UNAM at SemEval-2018 Task 10: Unsupervised Semantic Discriminative Attribute Identification in Neural Word Embedding Cones

Ignacio Arroyo-Fernández
Universidad Nacional
Autónoma de México (UNAM)
`iaf@ccg.unam.mx`

Ivan Meza
Instituto de Investigaciones
en Matemáticas Aplicadas
y en Sistemas – UNAM
`ivanvladimir@turing`
`.iimas.unam.mx`

Carlos F. Méndez-Cruz
Centro de Ciencias
Genómicas – UNAM
`cmendezc@ccg.unam.mx`

Abstract

In this paper we report an unsupervised method aimed to identify whether an attribute is discriminative for two words (which are treated as concepts, in our particular case). To this end, we use geometrically inspired vector operations underlying unsupervised decision functions. These decision functions operate on state-of-the-art neural word embeddings of the attribute and the concepts. The main idea can be described as follows: if attribute q discriminates concept a from concept b, then q is excluded from the feature set shared by these two concepts: the intersection. That is, the membership $q \in (a \cap b)$ does not hold. As a, b, q are represented with neural word embeddings, we tested vector operations allowing us to measure membership, i.e. fuzzy set operations (t-norm, for fuzzy intersection, and t-conorm, for fuzzy union) and the similarity between q and the convex cone described by a and b.

1 Introduction

There exist nowadays a number of arithmetic vector operations for computing word relationships interpreted as linguistic regularities. A very popular setting is solving word analogies (Lepage, 1998), which is mainly used to evaluate the quality of word embeddings (Mikolov et al., 2013). Recently other alternatives to solve word analogies have been proposed (Linzen, 2016), including supervised methods (Drozd et al., 2016).

Solving word analogies requires three word arguments, and a fourth one is inferred. Such an inference raises from the similarity between common or similar contexts shared by the two pairs of words. Thus, given words "queen", "woman", "king", "man", the following arithmetic operation holds for their corresponding embeddings $x_{(\cdot)}$:
$x_{king} - x_{man} + x_{woman} = x_{queen}$.

In this work, we explore similar approaches for Discriminative Attribute Identification (DAI). This task requires tree word arguments a, b, q, and a binary label $y \in \{0, 1\}$ is inferred from them (Cree and McRae, 2003; Lazaridou et al., 2016; McRae et al., 2005). Such a label indicates whether the third word, q, is identified as a discriminative (semantic) attribute between words (concepts) a, b. We observed that the task of identifying discriminative attributes between words, represented via word embeddings, evokes that of solving word analogies.

We propose geometrically inspired vector operations on word embeddings $x_a, x_b, x_q \in \mathbb{R}^n$ of the words a, b, q, respectively. The output of each of these operations is in turn operated by a unsupervised decision function aimed to predict the label y. The decision functions are based on the reasoning given originally in (Lepage, 1998) for solving word analogies. Under this reasoning, the important thing is to look for those items shared by the objects compared, and verify whether the item of interest is included among them.

In other words, in the case of DAI, if we are asked whether x_q, the attribute embedding, discriminates x_a from x_b, then an idea is to verify whether the attribute is contained in the set shared by the two concepts in question, i.e. does the set operation $q \in (a \cap b)$ hold? Our hypothesis, is that x_q discriminates x_a from x_b if the result of such an operation is false in terms of the subspace delimited by x_a and x_b, i.e. a convex cone. Thus, a number of vector operations and decision functions were tested as different vector versions of this set operation on state-of-the-art neural word embeddings.

The proposed method does not rely on language or knowledge resources (i.e. knowledge bases and graphs, PoS or any kind of taggers, etc.). Furthermore, with the help of the geometrical insight that our method provides, we also discuss the possibilities of it for being used to study measures of how concepts can be generated from attributes in the sense of vector space modeling of natural language. Thus, this study can be considered, e.g., for designing semantically driven word embedding methods or to explore alternatives for building knowledge resource applications.

Our results showed that the proposed approach hold coherence with respect to the semantic notions proposed in the DAI task. This approach reached 0.622 of F-measure in predicting discriminative attributes.

