Harvesting Common-sense Navigational Knowledge for Robotics from Uncurated Text Corpora

Nancy Fulda, Nathan Tibbetts, Zachary Brown and David Wingate

Abstract: As robotic systems are deployed into everyday situations, the need for abstract reasoning becomes more pronounced. The ideal robotic assistant should be able to understand verbal commands and work independently to fulfill human-prescribed goals, even if instructions are ambiguous or circumstances change. This paper presents a new algorithm for high-level reasoning based on Euclidean representations of words and their meanings. Rather than using ontologies or knowledge graphs, we model information about the world as a learned geometry of the contexts in which human beings tend to use each idea. Building on the analogy algorithms utilized by Mikolov et al. [1], we perform mathematical operations on the vector space to infer responses to previously unseen problems, and apply our method to a sequence of semantic reasoning tasks in order to answer questions such as 'Where can I find a dustpan?', 'Where do the crayons belong?', and 'What transportation method will bring me to the airport?'. Our Directional Scoring Method (DSM) returns a ranked list of possible responses, many of which are plausible answers to the query. Additionally, DSM's top-ranked response is significantly more likely to be correct than the top-ranked responses of naive analogy estimations.

1 Introduction

The field of robotics is transitioning from automation to autonomy. Rather than performing specialized tasks repeatedly, modern systems are expected to behave intelligently in situations where human input is not immediately available and where environmental circumstances cannot be predicted in advance. To excel under these conditions, a robotic system must be equipped with general-purpose knowledge about its environment, the components and prerequisites of potential objectives, and the behavior of other entities. Such knowledge is often represented in the form of ontologies and knowledge graphs [2, 3, 4], but although these structures are easily integrated with robotic reasoning systems, they fail to represent the full complexity of human thought. They also frequently require hand-coding, an expensive and time-consuming process that is susceptible to errors of omission.

In this paper, we take an alternate approach to common-sense reasoning. Following the example of machine learning researchers, we model knowledge about the world as geometric points extracted from uncurated text corpora [5, 6, 7]. *Word embeddings* are trained based on local context, producing a model in which words that tend to appear in similar contexts are proximate to one another. Although these embeddings are trained exclusively based on word co-occurance, prior work [1, 8] has demonstrated that general purpose knowledge about the world is implicitly encoded in the resulting vector space. For example, it is possible to perform an *analogy query* by providing input of the form A:B :: C:D, where A, B, and C are given words and D must be inferred. For example, given 'Madrid:Spain :: Paris:D', an algorithm should return D=*France*, but to do that, the algorithm must 'know' in some sense that Paris is associated with France, which is a common-sense fact.

There is a problem, however. When presented with queries that have more than one correct answer, such as 'microwave:kitchen :: pillow:?', the traditional analogy method breaks down (usually by returning a synonym for one of the source words). Perhaps for this reason, word embeddings have seldom been seriously considered as potential knowledge bases for real robots.

This paper presents a novel algorithm for performing analogy queries under conditions which present multiple possible correct answers. The results constitute a valuable proof of concept that word embeddings implicitly encode common-sense facts that are useful to a robot, such as where

1st Conference on Robot Learning (CoRL 2017), Mountain View, United States.

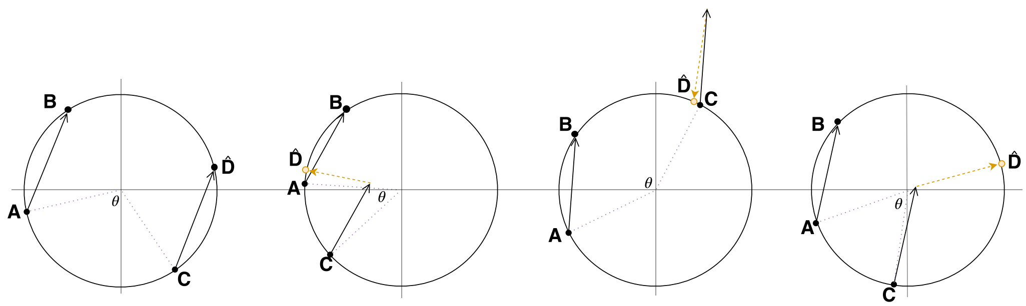

Figure 1: Vector operations in a 2-dimensional slice of an n-dimensional unit hypersphere. The three illustrations on the right represent a possible failure mode. In the middle two illustrations, when analogy vector $\overrightarrow{AB}$ is applied to point C, the resulting endpoint tends to lie close to one of A, B, or C. The specific failure depends on the angle θ and the length of $\overrightarrow{AB}$.

objects are commonly located and how one might best travel to a given destination. In particular, we show how *canonical vectors* can be used to capture high-level semantic queries, we illustrate that naive extrapolation is inappropriate for the hypersphere topology in which words are typically embedded, and we present an improved methodology for performing semantic queries about everyday objects and environments.

