Grounding Semantics in Olfactory Perception

Douwe Kiela, Luana Bulat and Stephen Clark
Computer Laboratory
University of Cambridge
douwe.kiela,ltf24,stephen.clark@cl.cam.ac.uk

Abstract

Multi-modal semantics has relied on feature norms or raw image data for perceptual input. In this paper we examine grounding semantic representations in olfactory (smell) data, through the construction of a novel bag of chemical compounds model. We use standard evaluations for multi-modal semantics, including measuring conceptual similarity and cross-modal zero-shot learning. To our knowledge, this is the first work to evaluate semantic similarity on representations grounded in olfactory data.

1 Introduction

Distributional semantics represents the meanings of words as vectors in a "semantic space", relying on the *distributional hypothesis*: the idea that words that occur in similar contexts tend to have similar meanings (Turney and Pantel, 2010; Clark, 2015). Although these models have been successful, the fact that the meaning of a word is represented as a distribution over other words implies they suffer from the *grounding problem* (Harnad, 1990); i.e. they do not account for the fact that human semantic knowledge is grounded in physical reality and sensori-motor experience (Louwerse, 2008).

Multi-modal semantics attempts to address this issue and there has been a surge of recent work on perceptually grounded semantic models. These models learn semantic representations from both textual and perceptual input and outperform language-only models on a range of tasks, including modelling semantic similarity and relatedness, and predicting compositionality (Silberer and Lapata, 2012; Roller and Schulte im Walde, 2013; Bruni et al., 2014). Perceptual information is obtained from either feature norms (Silberer and Lapata, 2012; Roller and Schulte im Walde, 2013; Hill and Korhonen, 2014) or raw data sources such as images (Feng and Lapata, 2010; Leong and Mihalcea, 2011; Bruni et al., 2014; Kiela and Bottou, 2014). The former are elicited from human annotators and thus tend to be limited in scope and expensive to obtain. The latter approach has the advantage that images are widely available and easy to obtain, which, combined with the ready availability of computer vision methods, has led to raw visual information becoming the de-facto perceptual modality in multi-modal models.

However, if our objective is to ground semantic representations in perceptual information, why stop at image data? The meaning of *lavender* is probably more grounded in its smell than in the visual properties of the flower that produces it. Olfactory (smell) perception is of particular interest for grounded semantics because it is much more primitive compared to the other perceptual modalities (Carmichael et al., 1994; Krusemark et al., 2013). As a result, natural language speakers might take aspects of olfactory perception "for granted", which would imply that text is a relatively poor source of such perceptual information. A multi-modal approach would overcome this problem, and might prove useful in, for example, metaphor interpretation (*the sweet smell of success; rotten politics*) and cognitive modelling, as well as in real-world applications such as automatically retrieving smells or even producing smell descriptions. Here, we explore grounding semantic representations in olfactory perception.

We obtain olfactory representations by constructing a novel bag of chemical compounds (BoCC) model. Following previous work in multi-modal semantics, we evaluate on well known conceptual similarity and relatedness tasks and on zero-shot learning through induced cross-modal mappings. To our knowledge this is the first work to explore using olfactory perceptual data for grounding linguistic semantic models.

Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers), pages 231–236,
Beijing, China, July 26-31, 2015.

Olfactory-Relevant Examples					
MEN		sim	SimLex-999		sim
bakery	bread	0.96	steak	meat	0.75
grass	lawn	0.96	flower	violet	0.70
dog	terrier	0.90	tree	maple	0.55
bacon	meat	0.88	grass	moss	0.50
oak	wood	0.84	beach	sea	0.47
daisy	violet	0.76	cereal	wheat	0.38
daffodil	rose	0.74	bread	flour	0.33

Table 1: Examples of pairs in the evaluation datasets where olfactory information is relevant, together with the gold-standard similarity score.

2 Tasks

Following previous work in grounded semantics, we evaluate performance on two tasks: conceptual similarity and cross-modal zero-shot learning.