2   Literature Review

Up to our knowledge, there is not work proposing unsupervised methods for discriminative attribute identification or extraction with a direct link to word embeddings. Most related work deals with semantic relation extraction or with labeling semantic relations in lexical semantics, e.g. given a hypernym, to perform hyponym extraction (Fu et al., 2014).

There is also work on using semantic attributes to classify images of objects in a supervised fashion (Chen et al., 2012; Lazaridou et al., 2016). In this case, dictionaries of discriminative attributes of objects are used (e.g. fruits by their *color or form*), but experiments are not performed on text data, e.g., a snippet describing the object. In more applicative cases, the use of dictionaries of object attributes has shown to be a good approach in clothing recommender systems. These systems group images of items sharing attributes the customers are usually interested in, e.g. images of jackets *with a hat* (Chen et al., 2012; Kalra et al., 2016; Zhou et al., 2016).

Other contributions provide methods for object classification by using multiple data sources, including text. In (Farhadi et al., 2009, 2010; Lampert et al., 2009) it is proposed supervised learning of semantic attributes and textual descriptions of objects. Their methods are aimed to generalize recognition and (template) textual description of unseen objects with similar and shared attributes. In the particular case of (Berg et al., 2010), a supervised algorithm learns to label object attributes

by fitting multinomial associations between text segments and recognized image segments as co-occurring objects within a Web corpus (Su and Jurie, 2012). After that, the learned attributes of the objects are detected and used as features for feeding an unsupervised method for categorizing images and text. In (Deeptimahanti and Sanyal, 2011; Overmyer et al., 2001) natural language descriptions of case uses of user requirements are parsed to obtain Unified Modeling Language (UML) diagrams including object attributes. This approach is aimed to facilitate design in software engineering by semi-automatically building code objects (Yue et al., 2011).

3   A Convex Combination

Lepage (1998) proposed solving word analogies based on characters shared by words and sentences. We extrapolated such an idea for feature similarities, similarly to what (Mikolov et al., 2013) did on vectors for solving word analogies. Thus, our method attempts taking into account these two ideas in the following way. Thinking about neural word embeddings as vectors generated by axis of attributes, our approach is to observe the linear subspace A delimited by embeddings of words a and b, and to see how the embedding of q is contained in it. This linear subspace has the properties of a convex cone. Thus, in geometrical terms, to assess whether an attribute can generate a pair of concepts or not, we propose to measure the degree, λ, to which the embedding x_q is a convex combination of the embeddings x_a and x_b. This measure can be derived from the convex combination

$$\lambda x_a + (1 - \lambda)x_b = x_q, \qquad (1)$$

where $\lambda \in [0, 1]$. The embeddings $x_a, x_b, x_q \in \mathbb{R}^d$ represent nouns a and b and the query attribute q, respectively (see Figure 1). The requirement of x_a, x_b to describe a convex cone A is due to the fact that, geometrically, features shared by these embeddings would be enclosed within such cone. This can be observed by testing extreme values in Eq. (1). Assume all embeddings are normalized in magnitude. Let us making q to be, simultaneously, as far as possible from a and b while keeping the volume of A greater than zero. Also make that $\langle x_a, x_b \rangle$ to be small. In this scenario, embeddings x_a, x_b delimit a cone of less than 90 degrees. As the embedding x_q is as far as possible from the
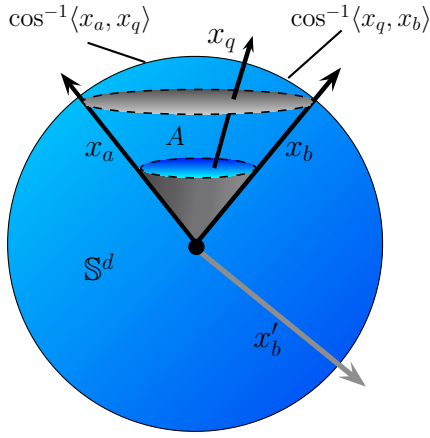
Figure 1: The $d-$dimensional vector space defined in the unitary-sphere $\mathbb{S}^d$.