2 Related Work

In 2013, Mikolov et al. [6] presented word2vec, a learning algorithm that produced vector representations of words based on the ability of a neural network to predict their contexts. They are not the first to perform such a feat, but their machine learning approach was more flexible than the statistical methods [9, 10] that preceded them and more efficient than similar approaches which used more complex network architectures [11, 5, 12]. Perhaps most importantly, Mikolov et al. were able to observe a surprising array of linguistic regularities in the word vectors produced by their training algorithm, including the ability to find gender-paired terms, map countries to the names of their capitols, and identify the present and past tenses of verbs [1].

The result quickly caught the attention of the research community, raising the profile of related algorithms like GLoVE [7] and skip-thought vectors [13]. Multiple approaches have been proposed to improve the quality of the embedding space [14, 15] and to improve algorithm performance on analogical reasoning tasks [16, 17]. In situations where a specific type of word relationship is to be applied repeatedly, the use of a centroid or averaged analogy vector has proven quite effective [8, 18]. Other researchers have focused on algorithms that blend euclidean distance or cosine similarity with directional filters to select desired elements from a vector space. Lukac et al. [19] created a directional filter that minimizes the sum of weighted angular distances to remove impulses and outliers from an image, while Chen [20] examines a variety of weight functions for spatial autocorrelation.

Recent work in our laboratory applies these concepts at a more abstract level by using word embeddings to automatically detect *affordances* [8], meaning the set of actions that can be performed on a given object. The quality of extracted affordances was evaluated using simulated text-based environments, in which the agent's maximum score increased by over 75%. In this paper, we build on prior work by improving analogy quality and expanding the types of useful information that can be harvested from unstructured text.

3 Embeddings and Analogies on Hyperspheres

When discussing Mikolov et al.'s results, enthusiasts often describe analogy relations as a matter of simple vector addition: take the vector for the word 'king', subtract from it the vector for 'man' and add the vector for 'woman', and, *presto*! The result is the vector for 'queen'. This description is incomplete. In the tagged embedding space used for our experiments, the closest word to the point $\overrightarrow{king} - \overrightarrow{man} + \overrightarrow{woman}$ is not $\overrightarrow{queen}$, but instead $\overrightarrow{king}$. ($\overrightarrow{queen}$ comes in a close second.)

This tendency of analogy operations to produce results close to their origins is well documented and easily verifiable [21, 17]. Take any analogy pair A:B::C:D, and the vector corresponding to $\vec{C}$+$\vec{B}$-$\vec{A}$ is likely to lie closer to A, B, or C than to any other word in the embedding space.

Mikolov et al. addressed this problem by explicitly excluding all three analogy source words from the analogy results - a straightforward tactic that was introduced in [6] and works remarkably well. However, while this approach yields impressive results on syntactic analogies, it is far less effective on more challenging derivative and lexicographic analogies [17]. We hypothesize that this is due to synonyms and morphological forms of the source words, which would naturally be located nearby, and which would not be eliminated by source word exclusion.

In this paper, we take a different approach by considering the shape of the embedding space: Word2vec analogies are trained using softmax, a normalization function that centers vectors at the origin, thus constraining the final word representations to the surface of a unit hypersphere. Under these conditions, many translations of relational vector $\vec{r}$ formed using the method $\vec{r} = \vec{A} - \vec{B}$ will cause the endpoint of $\vec{r}$ to point away from the surface of the hypersphere. When an algorithm queries for the closest words to the endpoint of $\vec{r}$ after performing a translation on $\vec{r}$, the operation is roughly equivalent to projecting $\vec{r}$ onto the surface of the hypersphere and then performing the query. Therefore, a sufficiently large angle between $\vec{A}$ and $\vec{C}$ will result in a vector whose projection onto the hypersphere lies close to one or more of $\vec{A}$, $\vec{B}$, or $\vec{C}$ (see Figure 1).

