

ELiRF-UPV at SemEval-2018 Task 10: Capturing Discriminative Attributes with Knowledge Graphs and Wikipedia

José-Ángel González, Lluís-F. Hurtado, Encarna Segarra, Ferran Pla

Universitat Politècnica de València

Camí de Vera sn, 46022, València

{jogonba2|lhurtado|esegarra|fpla}@dsic.upv.es

Abstract

This paper describes the participation of ELiRF-UPV team at task 10, Capturing Discriminative Attributes, of SemEval-2018. Our best approach consists of using ConceptNet, Wikipedia and NumberBatch embeddings in order to establish relationships between concepts and attributes. Furthermore, this system achieves competitive results in the official evaluation.

1 Introduction

Capturing Discriminative Attributes, task 10 of SemEval-2018 (Krebs et al., 2018), proposes working on semantic difference detection. The goal of the task is to predict whether a word is a discriminative attribute between two other words. This problem is known as semantic difference detection, which is a binary classification task: given a triple (*apple, banana, red*), it consists in determining whether it exemplifies a semantic difference. Regarding semantic difference, it is a ternary relation between two concepts, for instance, (*apple, banana*) and a discriminative feature (*red*) that characterizes the first concept but not the other.

As task 10 is related to the semantic relations among different words, knowledge graphs seems the most appropriate resources to be used. An interesting knowledge resource that we used for this task is ConceptNet. In particular, ConceptNet (Speer et al., 2016) is a knowledge graph that connects words and phrases of natural language using labeled edges. It was designed to represent some general knowledge involved in natural language and could be used in combination with other resources. The combination of ConceptNet with distributed representations such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014), is known as NumberBatch

embeddings (Speer et al., 2016).

Regarding the relations specified in ConceptNet, there are a total of 36 relations such as *IsA* (A banana is a dessert), *UsedFor* (A net is used for catching fish), or *FormOf* (“Leaves” is a form of the word “leaf”), intended to represent a relationship independently of the language or the source of the terms it connects.

In this work, we propose five knowledge based systems and one additional machine learning system based on Siamese networks. We used ConceptNet in order to determine if each input concept and the input attribute are related through a relation edge or path. When there is a relationship between the first concept and the attribute and there is no relationship between the second concept and the attribute, then the answer is 1, otherwise, the answer is 0. However, there are cases in which ConceptNet does not provide enough information to take a decision. In those cases, we have implemented a system that seeks the information in Wikipedia articles by using distances between NumberBatch embeddings.

2 Resources and Preprocess

As we stated in Section 1, ConceptNet 5 is used in order to find the relationships among concepts and attributes. ConceptNet 5 is freely available under the Creative Commons Attribution-ShareAlike license (CC BY SA 4.0) from <http://conceptnet.io>. Moreover, we use two additional resources next to ConceptNet.

On the one hand, in order to use more information of each concept we used the Wikipedia articles. In this way, we get the most related article for each concept, following the recommendation of the Wikipedia disambiguation system. Wikipedia articles had been preprocessed. First, we remove non relevant sections such as “See

Also”, “References” and “External Links” which links to other resources. After that, we normalized tokens like numbers or urls e.g. “678.2” → “number” y “https://en.wikipedia.org” → “url” and we made a tokenization of the articles.

On the other hand, we used distributed representations of words, more specifically, we used NumberBatch embeddings (Speer et al., 2016).

3 System Description

We tested several approaches to address this task, mostly knowledge-based. Let (c_1, c_2, at) be a triple where, c_1 and c_2 are concepts and at is an attribute. The goal of the task is to define a function d to decide whether at is a discriminative feature of c_1 (value 1 for d) or not (value 0 for d). That is, if at characterizes c_1 but not c_2 .

Although the training and the development sets were not very large, we wanted to test the performance of Machine Learning (ML) approaches for this task. We selected a siamese neural network (Bromley et al., 1993) because these kind of systems are suitable for similar tasks such as knowledge base completion (Yang et al., 2014). This system works as follow. First, the input to this network are the NumberBatch embeddings of c_1 , c_2 and at . From this, a shared Multilayer Perceptron is applied in order to extract a complex representation of each term, $f(c_1)$, $f(c_2)$ and $f(at)$. With these representations, we compute the differences s_1 and s_2 where $s_1 = f(c_1) - f(at)$ and $s_2 = f(c_2) - f(at)$ with the aim of establishing relationships between each concept and the attribute. Finally, we concatenate s_1 and s_2 and we apply a fully-connected layer with softmax activation functions to carry out a classification i.e. $d_{Sia} = 1$ if at is discriminative for c_1 and not for c_2 or $d_{Sia} = 0$ otherwise.