x_a and x_b and it is contained in A, then it passes close to the center of the circular basis of the cone. Thus, we have

$$\|x_q - x_a\| \approx \|x_q - x_b\|.$$

This geometrical scenario indicates that q is shared equally by a and b, so it is not discriminative for them. In the case of $\langle x_a, x_b \rangle \approx -1$ it means that the set A is not convex. This is because x_a and x_b describe a unique line and have opposite directions (they are anti-parallel). This geometrical scenario prevents the pair of word embeddings from being generated by linear combinations of attributes in common to them (see x_a and x_b' in Figure 1). In the case when a, b are semantically very similar, we have that[1] $\langle x_a, x_b \rangle \to 1$. It means that both vectors are (almost) parallel, so they refer concepts sharing most of their attributes. In this case, if x_q is far away from both x_a and x_b, then determining discriminativeness has not sense (probably x_q is not an attribute of either of them).

The geometrical scenario of identifying a discriminative attribute q can occur when $\langle x_a, x_b \rangle$ is small and either $\langle x_q, x_a \rangle \to 1$ or $\langle x_q, x_b \rangle \to 1$. For example, if $\langle x_q, x_b \rangle \to 1$, it means that x_q tends to be parallel to x_b and we can see x_b as a linear combination of x_q. As $\langle x_a, x_b \rangle$ is small (x_a and x_b are almost orthogonal), then $\langle x_a, x_q \rangle$ is also small. Therefore, this analysis leads us to think that q discriminates a from b, and that q is an attribute of b rather than of a.

---

[1]The symbol '$\to$' denotes tendency ("*tends to*").

4 The Convex Cone Method

The scenarios depicted in Section 3 overall show how projections among word embeddings form convex combinations and how these projections can be exploited in DAI. Without loss of generality, these projections can be seen as distances. In this sense, the convex parameter λ in Eq. (1) indeed weighs distances involving x_a, x_b, x_q. Now, notice that Eq. (1) expresses x_q in terms of x_a and x_b. However, in DAI they are know and we would like to measure the relationship among them given they are $d-$dimensional vectors. This measure is can be given by λ, which now becomes into the unknown. In this case, λ acts as a bounded measure of how much a given pair of concepts a, b shares a given attribute q. Thus, by performing some comprehensive algebra starting from Eq. (1), we arrive at

$$\lambda(x_a - x_b) + x_b = x_q,$$

which leads to the $d-$dimensional Euclidean (*cone*) version of the convex parameter

$$\lambda = \frac{\|x_q - x_b\|}{\|x_a - x_b\|}. \tag{2}$$

Furthermore, in addition to (2), we consider an alternative distance criterion. That is, it is possible measuring distance in terms of arcs instead of doing it in terms of straight line segments. Therefore, we have the arc (*arcone*) version of the convex parameter:

$$\lambda = \frac{\cos^{-1}\langle x_q, x_b \rangle}{\cos^{-1}\langle x_a, x_b \rangle}, \tag{3}$$

where $\langle x, x' \rangle \in [-1, 1]$ given that $\|x\| = 1$ for all $x \in \mathbb{S}^d$ (the unitary sphere). Both arcs $\cos^{-1}\langle x, x' \rangle$ in the numerator and in the denominator of (3) are in the interval $[0, 2\pi]$.

The convex parameter λ measures the degree to which x_q is a convex combination of x_a and x_b. Form the point of view of the combination degree, rather than from the point of view of the absolute value of λ, some function $f(\lambda)$ must be maximum at $\lambda = 0.5$ (see Figure 2). When it occurs, x_q passes close to the axis of the cone A (so it also passes close to the center of the shaded circular area of radius $0.5\|x_a - x_b\|$ in Figure 1). Therefore, $\lambda \to 0.5$ indicates that the attribute q is highly shared by both concepts a and b.