Accordingly, we present DSM, a directional scoring method that compensates for this tendency to return to the source words of the analogy. Rather than excluding source words explicitly, DSM gives precedence to words that lie along an extension of the canonical analogy vector, thus decreasing the likelihood that source words or their semantic neighbors will be selected as an analogy response.

4 Improving Analogies with Canonical Vectors and Directional Scoring

The primary contribution of this work lies in the combination of an averaged relation vector with a directional scoring method (DSM) in order to navigate the embedding space more effectively. Rather than accepting Euclidean proximity as the sole determinant of relevance, we instead evaluate candidate solutions based on a combination of proximity to the the vector offset endpoint and orientation with respect to the analogy vector's trajectory. We show that DSM matches the performance of traditional offset methods on the Google Analogy Test Set, and that it outperforms them by a factor of 10% to 50% on a more challenging set of analogical reasoning tasks with multiple correct answers.

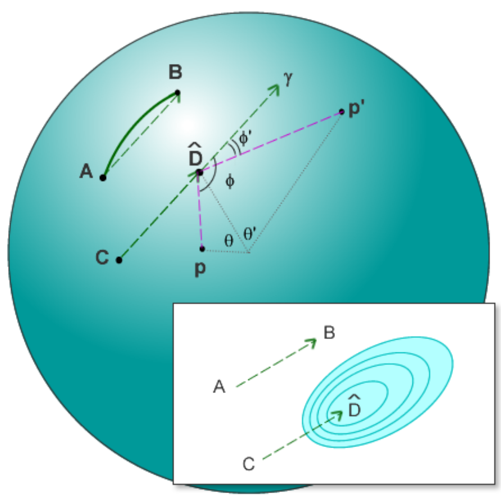

Figure 2: DSM calculation. If we take $\vec{z}$ = $\overrightarrow{AB}$ as a (degenerate) canonical analogy vector, then candidate analogy responses are scored based on the criteria $S_{dsm} = \alpha D_C(\vec{p} - \hat{D}, \vec{z}) + D_C(\vec{p}, \hat{D}) = \alpha(1 - cos(\phi)) + (1 - cos(\theta))$. Inset: DSM search pattern. Rather than selecting candidate points based solely on proximity, DSM searches in expanding rings that are elongated in the direction of the analogy vector.

4.1 Canonical analogy vectors

As noted by [8, 18, 22], analogy performance can be improved by averaging multiple examples of the type of relationship that is sought. DSM applies this principle as an initial step prior to directional scoring. A set of canonical examples A_i,B_i ϵ V is compiled, where A_i and B_i are natural language source words and V is the model's vocabulary. Let $\vec{A}_i$ and $\vec{B}_i$ be the vector representations of A_i and B_i and $\overrightarrow{AB}_i$ = $\vec{B}_i$ - $\vec{A}_i$. The canonical analogy vector $\vec{z}$ is defined as $\vec{z} = 1/n \sum_i^n \overrightarrow{AB}_i$ where n is the number of canonical examples used.

Algorithm 1 - Directional Scoring Method (DSM) Analogy

Inputs:
W = Tensor containing all word vectors in the embedding space. Note that all elements of W are already unit length.
c = Word-vector on which to apply analogy, following the convention **a:b :: c:d**.
γ = Canonical analogy vector
Parameters:
α = Scaling factor. In general practice, alpha=0.3 yields good results.
Output:
d' = Directionally scored analogy response, an element of W.

$\hat{\mathbf{d}} \leftarrow \mathbf{c} + \gamma$ // Analogy response for a standard offset analogy; center of search
$D_C \leftarrow 1 - \frac{W \cdot \hat{\mathbf{d}}^T}{\|\hat{\mathbf{d}}\|}$
$W' \leftarrow W - \hat{\mathbf{d}}$
for $w'_i \in W'$ **do**
  $\mathbf{h}_i \leftarrow \|w'_i\|$ // Building **h**, a list of vector norms
end for
$D'_C \leftarrow 1 - \frac{W' \cdot \gamma^T}{\mathbf{h}\|\gamma\|}$ // Element-wise division
$D_{dsm} \leftarrow \alpha D'_C + D_C$
$m \leftarrow min(D_{dsm})$
$i \leftarrow$ index of m in D_{dsm}
return W_i // This is our response word, **d'**

Ideally, the canonical examples should be high-quality exemplars of the relationship sought, and should use words whose meaning is unambiguous and not clouded by multiple possible interpretations. However, as demonstrated in Section 5, a group of less stringently selected canonical examples can also function effectively.