The rest of the systems were knowledge-based. As first knowledge-based approach, we use the relationships between each concept and the attribute to determine if the attribute is discriminant. To do this, we use the ConceptNet relations. Note that, ConceptNet contains both positive (*IsA*, *FormOf*, *DerivedFrom*, *SimilarTo*, ...) and negative relations (*DistinctFrom*, *NotHasProperty*). In our proposals, we only consider positive edges, that is, those that denote positive relationships.

We look for positive edges between each concept and the attribute. If an edge between c_1 and at exists but there is not any edge between c_2 and

at , we assume that at is discriminant. In other words, at is discriminant if it is reachable from c_1 but not from c_2 . In this way, the function d_{CN} that determines if the attribute at is discriminant for concepts c_1 and c_2 can be defined as shown in Equation 1.

$$d_{CN} = \begin{cases} 1 & : \text{ if } at \in R(c_1) \wedge at \notin R(c_2) \\ 0 & : \text{ otherwise} \end{cases} \quad (1)$$

where, $R(c_1)$ and $R(c_2)$ are the sets of reachable nodes from c_1 and c_2 respectively using positive edges.

The main problem of this proposal is its low coverage. In many cases, there is no edge between any of the two concepts and the attribute, and therefore, it is decided that the attribute is not discriminant. In order to increase the coverage of d_{CN} , we extend the set of reachable nodes from a concept to those nodes reachable from other concepts closely related to the original concept. We have considered as related concepts those that are linked by the *FormOf* relation in ConceptNet. Considering the extended sets of reachable nodes, we can redefine the function d_{CN} of Equation 1 as shown in Equation 2.

$$d_{CN^2} = \begin{cases} 1 & : \text{ if } at \in R^2(c_1) \wedge at \notin R^2(c_2) \\ 0 & : \text{ otherwise} \end{cases} \quad (2)$$

where $R^2(c)$ is the set of nodes reachable from c or from any concept closely related to c .

Nevertheless, this approach still has coverage problems. In order to mitigate this problem, we proposed two new approaches based on Wikipedia. The main idea is simple, we search for the attribute at in the Wikipedia article of concepts c_1 and c_2 ($doc(c_1)$ and $doc(c_2)$) and, we decide if at is discriminant based on the result of this search. In this way, function d_{We} can be defined as shown in Equation 3.

$$d_{We} = \begin{cases} 1 & : \text{ if } at \in doc(c_1) \wedge at \notin doc(c_2) \\ 0 & : \text{ otherwise} \end{cases} \quad (3)$$

Further, we can relax the exact match criterion. Concretely, if we use NumberBatch distributed representations of words (h), we can compute similarities between at and all the tokens of $doc(c_1)$ and $doc(c_2)$ in order to decide which concept is the closest to at .

We define a threshold ϵ to ensure that there is enough difference between the maximum similarity $\max_{w \in doc(c_1)} \cos(h(w), h(at))$ and $\max_{w \in doc(c_2)} \cos(h(w), h(at))$. Using the development set, the value of ϵ was fixed to 0.2. This new decision function d_{Wt} is presented in the Equation 4.

$$d_{Wt} = \begin{cases} 1 & : \text{if } \max_{w \in doc(c_1)} \cos(h(w), h(at)) \\ & - \max_{w \in doc(c_2)} \cos(h(w), h(at)) \geq \epsilon \\ 0 & : \text{otherwise} \end{cases} \quad (4)$$

Finally, in order to explore the joint behavior of the knowledge-based approaches, we propose the combination of d_{CN^2} and d_{Wt} . When there is a relationship between at and any concept –it does not matter if it is c_1 , c_2 or both– we decide if at is discriminant using d_{CN^2} . But if there is no relationship in ConceptNet between them, we *smooth* the decision using d_{Wt} . Thus, we only use d_{CN^2} when we really have information in ConceptNet. The definition of this new decision function d_{CN^2+Wt} is shown in Equation 5.

$$d_{CN^2+Wt} = \begin{cases} d_{CN^2} & : \text{if } at \in R^2(c_1) \\ & \vee at \in R^2(c_2) \\ d_{Wt} & : \text{otherwise} \end{cases} \quad (5)$$

4 Experimental Results

In order to validate the correctness of the proposed approaches and also to select the one with the best performance for the competition, we carried out an evaluation of the approaches using the development set provided by the organizers. The results obtained are shown in Table 4.

d	Approach	Macro F_1
d_{Sia}	Siamese network	57.07
d_{CN}	ConceptNet	58.51
d_{CN^2}	ConceptNet ²	61.94
d_{We}	Wiki. exact match	58.46
d_{Wt}	Wiki. threshold	60.34
d_{CN^2+Wt}	ConceptNet + Wiki.	68.20

Table 1: Results on the development set.