The extreme values of λ must be interpreted contrarily by f, i.e. $\lambda \to 0$ means that, on the
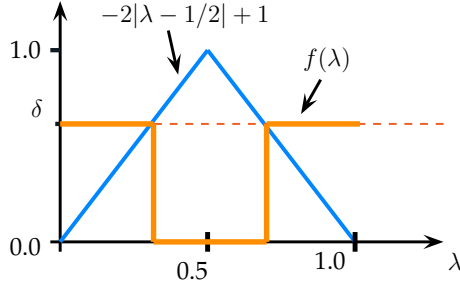
Figure 2: The decision function $f(\lambda)$.

one hand, the attribute q uniquely characterizes (or generates) the concept a, so x_a is approximately parallel to x_q. On the other hand, $\lambda \to 1$ means that the attribute q uniquely characterizes the concept b. Thus, we need that some decision function f to take advantage of extreme values of λ for making decision on whether an attribute q is discriminative of a pair of concepts a and b.

Therefore, we define our decision criterion subject to some threshold $\delta \in [0, 1]$ (say $\delta = 0.7$):

$$f(\lambda) = \begin{cases} 1 & \text{if } -2|\lambda - 1/2| + 1 < \delta \\ 0 & \text{if } -2|\lambda - 1/2| + 1 \geq \delta \end{cases} \quad (4)$$

where if upper inequality (4) holds, it means either that $\lambda \to 0$ or that $\lambda \to 1.0$, so $f(\lambda) = 1$ and therefore attribute q discriminates concepts a and b. Conversely, if lower inequality (4) holds, it means that $\lambda \to 0.5$. Therefore, $f(\lambda) = 0$ and the decision function determines that q does not discriminate a and b. See Figure 2.

5 Other Geometrical Methods

In addition to the convex cone method, we also tested mean-based, sum-based and fuzzy methods for quantifying the containment $q \in (a \cap b)$.

5.1 Similarity with Respect to the Sum and to the Mean

The sum-based method computes the resultant vector of x_a and x_b. The similarity between such a vector and the candidate attribute x_q should be smaller than some threshold δ so as to consider that q discriminates a from b, that is:

$$f(x_q; x_a, x_b) = \begin{cases} 1 & \text{if } \langle x_q, x_a + x_b \rangle < \delta \\ 0 & \text{if } \langle x_q, x_a + x_b \rangle \geq \delta \end{cases} \quad (5)$$

Unlike to the convex cone method, Eq. (5) indicates that the sum-based method measures directly the similarity between the resultant vector $x_a + x_b$

and x_q. The motivation of this operation is similar to that of the convex cone method. That is, $x_a + x_b$ is an embedding that embeddings x_a and x_b have in common. Therefore, probably such embedding is similar to x_q if this latter also is common to x_a and x_b.

The mean-based method follows exactly the same principle, but only requires multiplying $x_a + x_b$ in (5) by 0.5.

5.2 Similarity with Respect to a Fuzzy Connective

The fuzzy method computes the connective:

$$x_{\{a,b\}} = \alpha \min\{x_a, x_b\} + (1 - \alpha) \max\{x_a, x_b\}$$

between the fuzzy intersection ($\min\{\cdot\}$) and the fuzzy union ($\max\{\cdot\}$) of the embeddings x_a, x_b (Zadeh, 1965). These set operations are known as the Gödel's t-norm and t-conorm (Klement et al., 2013), respectively and they are defined element-wise for vectors. α is known as the compensation parameter and controls the mixture between union and intersection. Thus, the connective acts as a convex combination of the fuzzy union and the fuzzy intersection operators, so if $\alpha \to 0$ it causes that the intersection ($\min\{\cdot\}$) vanishes whereas the union ($\max\{\cdot\}$) survives. The contrary effect can be induced if $\alpha \to 1$.

Fuzzy set operations are conceptually more akin to the idea of observing whether the intersection set of concept attributes contains some query attribute. To contextualize word embeddings with fuzzy sets, we assume the embedding $x_a \in \mathbb{R}^d$ is given by a membership function $x_a = \mu(\mathcal{A})$. Herein, $\mathcal{A}$ is the set of items in some subset (of cardinality d) of the contexts of the word a. We also assume that the subset of contexts was statistically estimated by the word embedding method, which is in this case though as the membership function defined on the set $\mathcal{C} \supset \mathcal{A}$ of all contexts in the corpus $\mu : \mathcal{C} \to \mathbb{S}^d$.