2.1 Conceptual similarity

We evaluate the performance of olfactory multi-modal representations on two well-known similarity datasets: SimLex-999 (Hill et al., 2014) and the MEN test collection (Bruni et al., 2014). These datasets consist of concept pairs together with a human-annotated similarity score. Model performance is evaluated using the Spearman ρ_s correlation between the ranking produced by the cosine of the model-derived vectors and that produced by the gold-standard similarity scores.

Evidence suggests that the inclusion of visual representations only improves performance for certain concepts, and that in some cases the introduction of visual information is detrimental to performance on similarity and relatedness tasks (Kiela et al., 2014). The same is likely to be true for other perceptual modalities: in the case of a comparison such as *lily-rose*, the olfactory modality certainly is meaningful, while this is probably not the case for *skateboard-swimsuit*. Some examples of relevant pairs can be found in Table 1.

Hence, we had two annotators rate the two datasets according to whether smell is relevant to the pairwise comparison. The annotation criterion was as follows: if both concepts in a pairwise comparison have a distinctive associated smell, then the comparison is relevant to the olfactory modality. Only if both annotators agree is the comparison deemed olfactory-relevant. This annotation leads to a total of four evaluation sets: the MEN test collection **MEN** (3000 pairs) and its olfactory-relevant subset **OMEN** (311 pairs); and the SimLex-999 dataset **SLex** (999 pairs) and its olfactory-relevant subset **OSLex** (65 pairs). The inter-annotator agreement on the olfactory relevance judgments was high ($\kappa = 0.94$ for the MEN test collection and $\kappa = 0.96$ for SimLex-999).[1]

2.2 Cross-modal zero-shot learning

Cross-modal semantics, instead of being concerned with improving semantic representations through grounding, focuses on the problem of reference. Using, for instance, mappings between visual and textual space, the objective is to learn which words refer to which objects (Lazaridou et al., 2014). This problem is very much related to the object recognition task in computer vision, but instead of using just visual data and labels, these cross-modal models also utilize textual information (Socher et al., 2014; Frome et al., 2013). This approach allows for *zero-shot learning*, where the model can predict how an object relates to other concepts just from seeing an image of the object, but without ever having seen the object previously (Lazaridou et al., 2014).

We evaluate cross-modal zero-shot learning performance through the average percentage correct at N (P@N), which measures how many of the test instances were ranked within the top N highest ranked nearest neighbors. A chance baseline is obtained by randomly ranking a concept's nearest neighbors. We use partial least squares regression (PLSR) to induce cross-modal mappings from the linguistic to the olfactory space and vice versa.[2]

Due to the nature of the olfactory data source (see Section 3), it is not possible to build olfactory representations for all concepts in the test sets. However, cross-modal mappings yield an additional benefit: since linguistic representations have full coverage over the datasets, we can project from linguistic space to perceptual space to also obtain full coverage for the perceptual modalities. This technique has been used to increase coverage for feature norms (Fagarasan et al., 2015). Consequently, we are in a position to compare perceptual spaces directly to each other, and to linguistic

[1] To facilitate further work in multi-modal semantics beyond vision, our code and data have been made publicly available at http://www.cl.cam.ac.uk/~dk427/aroma.html.

[2] To avoid introducing another parameter, we set the number of latent variables in the cross-modal PLSR map to a third of the number of dimensions of the perceptual representation.

Smell label	Phenethyl acetate	Isoamyl butyrate	Anisyl butyrate	Myrcene	Syringaldehyde
	Chemical Compound				
Melon	✓	✓			
Pineapple	✓				✓
Licorice			✓		
Anise			✓	✓	
Beer				✓	✓

Table 2: A BoCC model.

space, over the entire dataset, as well as on the relevant olfactory subsets. When projecting into such a space and reporting results, the model is prefixed with an arrow (→) in the corresponding table.

