.J: Erlbaum.

Turney, P, & Pantel, P. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research (JAIR)*, 37, 141-188.

Turney,P., Neuman, Y., Assaf,.D, Cohen, Y. 2011. Literal and Metaphorical Sense Identification through Concrete and Abstract Context. *EMNLP 2011*: 680-690

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. 2004. Reprssenting the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48, 422–488.

Warrington, E. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology* 27(4), 635-657.

Wu, Z., Palmer, M. 1994. Verb semantics and lexical selection. *In: Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics. 133–138.*