As a first attempt to explore a relationship between fuzzy sets and word embeddings, in this paper we induced bias α to a decision function f based on the inner product between the connective $x_{\{a,b\}}$ (a biased version of $x_a + x_b$) and the query attribute x_q. In this way, the decision of DAI is made according to the threshold δ, i.e.:

$$f(x_q; \alpha; x_a, x_b) = \begin{cases} 1 & \text{if } \langle x_q, x_{\{a,b\}} \rangle < \delta \\ 0 & \text{if } \langle x_q, x_{\{a,b\}} \rangle \geq \delta \end{cases}$$
$$(6)$$

where α is the tolerance parameter of the fuzzy connective and it must be manually set.

6 Experiments and Results

For our experiments, we computed our decision functions $f(\cdot)$ on tuples of the form $\{x_a, x_b, x_q, y\}$. To this end, we used state-of-the-art word embeddings, i.e. Glove (Pennington et al., 2014), FastText (Bojanowski et al., 2016), Word2Vec (Mikolov et al., 2013) and Dependency-Based Word2Vec (DBW2V) (Levy and Goldberg, 2014). We also explored embeddings tanking into account external knowledge. This is the case of ConceptNet embeddings (Speer and Lowry-Duda, 2017). DBW2V embeddings are W2V embeddings enriched by using syntactic dependencies and Conceptnet are embeddings enriched with both syntactic dependencies and knowledge graphs (Faruqui et al., 2015). We trained W2V and FastText by using the Wikipedia dataset[2]. In the case of Glove[3], ConceptNet[4] and DBW2V[5] we downloaded pretrained embeddings from authors' websites. For Word2Vec and FastText we trained models of $200, 300, 400, 500$ and 1000 dimensions. In our results we only report the dimensionality that performed best.

As our approach is unsupervised, we report experiments on the validation dataset available on the competition's repository[6]. We can see in Table 1 that the arcone operation defined in (3) provided the best results for all word embedding methods. Our general best result was obtained by using Glove embeddings of 300 dimensions. We expected a good result from these embeddings as they specifically learn from mutual information statistics of word pairs. This enables Glove to encode feature contrasts, which also allows it for being the state-of-the-art method in word analogy tasks. During the competition we submitted our best configuration as unique run (Glove $300d$ and arcone operation with $\delta = 0.4$), which gave us $F1 = 0.60$ (place 19/26).

Regarding to the threshold δ of the decision functions $f(\cdot)$, we tested a set of values $\delta \in \{0.0, 0.1, 0.4, 0.7, 1.0\}$. Our best result was obtained when $\delta = 0.4$ for almost all embedding methods, excepting DBW2V. This means that the convex parameter λ can vary 60% around the maximum (0.5) in order to consider that an attribute q is shared by (or to generate) both concepts a, b. Thus, by evaluating δ in (4) we see either that x_q is too biased towards x_a if it holds that $\lambda < 0.4(0.5) = 0.2$, or that x_q is too biased towards x_b if it holds that $\lambda > 0.3 + 0.5 = 0.8$. In these cases we can say that q is discriminative for the concepts a, b as it is an attribute only (or mainly) of one of them.

Embedding	Dim.	$f(\cdot)$	δ	F1-score
Glove	300	arcone	0.4	**0.622**
		cone	0.4	0.615
		sum	-0.4	0.457
		fuzzy	-0.4	0.438
		mean	-0.4	0.426
ConceptNet	300	arcone	0.4	0.585
		cone	0.4	0.581
		mean	0.1	0.469
		fuzzy	-0.1	0.465
		sum	-0.4	0.451
Word2Vec	300	arcone	0.4	0.584
		cone	0.4	0.577
		fuzzy	-0.1	0.448
		mean	-0.1	0.444
		sum	-0.4	0.439
FastText	500	arcone	0.4	0.570
		cone	0.4	0.568
		fuzzy	0.4	0.451
		mean	0.4	0.450
		sum	0.7	0.437
DBW2V	300	arcone	0.7	0.541
		cone	0.7	0.536
		sum	-0.7	0.498
		fuzzy	-0.4	0.485
		mean	-0.4	0.475

Table 1: Best results for all the word embedding and vector operation methods on the validation dataset.