As can be seen in Table 4, the knowledge-based systems which use knowledge resources obtain among 1.31 and 11.13 points of macro F_1 more

than the Siamese network which uses only NumberBatch embeddings. The approaches that use only ConceptNet (d_{CN} and d_{CN^2}) achieved as good results as those based on Wikipedia (d_{We} and d_{Wt}). Note that the coverage of d_{CN} and d_{CN^2} is very low, 48.71% for d_{CN} and 56.61% for d_{CN^2} . In cases where there are no links in ConceptNet –more than fifty percent of the time for d_{CN} – it is decided that the attribute is not discriminant. Moreover, the more knowledge incorporated into the system, the better results are obtained. We achieved the best results using the combination of ConceptNet graph and Wikipedia articles (d_{CN^2+Wt}), achieving 68.20 macro F_1 .

Regarding the evaluation with the test set, we used d_{CN^2+Wt} as decision function for the SemEval competition. Our system achieved competitive results (69.00 macro F_1 , 6 points of macro F_1 below the best system that obtains 75.00 macro F_1). Our proposal was ranked in 5th place out of a total of 26 participating teams. Several results from the official evaluation are shown in Table 4.

Team	Macro F_1
Sunnylnp (1/26)	75.00
Esantus (2/26)	73.00
ELiRF-UPV (5/26)	69.00
...	
Amrita_student (25/26)	49.00
Luminoso (26/26)	49.00

Table 2: Official results

5 Analysis of Results

Once the evaluation is finished, we want to carry out an analysis of the behavior of our system. Our goal is to detect in which types of attribute the system works worse. This way, we could add specific knowledge resources to deal with these attributes.

Although we have not completed this analysis, a group of attributes that caught our attention were the colors. While the overall error rate of our system was about 30%, the error in samples with attributes related to colors was about 50%. More details about the behavior of our system can be seen in Table 5.

Therefore, it would be possible to improve the system behavior by treating the color attributes in a specific way. For instance, by using image resources, such as ImageNet (Deng et al., 2009), to compute the color palette of images of each concept and compare it with the color attribute.

Attributes	Errors	Occurrences
Black	56	116
Brown	41	56
Yellow	11	46
Red	12	19
Blue	1	3
Color	121 (50.42%)	240
Other	586 (27.90%)	2100
Total	707	2340

Table 3: Error analysis in samples with attributes related to colors.

6 Conclusions and Future Work

In this work, we proposed a knowledge-based system for the discriminative attributes task. This system is based on the combination of two knowledge resources: a knowledge graph with semantic links such as ConceptNet and a general resource such as Wikipedia.

With this system, we achieved good results in the development set, compared to a supervised learning approach like siamese neural networks. That is, a combination of knowledge-based approaches produces significant improvements compared to the supervised approach. Regarding the evaluation with the test set, we obtained competitive results.

As future work, we propose an extension of our system based on the addition of more knowledge resources such DBpedia (Lehmann et al., 2015), Wordnet (Fellbaum, 1998) or Microsoft Concept Graph (Wang et al., 2015). Moreover, it could be interesting to consider the sections of Wikipedia with links to other resources in order to extract more information.

Finally, we propose the incorporation of knowledge resources into Deep Learning systems, beyond using only distributed representations of words. This offers us end-to-end systems capable of learning more complex decision algorithms. Concretely, the siamese neural networks seems to be promising for this work due to their good results in related fields such as knowledge-based completion (Yang et al., 2014).

7 Acknowledgements

This work has been partially supported by the Spanish MINECO and FEDER funds under projects ASLP-MULAN: Audio, Speech and Language Processing for Multimedia Analytics

(TIN2014-54288-C4-3-R); and AMIC: Affective Multimedia Analytics with Inclusive and Natural Communication (TIN2017-85854-C4-2-R). Work of José-Ángel González is also financed by Universitat Politècnica de València under grant PAID-01-17.

References

- Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1993. [Signature verification using a "siamese" time delay neural network](#). In *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS'93, pages 737–744, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. [ImageNet: A Large-Scale Hierarchical Image Database](#). In *CVPR09*.
- Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.
- Alicia Krebs, Alessandro Lenci, and Denis Paperno. 2018. [SemEval-2018 Task 10: Capturing discriminative attributes](#). In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.
- Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2015. [DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia](#). *Semantic Web Journal*, 6(2):167–195.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). *CoRR*, abs/1310.4546.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. [Glove: Global vectors for word representation](#). In *EMNLP*, volume 14, pages 1532–1543.
- Robert Speer, Joshua Chin, and Catherine Havasi. 2016. [Conceptnet 5.5: An open multilingual graph of general knowledge](#). *CoRR*, abs/1612.03975.
- Zhongyuan Wang, Haixun Wang, Ji-Rong Wen, and Yanghua Xiao. 2015. [An inference approach to basic level of categorization](#). In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, pages 653–662, New York, NY, USA. ACM.
- Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. [Embedding entities and relations for learning and inference in knowledge bases](#). *CoRR*, abs/1412.6575.