

# Instructions for EMNLP 2017 Proceedings

## Anonymous EMNLP submission

### 1 introduction

Personalization, as well as comparing human to animals, is a ubiquitous phenomenon in natural languages. In such cases, each kind of animal has its specific reasons to be mentioned. For example, the sentence “he is like a tiger” is more likely to refer to “he is strong”, rather than “he is sly”. The reason of such fact is we have already kept a prototype of tiger in our mind: all tigers are strong, but they are not necessary to be sly. We call properties like strong, ferocious and brave “central properties (or central attributes)” of tigers. To be specific, if the sentence “All *N(s)* are *A*.” is reasonable and common in a non-specific situation, we could treat *A* as a central property of *N* (there, *N* is a noun and *A* is an adjective).

The central properties of nouns can be applied to metaphor generation, making natural language generated by computer more vivid. It can be helpful to the collection of a metaphor knowledge base, in which each item could be stored as a triplet  $(n_{source}, n_{target}, adj_{reason})$ . As a example, most people agree that fat can be has a central property of pigs, but not central one of cats. Therefore, we could have (cat, pig, fat) as an item in the knowledge base, with which computer could generate the sentence like “the cat is like a pig” to express “the cat is fat”. Such generation can be interpreted easily.

More generally, not only animals have their central properties, the collection of our knowledge base could be expand into other domains rather than animals.

### 2 Related Work

Veale and Hao (2007) built an English knowledge base “sardonicus” with simile queries. Li et al. (2015) have built a Chinese one and indicated that would shed light to metaphor related tasks. Bulat

et al. have shown that attributes (or properties) are helpful to improve computational metaphor modeling.

For feature extraction, we used Linggle (Boisson et al., 2013). It provides an efficient API to rapidly query the frequency and collocations of words.

### 3 Experiments

#### 3.1 Feature Collection

To indicate whether an adjective is a central property of a noun, we apply the following features. We extracted the features below between 2,816 adjectives and 54 nouns.

##### 3.1.1 Frequency and Collocation

Word frequency and collocation are intuitive descriptors of a document. Linggle (Boisson et al., 2013) makes it easy for us to extract a word’s frequency and two words’ collocation frequency in web-scale corpus.

##### 3.1.2 PMI

Pairwise mutual information (or point mutual information), known as PMI (Church and Hanks, 1990; Bouma, 2009), is a widely accepted measure that illustrates the association between two words. The PMI between two words  $x$  and  $y$  can be computed by

$$PMI(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)} \quad (1)$$

where  $P(x)$  is the probability of word  $x$  to appear,  $P(x, y)$  is the joint probability of  $x$  and  $y$ , namely the probability of the collocation of  $x$  and  $y$ . Based on the statistics, it can be estimated by the frequency of word  $x$  and  $y$ , with the collocation frequency of them. In application, it is usually

computed by

$$\begin{aligned}
 PMI(x, y) &= \log_2 \frac{C(x, y) \cdot N}{C(x)C(y)} \\
 &= \log_2 C(x, y) - \log_2 C(x) - \log_2 C(y) + \log_2 N
 \end{aligned} \quad (2)$$

where  $N$  is the number of unigrams in total,  $C(x)$  and  $C(y)$  represent the frequency of unigram  $x$  and  $y$ , and  $C(x, y)$  represents the frequency of bigram  $(x, y)$ . With no loss of generality, we omit the constant  $\log_2 N$  and compute  $PMI^*(x, y) = \log_2 C(x, y) - \log_2 C(x) - \log_2 C(y)$  instead of  $PMI(x, y)$ .

### 3.1.3 Concreteness of words

Turney et al. (2011) proposed a method to compute the concreteness and abstractness of words, and suggested that concreteness of words are strong indicators of metaphors. We replicated the experiments for concreteness computation with the method on a large web scale corpus, and got a list with 75,626 words' concreteness. We also treat concreteness as a feature for central attributes extraction.

### 3.1.4 "As ... as" pattern

Previous works have shown that the simile template "as adjective as noun" can be applied to extract properties of nouns (Roncero et al., 2006; Veale and Hao, 2007; Li et al., 2015). Veale and Hao (2007) also indicated that simile is widely viewed as a less sophisticated conceptual device than metaphor, and simile detection would be helpful to extract metaphors. Thus, the frequency of cooccurrence of an adjective and a noun pair in the pattern of "as adj. as a/an noun." would be a strong feature to detect central properties of nouns. We used Linggle API (Boisson et al., 2013)<sup>1</sup> to extract such patterns, with the queries like "as strong as a/an n." and "as adj. as a/an tiger", and it would return 50 most frequent matched phrases.

## 3.2 Annotation

We annotated the extract

## 4 Future Work

- The concept of central property should not be limited in only adjectives. For example, "have sharp teeth" should be a central property of tigers or lions, so that something could

be compared to them in order to indicate such a property.

## References

- Joanne Boisson, Ting-Hui Kao, Jian-Cheng Wu, Tzu-Hsi Yen, and Jason S Chang. 2013. Linggle: a web-scale linguistic search engine for words in context. In *ACL (Conference System Demonstrations)*. pages 139–144.
- Gerlof Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the Biennial GSCL Conference*. volume 156.
- Luana Bulat, Stephen Clark, and Ekaterina Shutova. ????. Modelling metaphor with attribute-based semantics .
- Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics* 16(1):22–29.
- Bin Li, Xiaopeng Bai, Siqi Yin, and Jie Xu. 2015. Chinese cogbank: Where to see the cognitive features of chinese words. In *Proceedings of the Third Workshop on Metaphor in NLP*. pages 77–86.
- Carlos Roncero, John M Kennedy, and Ron Smyth. 2006. Similes on the internet have explanations. *Psychonomic Bulletin & Review* 13(1):74–77.
- Peter D Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 680–690.
- Tony Veale and Yanfen Hao. 2007. Learning to understand figurative language: from similes to metaphors to irony. In *Proceedings of the Cognitive Science Society*. volume 29.

<sup>1</sup>add the queries after the url <http://linggle.com/query/>