4.2 Scoring Algorithm

DSM uses a weighted sum of (a) cosine similarity and (b) alignment with the analogy vector in order to select candidate analogy solutions. The DSM score S_{dsm} of a word vector $\vec{p}$ is calculated as:

$$S_{dsm} = \alpha D_C(\vec{p} - \hat{D}, \vec{z}) + D_C(\vec{p}, \hat{D}) \tag{1}$$

where D_C is the cosine distance between two vectors, $\vec{z}$ is the canonical analogy vector, $\hat{D} = \vec{C} + \vec{z}$ is the endpoint of the offset operation, $\vec{C}$ is the vector representation of the natural language source word to which the offset operation was applied, and α is a scale factor. Intuitively, this equation can be viewed as attempting to simultaneously minimize both $\vec{p}$'s distance from endpoint $\hat{D}$ and the divergence of $\vec{p} - \hat{D}$ from the canonical analogy vector. The algorithm's preferred response word minimizes DSM score, as shown in in Algorithm 1. If a ranked list of response words is desired, the algorithm orders words by increasing DSM score and returns the top k candidates.

5 Quantitative Analysis

To evaluate the DSM algorithm, we compared four analogy variants on two test sets: The Google Analogy Test Set and the BYU Analogical Reasoning Dataset. Both of these evaluation tasks require the algorithm to infer the correct answer to analogies having the format A:B::C:D. Each task is broken into subtasks based on the type of analogy being evaluated.

5.1 Algorithms tested

We tested four algorithm variants:

Offset: The averaged canonical vector $\vec{z}$ is added to source word $\vec{C}$ to obtain point $\hat{D}$. Candidate solution words p_i are returned in order of increasing cosine distance from $\hat{D}$.

Offset with exclusion: The averaged canonical vector $\vec{z}$ is added to source word $\vec{C}$ as above, but the source word C is excluded from consideration when evaluating candidate solution words.

DSM: The averaged canonical vector $\vec{z}$ is added to source word $\vec{C}$ and candidate solution words are selected using the scoring method described in Section 4.2. No explicit exclusion of source words is applied.

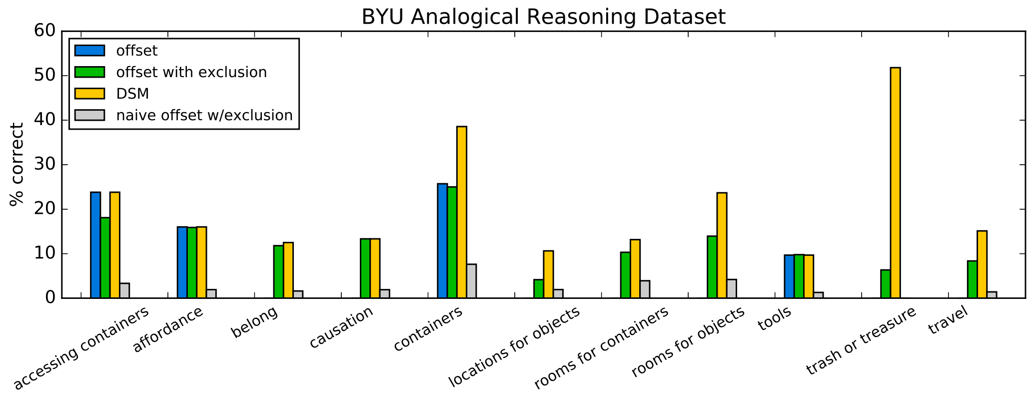

Figure 3: Algorithm performance on a sequence of analogical reasoning tasks, showing the % of queries for which the algorithm's *first* response word was the correct answer. Algorithms utilizing a canonical analogy vector are dramatically superior to the naive application of vector $\overrightarrow{AB}$ to $\overrightarrow{C}$. Directional scoring improves performance even further.