3 Olfactory Perception

The Sigma-Aldrich Fine Chemicals flavors and fragrances catalog[3] (henceforth SAFC) is one of the largest publicly accessible databases of semantic odor profiles that is used extensively in fragrance research (Zarzo and Stanton, 2006). It contains organoleptic labels and the chemical compounds—or more accurately the perfume raw materials (PRMs)—that produce them. By automatically scraping the catalog we obtained a total of 137 organoleptic smell labels from SAFC, with a total of 11,152 associated PRMs. We also experimented with Flavornet[4] and the LRI and odour database[5], but found that the data from these were more noisy and generally of lower quality.

For each of the smell labels in SAFC we count the co-occurrences of associated chemical compounds, yielding a bag of chemical compounds (BoCC) model. Table 2 shows an example subspace of this model. Although the SAFC catalog is considered sufficiently comprehensive for fragrance research (Zarzo and Stanton, 2006), the fact that PRMs usually occur only once per smell label means that the representations are rather sparse. Hence, we apply dimensionality reduction to the original representation to get denser

[3]http://www.sigmaaldrich.com/industries/flavors-and-fragrances.html

[4]http://www.flavornet.org

[5]http://www.odour.org.uk

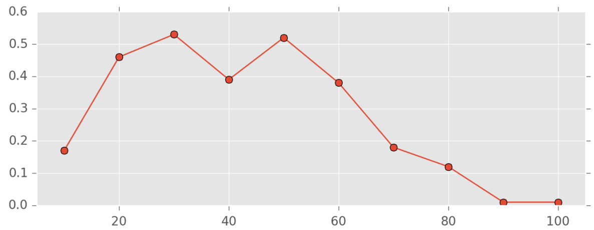

Figure 1: Performance of olfactory representations when using SVD to reduce the number of dimensions.

Dataset	Linguistic	BoCC-Raw	BoCC-SVD
OMEN (35)	0.40	0.42	0.53

Table 3: Comparison of olfactory representations on the covered OMEN dataset.

vectors. We call the model without any dimensionality reduction BoCC-Raw and use singular value decomposition (SVD) to create an additional BoCC-SVD model with reduced dimensionality. Positive pointwise mutual information (PPMI) weighting is applied to the raw space before performing dimensionality reduction.

The number of dimensions in human olfactory space is a hotly debated topic in the olfactory chemical sciences (Buck and Axel, 1991; Zarzo and Stanton, 2006). Recent studies involving multi-dimensional scaling on the SAFC catalog revealed approximately 32 dimensions in olfactory perception space (Mamlouk et al., 2003; Mamlouk and Martinetz, 2004). We examine this finding by evaluating the Spearman ρ_s correlation on the pairs of **OMEN** that occur in the SAFC database (35 pairs). The coverage on SimLex was not sufficient to also try that dataset (only 5 pairs). Figure 1 shows the results. It turns out that the best olfactory representations are obtained with 30 dimensions. In other words, our findings appear to corroborate recent evidence suggesting that olfactory space (at least when using SAFC as a data source) is best modeled using around 30 dimensions.

3.1 Linguistic representations

For the linguistic representations we use the continuous vector representations from the log-linear skip-gram model of Mikolov et al. (2013), specifically the 300-dimensional vector representations trained on part of the Google News dataset (about 100 billion words) that have been released on the

	MEN	OMEN	SLex	OSLex
Linguistic	0.78	0.38	0.44	0.30
→BoCC-Raw	0.38	0.36	0.19	0.23
→BoCC-SVD	0.46	0.51	0.23	0.48
Multi-modal	0.69	0.53	0.40	0.49

Table 4: Comparison of linguistic, olfactory and multi-modal representations.

Mapping	P@1	P@5	P@20	P@50
Chance	0.0	3.76	13.53	36.09
Olfactory ⇒ Ling.	1.51	8.33	24.24	47.73
Ling. ⇒ Olfactory	4.55	15.15	43.18	67.42

Table 5: Zero-shot learning performance for BoCC-SVD.