Given that it was needed $\delta = 0.7$ for DBW2V, we inferred that these embeddings allowed much less bias from the center of the cone and λ must be within 30% of its maximum in order to decide that an attribute is shared by two concepts. In other

words, with DBW2V, it is more difficult to distinguish whether the attribute q is discriminative of a, b because it is allowed to be distant from both them even when it can be discriminative. This condition allows for much more feature overlapping and therefore the ranking on bottom of these embeddings can be explained.

Notice that the Euclidean version of the cone vector operation was the second best method for all word embedding methods. In fact, no difference was registered greater than 0.7% between cone and arcone operations.

The fuzzy approach did not show noticeable results. The variation of both, the threshold and the compensation parameter.

7   Discussion

We consider a bit surprising the difference in performance of Glove with respect to knowledge-based (ConceptNet) and Dependency-based (DBW2V) embeddings: 5.9% with respect to ConceptNet and 13.0% with respect to DBW2V. Such embeddings were expected to provide much more information about discriminative features because they are trained by taking into account semantic features explicitly by using knowledge and language resources for training.By using our arcone vector operation, W2V was ranked barely next to ConceptNet with a small difference of 0.17%. We think there are three possible motivations for this behavior. The first one is that the nature of our decision functions did not allowed to capture semantic features embedded into ConceptNet and into DBW2V. The second possibility is that semantic features are better embedded by Glove and, the third possibility, is that embedding semantic features explicitly can lead to overfitting of the resulting word representations. This latter possibility could be an additional explanation that DBW2V ended at bottom of our ranking.

In the case of FastText, these embeddings have been tested in word analogy tasks with success. However, as in the case of DBW2V, they are better than W2V or Glove mainly for syntactic analogies, which probably makes better FastText (and probably DBW2V) for NLP tasks other than DAI, e.g. sentence representation (Arroyo-Fernández et al., 2017).

Some assumptions were made for practical reasons in the case of fuzzy set operations. We are aware that this could affected drastically the results. The first assumption was that word embeddings were produced by membership functions, which take values in $[0, 1] \subset \mathbb{R}$ exclusively. This is not the case of word embeddings and they cannot directly mapped to identifiable textual items. Therefore, applying the t-norm and the t-conorm to these vectors is not completely intuitive. Nevertheless, with real-valued vectors we still had: as both embeddings tend to be in the same quadrant, the larger the magnitude of the connective embedding $x_{\{a,b\}}$. This latter embedding is somewhat oriented to the direction of the resultant $x_a + x_b$, which can be regulated by α, inducing a bias with respect to that direction. Although, this interpretation was worth exploring it did not gave us interesting results. Thus a better version of this fuzzy approach is pending.

At this moment, we have not clear what was the reason several of our results were contradictory with respect to the F-measure. Particularly for distributed representations. That is, we have balanced binary labels in the gold standard, but some scores resulted less than 50%. It is difficult to figure out how it happened analyzing directly distributed representations. Therefore, it remains an open issue proposing an alternative geometrical approach to tackle this inconsistency with respect to the main hypothesis of this paper.

8   Conclusions

The results of our experiments showed that the arcone vector operation is a simple method for quantifying discriminativeness. This operation showed to be correlated with respect to human judgments annotated in the validation dataset when Glove word embeddings were used. From the vector operations presented in this paper, the arcone operation, Eq. (3), best represents the abstract operation between sets $a \cap b = A$. Notice that the concept of cone is limited to euclidean metrics neither on $\mathbb{R}^d$ nor on $\mathbb{S}^d$. Therefore, other kind of transformations and related theories can be explored.

The effectiveness of our approach can be further explored as part of a learning algorithm aimed to obtain specialized (or enriched) word embeddings such that their geometrical structure is fitted in sets of convex volumes. An immediate experiment is using vector operations proposed in this paper as restrictions or as objectives for learning such embeddings for building knowledge resources.