Naive offset with exclusion: Rather than using a canonical analogy vector, this method naively applies the vector $\overrightarrow{AB}$ to the source word $\overrightarrow{C}$ to obtain an offset endpoint $\hat{D}$. Candidate source words are selected in order of increasing cosine distance from $\hat{D}$, with source words A,B, and C excluded from consideration.

The embedding space used for this experiment is identical to that used in [8]. It was trained on a part-of-speech tagged Wikipedia text corpus using Mikolov et al.'s skip-gram method [6]. The final embedding space has 100 vector dimensions and a vocabulary size of approximately 1.5 million words and symbols. Canonical examples for the construction of $\vec{z}$ were taken from the AB and CD pairs of the first 10 entries of each analogy test set, with duplicates removed. After a coarse parameter search, we set DSM's α parameter to 0.3, a value that maximized performance for most (but not all) analogical subsets. We are currently investigating ways to initialize α programmatically as a function of the canonical examples.

5.2 Datasets used

The BYU Analogical Reasoning Dataset[1] is a newly created challenge task containing 11,846 analogies of the form A:B::C:D. Subtasks include analogies relevant to robotic navigation and object-based interaction. For example, the 'Containers' subtask requires an algorithm to correctly predict that 'brooms are in closets', 'silverware is in a drawer', 'milk is in a refrigerator', and so forth. The 'Rooms for Objects' subtask requires predictions about such things as 'refrigerator is in the kitchen', 'potatoes are in the cellar', and 'beds are in the bedroom', while the 'Tools' subtask requires knowledge about which objects can be used to enable certain actions, such as 'cutting requires a knife', 'digging requires a shovel', or 'baking requires an oven'. These hand-coded analogies are highly abstract and frequently include many-to-one relationships, making them particularly challenging.

The Google analogy corpus was introduced in [6], and is a standard benchmark.

5.3 Results

Our quantitative results highlight the difference in difficulty between the Google corpus and the BYU Analogical Reasoning Dataset. Whereas the Google dataset consists primarily of clearly-defined relationships with strict one-to-one correspondences[2], the BYU corpus contains abstract relationships for which multiple answers may seem equally correct. For example, milk can be contained in a jug, but it can also be contained in a bottle. The resulting analogy queries are much

[1] https://github.com/NancyFulda/BYU-Analogical-Reasoning-Dataset

[2] The countries/currency subcorpus is a notable exception, as many countries often share a monetary unit.

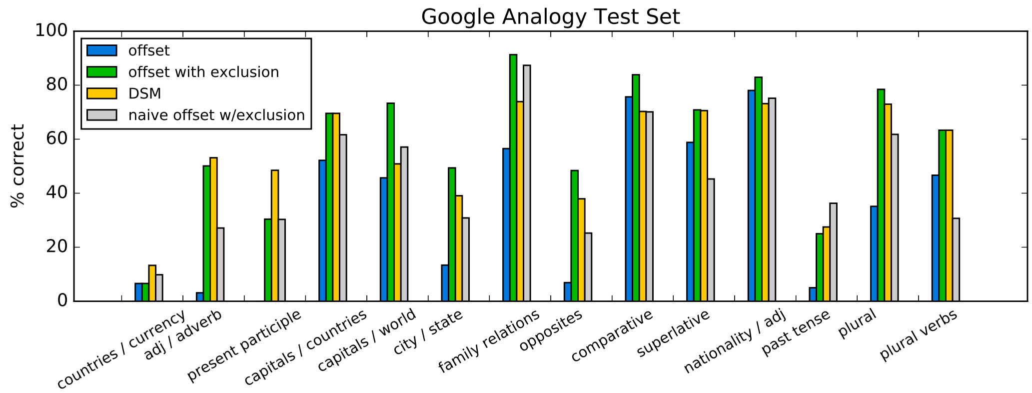

Figure 4: Algorithm performance on the Google Analogy Test Set, showing the % of queries for which the algorithm's *first* response word was the correct answer. On these primarily syntactic datasets, it is more difficult to identify a superior algorithm.

more difficult to answer correctly, although as we will see in Section 6, our DSM algorithm is able to return a high percentage of plausible responses.