Word2vec website.[6]

3.2 Conceptual Similarity

Results on the 35 covered pairs of **OMEN** for the two BoCC models are reported in Table 3. Olfactory representations outperform linguistic representations on this subset. In fact, linguistic representations perform poorly compared to their performance on the whole of **MEN**. The SVD model performs best, improving on the linguistic and raw models with a 33% and 26% relative increase in performance, respectively.

We use a cross-modal PLSR map, trained on all available organoleptic labels in SAFC, to extend coverage and allow for a direct comparison between linguistic representations and cross-modally projected olfactory representations on the entire datasets and relevant subsets. The results are shown in Table 4. As might be expected, linguistic performs better than olfactory on the full datasets. On the olfactory-relevant subsets, however, the projected BoCC-SVD model outperforms linguistic for both datasets. Performance increases even further when the two representations are combined into a multi-modal representation by concatenating the L2-normalized linguistic and olfactory (→BoCC-SVD) vectors.

3.3 Zero-shot learning

We learn a cross-modal mapping between the two spaces and evaluate zero-shot learning. We use all 137 labels in the SAFC database that have corresponding linguistic vectors for the training data.

[6]https://code.google.com/p/word2vec/

apple	bacon	brandy	cashew
pear	smoky	rum	hazelnut
banana	roasted	whiskey	peanut
melon	coffee	wine-like	almond
apricot	mesquite	grape	hawthorne
pineapple	mossy	fleshy	jam
chocolate	**lemon**	**cheese**	**caramel**
cocoa	citrus	grassy	nutty
sweet	geranium	butter	roasted
coffee	grapefruit	oily	maple
licorice	tart	creamy	butterscotch
roasted	floral	coconut	coffee

Table 6: Example nearest neighbors for BoCC-SVD representations.

For each term, we train the map on all other labels and measure whether the correct instance is ranked within the top N neighbors. We use the BoCC-SVD model for the olfactory space, since it performed best on the conceptual similarity task. Table 5 shows the results. It appears that mapping linguistic to olfactory is easier than mapping olfactory to linguistic, which may be explained by the different number of dimensions in the two spaces. One could say that it is easier to find the chemical composition of a "smelly" word from its linguistic representation, than it is to linguistically represent or describe a chemical composition.

3.4 Qualitative analysis

We also examined the BoCC representations qualitatively. As Table 6 shows, the nearest neighbors are remarkably semantically coherent. The nearest neighbors for *bacon* and *cheese*, for example, accurately sum up how one might describe those smells. The model also groups together nuts and fruits, and expresses well what *chocolate* and *caramel* smell (or taste) like.

4 Conclusions

We have studied grounding semantic representations in raw olfactory perceptual information. We used a bag of chemical compounds model to obtain olfactory representations and evaluated on conceptual similarity and cross-modal zero-shot learning, with good results. It is possible that the olfactory modality is well-suited to other forms of evaluation, but in this initial work we chose to follow standard practice in multi-modal semantics to allow for a direct comparison.

This work opens up interesting possibilities in analyzing smell and even taste. It could be applied in a variety of settings beyond semantic similarity, from chemical information retrieval to metaphor interpretation to cognitive modelling. A speculative blue-sky application based on this, and other multi-modal models, would be an NLG application describing a wine based on its chemical composition, and perhaps other information such as its color and country of origin.

Acknowledgements

DK is supported by EPSRC grant EP/I037512/1. LB is supported by an EPSRC Doctoral Training Grant. SC is supported by ERC Starting Grant DisCoTex (306920) and EPSRC grant EP/I037512/1. We thank the anonymous reviewers for their helpful comments and Flaviu Bulat for providing useful feedback.

References

Elia Bruni, Nam-Khanh Tran, and Marco Baroni. 2014. Multimodal distributional semantics. *Journal of Artifical Intelligence Research*, 49:1–47.

Linda Buck and Richard Axel. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, 65(1):175–187.