References

Ignacio Arroyo-Fernández, Carlos-Francisco Méndez-Cruz, Gerardo Sierra, Juan-Manuel Torres-Moreno, and Grigori Sidorov. 2017. Unsupervised sentence representations as word information series: Revisiting TF–IDF. *arXiv preprint arXiv:1710.06524* .

Tamara L. Berg, Alexander C. Berg, and Jonathan Shih. 2010. Automatic attribute discovery and characterization from noisy web data. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*. Springer Berlin Heidelberg, Berlin, Heidelberg, pages 663–676.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606* .

Huizhong Chen, Andrew Gallagher, and Bernd Girod. 2012. Describing clothing by semantic attributes. *Computer Vision–ECCV 2012* pages 609–623.

George S Cree and Ken McRae. 2003. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General* 132(2):163.

Deva Kumar Deeptimahanti and Ratna Sanyal. 2011. Semi-automatic generation of UML models from natural language requirements. In *Proceedings of the 4th India Software Engineering Conference*. ACM, New York, NY, USA, ISEC '11, pages 165–174.

Aleksandr Drozd, Anna Gladkova, and Satoshi Matsuoka. 2016. Word embeddings, analogies, and machine learning: Beyond king-man+ woman= queen. In *COLING*. pages 3519–3530.

Ali Farhadi, Ian Endres, and Derek Hoiem. 2010. Attribute-centric recognition for cross-category generalization. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* pages 2352–2359.

Ali Farhadi, Ian Endres, Derek Hoiem, and David A. Forsyth. 2009. Describing objects by their attributes. *2009 IEEE Conference on Computer Vision and Pattern Recognition* pages 1778–1785.

Manaal Faruqui, Jesse Dodge, Sujay K. Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of NAACL*.

Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu. 2014. Learning semantic hierarchies via word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. volume 1, pages 1199–1209.

Bhavya Kalra, Kingshuk Srivastava, and Manish Prateek. 2016. Computer vision based personalized clothing assistance system: A proposed model. In *Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on*. IEEE, pages 341–346.

Erich Peter Klement, Radko Mesiar, and Endre Pap. 2013. *Triangular norms*, volume 8. Springer Science & Business Media.

C. H. Lampert, H. Nickisch, and S. Harmeling. 2009. Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. pages 951–958.

Angeliki Lazaridou, Nghia The Pham, and Marco Baroni. 2016. The red one!: On learning to refer to things based on their discriminative properties. *arXiv preprint arXiv:1603.02618* .

Yves Lepage. 1998. Solving analogies on words: an algorithm. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, pages 728–734.

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings. In *ACL (2)*. pages 302–308.

Tal Linzen. 2016. Issues in evaluating semantic spaces using word analogies. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*. Association for Computational Linguistics, Berlin, Germany, pages 13–18.

Ken McRae, George S Cree, Mark S Seidenberg, and Chris McNorgan. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods* 37(4):547–559.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. pages 3111–3119.

Scott P. Overmyer, Benoit Lavoie, and Owen Rambow. 2001. Conceptual modeling through linguistic analysis using LIDA. In *Proceedings of the 23rd International Conference on Software Engineering*. IEEE

Computer Society, Washington, DC, USA, ICSE '01, pages 401–410.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. pages 1532–1543. http://www.aclweb.org/anthology/D14-1162.

Robert Speer and Joanna Lowry-Duda. 2017. ConceptNet at semeval-2017 task 2: Extending word embeddings with multilingual relational knowledge. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, pages 85–89.

Yu Su and Frédéric Jurie. 2012. Improving image classification using semantic attributes. *International journal of computer vision* 100(1):59–77.

Tao Yue, Lionel C. Briand, and Yvan Labiche. 2011. A systematic review of transformation approaches between user requirements and analysis models. *Requirements Engineering* 16(2):75–99.

L.A. Zadeh. 1965. Fuzzy sets. *Information and Control* 8(3):338 – 353.

Jingjin Zhou, Zhengzhong Zhou, and Liqing Zhang. 2016. Hierarchical semantic classification and attribute relations analysis with clothing region detection. In *Advanced Multimedia and Ubiquitous Engineering*, Springer, pages 429–435.