On the Google dataset, naive offset methods perform almost as well as canonical methods (Fig. 4), whereas the more challenging analogical reasoning tasks require a canonical vector $\vec{z}$ in order to obtain passable results (Fig. 3). We are particularly intrigued to note that the performance of our directional scoring method is highly dependent on the specific reasoning task. Although DSM matches or exceeds the performance of traditional offset methods on all subcorpora of the BYU dataset, it sometimes produces improvements of 10% or more. Ongoing research in our laboratory suggests that this may result from distinct differences in the geometry of the source words A,B,C and D, as well as from the clustering behaviors of words within the hypersphere. Further research is required to determine the specific conditions under which DSM produces superior results.

6 Qualitative Analysis

Consider the thought experiment of a robotic household assistant, designated without loss of generality as 'RoButler'. To interact effectively with his human controller, RoButler requires common-sense reasoning abilities. A command to 'tidy up the living room' presupposes that he understands where common items like sofa pillows or magazines should be placed. When asked to 'bring me a glass of milk' he would need to determine that (a) milk is typically in the refrigerator, and (b) the refrigerator is typically in the kitchen. This information can then be connected to a planner.

We apply DSM as a tool to facilitate common-sense reasoning and display several sample queries below. Between 10 and 20 canonical examples were hand-selected for each common-sense reasoning task using entries from our Analogical Reasoning Dataset. None of the words in our canonical example set were used as source words for any of the sample queries below.

Query #1: Our first query uses canonical vectors extracted from the *travel* dataset, such as:

airport:car :: park:bike • hotel:taxi :: work:train • school:bus :: store:car

In this example, the offset method returns close synonyms of the source word, while DSM proposes candidate words that fit within the category 'methods of transportation'. Of those, five words (airplane, speedboat, helicopter, jet, boat) represent plausible answers to the original query. Overall, the candidates proposed by DSM are qualitatively superior, a pattern which holds across other queries we've tried using this analogy set. Even in cases when offset methods provide a correct first response, most of the other responses fail to fall into the correct word category.

Query #2: The next query uses canonical vectors extracted from the *locations for objects* dataset, which encodes knowledge about the locations of common household objects.

milk:refrigerator :: broom:closet • toaster:counter :: book:bookshelf

Traditional methods perform acceptably on this analogy, delivering a valid answer (‘refrigerator’) as the fourth response. Meanwhile, DSM scores a home run by delivering ‘refrigerator’ as its *first* response, followed by a list of plausible locations for alcoholic beverages.

Query #3: We now apply a query that uses analogical reasoning to perform a simple classification task: ‘Is this item garbage that should be disposed of, or is it a precious item which must be retained?’

wrapper:trash :: cup:treasure • peel:trash :: dirt:trash • toy:treasure :: sand:trash

The complete responses for two algorithms on the three queries are shown below. We note that similar results hold for other types of queries, such as ‘How do I get to the hospital?’ and ‘Are wood shavings trash?’ In addition, Table 1 shows proposed responses for three analogy sets given by DSM vs offset methods using a canonical analogy vector.

Query #1: How do I get to Hawaii?
source word: ‘Hawaii_NN’

OFFSET WITH EXCLUSION:	‘hawaii_NNS’, ‘hawaii_NNP’, ‘oahu_NN’, ‘hawaii_FW’, ‘hawaii_DT’, ‘hawaii_RB’, ‘honolulu_NN’, ‘hilo_NN’, ‘oahu_NNP’, ‘hawaii_ADD’, ‘maui_NN’, ‘kahului_NN’, ‘hilo_NNP’, ‘alaska_NN’
DSM:	‘airplane_NN’, ‘truck_NN’, ‘speedboat_NN’, ‘hawaii_NN’, ‘helicopter_NN’, ‘oahu_NN’, ‘jet_NN’, ‘hawaii_NNS’, ‘oahu_NNP’, ‘hawaii_NNP’, ‘hawaii_RB’, ‘kahului_NN’, ‘boat_NN’, ‘shuttle_NN’

Query #2: Where can I get a beer?
source word: ‘beer_NN’

OFFSET WITH EXCLUSION:	‘coffee_NN’, ‘keg_NN’, ‘schnapps_NNS’, ‘thermos_NN’, ‘refrigerator_NN’, ‘bottle_NN’, ‘bagel_NN’, ‘fridge_NN’, ‘drink_NN’, ‘coffeeshop_NN’, ‘chocolate_NN’, ‘brewed_VBD’, ‘delicatessen_NN’, ‘cask_NN’
DSM:	‘refrigerator_NN’, ‘fridge_NN’, ‘cellar_NN’, ‘kitchen_NN’, ‘jacuzzi_NN’, ‘pantry_NN’, ‘shop_NN’, ‘sauna_NN’, ‘restaurant_NN’, ‘parlor_NN’, ‘brewhouse_NN’, ‘luncheonette_NN’, ‘thermos_NN’, ‘coffeeshop_NN’,