S. Thomas Carmichael, M.-C. Clugnet, and Joseph L. Price. 1994. Central olfactory connections in the macaque monkey. *Journal of Comparative Neurology*, 346(3):403–434.

Stephen Clark. 2015. Vector Space Models of Lexical Meaning. In Shalom Lappin and Chris Fox, editors, *Handbook of Contemporary Semantics*, chapter 16. Wiley-Blackwell, Oxford.

Luana Fagarasan, Eva Maria Vecchi, and Stephen Clark. 2015. From distributional semantics to feature norms: grounding semantic models in human perceptual data. In *Proceedings of the 11th International Conference on Computational Semantics (IWCS 2015)*, pages 52–57, London, UK.

Yansong Feng and Mirella Lapata. 2010. Visual information in semantic representation. In *Proceedings of NAACL*, pages 91–99.

Andrea Frome, Gregory S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. 2013. DeViSE: A Deep Visual-Semantic Embedding Model. In *Proceedings of NIPS*, pages 2121–2129.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D*, 42:335–346.

Felix Hill and Anna Korhonen. 2014. Learning abstract concept embeddings from multi-modal data: Since you probably can't see what I mean. In *Proceedings of EMNLP*, pages 255–265.

Felix Hill, Roi Reichart, and Anna Korhonen. 2014. SimLex-999: Evaluating semantic models with (genuine) similarity estimation. *CoRR*, abs/1408.3456.

Douwe Kiela and Léon Bottou. 2014. Learning image embeddings using convolutional neural networks for improved multi-modal semantics. In *Proceedings of EMNLP*, pages 36–45.

Douwe Kiela, Felix Hill, Anna Korhonen, and Stephen Clark. 2014. Improving multi-modal representations using image dispersion: Why less is sometimes more. In *Proceedings of ACL*, pages 835–841.

Elizabeth A Krusemark, Lucas R Novak, Darren R Gitelman, and Wen Li. 2013. When the sense of smell meets emotion: anxiety-state-dependent olfactory processing and neural circuitry adaptation. *The Journal of Neuroscience*, 33(39):15324–15332.

Angeliki Lazaridou, Elia Bruni, and Marco Baroni. 2014. Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world. In *Proceedings of ACL*, pages 1403–1414.

Chee Wee Leong and Rada Mihalcea. 2011. Going beyond text: A hybrid image-text approach for measuring word relatedness. In *Proceedings of IJCNLP*, pages 1403–1407.

Max M. Louwerse. 2008. Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, 59(1):617–645.

Amir Madany Mamlouk and Thomas Martinetz. 2004. On the dimensions of the olfactory perception space. *Neurocomputing*, 58:1019–1025.

Amir Madany Mamlouk, Christine Chee-Ruiter, Ulrich G Hofmann, and James M Bower. 2003. Quantifying olfactory perception: Mapping olfactory perception space by using multidimensional scaling and self-organizing maps. *Neurocomputing*, 52:591–597.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Proceedings of ICLR*, Scottsdale, Arizona, USA.

Stephen Roller and Sabine Schulte im Walde. 2013. A multimodal LDA model integrating textual, cognitive and visual modalities. In *Proceedings of EMNLP*, pages 1146–1157.

Carina Silberer and Mirella Lapata. 2012. Grounded models of semantic representation. In *Proceedings of EMNLP*, pages 1423–1433.

Richard Socher, Andrej Karpathy, Quoc V. Le, Christopher D. Manning, and Andrew Y. Ng. 2014. Grounded compositional semantics for finding and describing images with sentences. *Transactions of ACL*, 2:207–218.

Peter D. Turney and Patrick Pantel. 2010. From Frequency to Meaning: vector space models of semantics. *Journal of Artifical Intelligence Research*, 37(1):141–188, January.

Manuel Zarzo and David T. Stanton. 2006. Identification of latent variables in a semantic odor profile database using principal component analysis. *Chemical Senses*, 31(8):713–724.