Query #3: Is jewelry trash?
source word: ‘jewelry_NN’

OFFSET WITH EXCLUSION:	‘jewellery_NN’, ‘jewelery_NN’, ‘jewels_NNS’, ‘trinkets_NNS’, ‘jewellery_NNP’, ‘priceless_JJ’, ‘paraphernalia_NNS’, ‘valuables_NNS’, ‘mementos_NNS’, ‘antiques_NNS’, ‘furniture_NN’, ‘souvenirs_NNS’, ‘keepsakes_NNS’, ‘memorabilia_NNS’
DSM:	‘treasure_NN’, ‘treasures_NNS’, ‘priceless_JJ’, ‘valuables_NNS’, ‘jewelry_NN’, ‘jewels_NNS’, ‘jewellery_NN’, ‘trinkets_NNS’, ‘mementos_NNS’, ‘jewelery_NN’, ‘paraphernalia_NNS’, ‘souvenirs_NNS’, ‘memorabilia_NNS’, ‘antiques_NNS’

Overall, our directional scoring method provides valid responses approximately twice as often as offset methods. Since Table 1 shows that our directional scoring method will quite literally throw the baby out with the bathwater, DSM clearly is not a production-ready system. However, it does provide a valuable proof of concept: Critical common-sense knowledge about a wide variety of topics is implicitly encoded in the vector space. If methodologies can be improved far enough to extract this information reliably, it is potentially superior to hand-coded ontologies or knowledge graphs. Such a system would require virtually no maintenance, would not be prone to errors of

Analogy Set	Source Word	Offset + Exclusion	DSM
Travel	hawaii	hawaii* (0)	**airplane (5)**
	hospital	**ambulance** (5)	**train (8)**
	theater	scooter (2)	scooter **(4)**
	germany	germany* (0)	**truck (8)**
	mall	**car** (8)	**car (10)**
	footbridge	gondola (3)	gondola (3)
	australia	australia* (0)	**car (5)**
	office	ticket (5)	**cab (6)**
	moon	moon* (2)	moon **(3)**
	yosemite	yosemite* (2)	yosemite **(4)**
Locations for objects	beer	coffee (8)	**refrigerator (13)**
	bagel	doughnut (5)	cubicle **(8)**
	knife	noose (1)	noose (3)
	shoe	handbag (0)	drawer **(1)**
	spoon	**tray** (2)	bookcase **(3)**
	remote	nunchuk **(1)**	remote (0)
	necklace	locket (2)	locket **(3)**
	marbles	**bookcase** (5)	**bookcase** (7)
	backpack	suitcase (0)	suitcase **(2)**
	bracelet	locket (3)	wallet (3)
Trash or Treasure	jewelry	jewellery (0)	**treasure (1)**
	pit	pits (1)	pit **(1)**
	dust	**trash** (1)	**trash** (1)
	iphone	iphone* (0)	wii **(1)**
	shavings	scraps (1)	**trash** (1)
	apple	sundog (0)	**treasure (1)**
	baby	mommy (0)	trash **(1)**
	bathwater	**trash** (1)	bathwater (1)
	money	loot (0)	money **(1)**
	dictionary	cyclopaedia (0)	dictionary **(1)**
Number Correct	-	6 (58)	12 (109)

Table 1: Analogy solutions provided by traditional offset methods and DSM when using a canonical averaged analogy vector. The first response provided by each algorithm is shown alongside the number of plausible responses returned (out of a total of 15). Entries are bolded if either the first response was valid or the number of valid responses was greatest. POS tags have been omitted for clarity. *In these cases, the method returned a differently-tagged version of the source word.

omission, could be updated automatically in response to recent news and scientific breakthroughs, and offers possibilities for domain-specific knowledge via the selection of training corpus.

7 Conclusion

Robotic systems typically rely on ontologies and knowledge graphs for common-sense reasoning. This paper presents an alternative option - word embeddings trained using uncurated text corpora - and demonstrates that a broad spectrum of common-sense knowledge is implicitly encoded within the vector space. This information could potentially be used by a robot that relies on natural language to reason about planning tasks. We have introduced a directional scoring method that nearly doubles rates of correct first responses while simultaneously increasing overall response accuracy. This suggests that with further development of these semantic query algorithms, including adjustments made to compensate for the spherical structure of word embeddings trained using softmax, it may be possible to improve the query performance even further. Finally, we have introduced a new Analogical Reasoning Dataset that can be used to benchmark progress in this area, with the hope that other researchers will join us in seeking to create robotic systems that are able to accept high-level commands, adapt these instructions to changing environments, and behave appropriately without micromanagement.

References

[1] T. Mikolov, W. tau Yih, and G. Zweig. Linguistic regularities in continuous space word representations. Association for Computational Linguistics, May 2013.

[2] M. Tenorth and M. Beetz. Knowrob: A knowledge processing infrastructure for cognition-enabled robots. 32:566–590, 04 2013.

[3] A. Saxena, A. Jain, O. Sener, A. Jami, D. K. Misra, and H. S. Koppula. Robobrain: Large-scale knowledge engine for robots. *CoRR*, abs/1412.0691, 2014. URL `http://arxiv.org/abs/1412.0691`.

[4] M. Waibel, M. Beetz, R. D'Andrea, R. Janssen, M. Tenorth, J. Civera, J. Elfring, D. Gálvez-López, K. Häussermann, J. Montiel, A. Perzylo, B. Schießle, O. Zweigle, and R. van de Molengraft. RoboEarth - A World Wide Web for Robots. *Robotics & Automation Magazine*, 18(2): 69–82, 2011.

[5] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, 2003.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.

[7] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.

[8] N. Fulda, D. Ricks, B. Murdoch, and D. Wingate. What can you do with a rock? affordance extraction via word embeddings. In *Proceedings of the International Joint Conference on Artifial Intelligence (IJCAI)*, 2017.

[9] P. D. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *J. Artif. Int. Res.*, 37(1):141–188, 2010. ISSN 1076-9757.

[10] Y. Liu, Z. Liu, T. seng Chua, and M. Sun. Topical word embeddings. In *In TwentyNinth AAAI Conference on Artificial Intelligence*, 2015.

[11] G. E. Hinton. Learning distributed representations of concepts. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pages 1–12. Hillsdale, NJ: Erlbaum, 1986.

[12] M. Baroni and G. Dinu. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *In ACL*, 2014.

[13] R. Kiros, Y. Zhu, R. Salakhutdinov, R. S. Zemel, A. Torralba, R. Urtasun, and S. Fidler. Skip-thought vectors. *CoRR*, abs/1506.06726, 2015.

[14] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *NIPS*, pages 3111–3119. Curran Associates, Inc., 2013.

[15] Q. V. Le and T. Mikolov. Distributed representations of sentences and documents. *CoRR*, abs/1405.4053, 2014.

[16] O. Levy, Y. Goldberg, and I. Dagan. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3: 211–225, 2015.

[17] A. Gladkova, A. Drozd, and S. Matsuoka. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In *Proceedings of the NAACL Student Research Workshop*, 2016.

[18] G. Brokos, P. Malakasiotis, and I. Androutsopoulos. Using centroids of word embeddings and word mover's distance for biomedical document retrieval in question answering. *CoRR*, abs/1608.03905, 2016. URL `http://arxiv.org/abs/1608.03905`.

[19] R. Lukac, B. Smolka, K. N. Plataniotis, and A. N. Venetsanopoulos. Selection weighted vector directional filters. *Computer Vision and Image Understanding*, 94(1):140 – 167, 2004. ISSN 1077-3142. doi:http://dx.doi.org/10.1016/j.cviu.2003.10.013. URL http://www.sciencedirect.com/science/article/pii/S107731420300198X. Special Issue: Colour for Image Indexing and Retrieval.

[20] Y. Chen. On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5(2):65–72, 2012.

[21] T. Linzen. Issues in evaluating semantic spaces using word analogies. *CoRR*, abs/1606.07736, 2016. URL http://arxiv.org/abs/1606.07736.

[22] A. Drozd, A. Gladkova, and S. Matsuoka. Word embeddings, analogies, and machine learning: Beyond king - man+ woman= queen. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, 2016.