ANLP Seminar 1: Word Sense Disambiguation

Julie Weeds

January 9, 2019

Figure 1 shows part of the WordNet noun hypernym hierarchy. This is an **ISA** hierarchy where each concept in the tree **IS A** type of its parent. The parent concept is referred to as a hypernym (of the child) and the child concept is referred to as a hyponym (of the parent). Pedersen (2010) presents an empirical comparison of similarity measures for pairs of concepts in WordNet based on Information Content. Read the paper and answer the following questions.

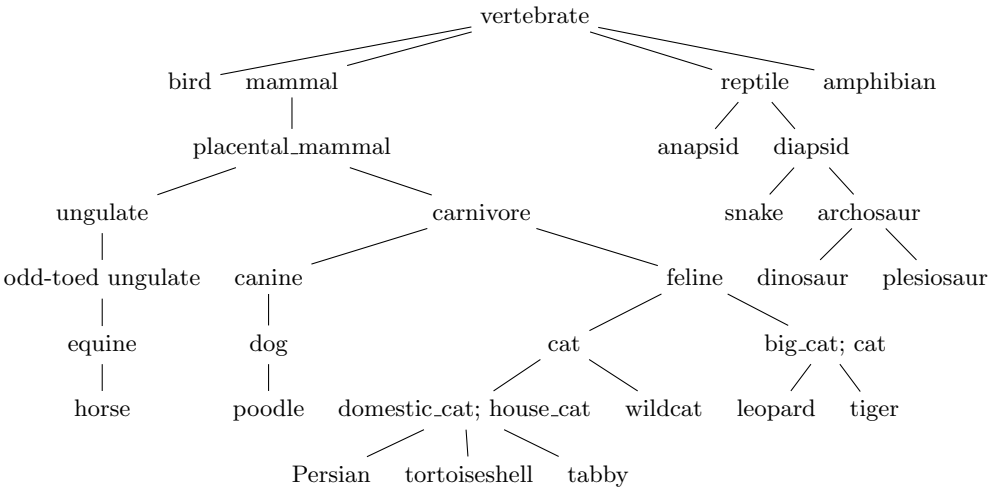

Figure 1: Part of the WordNet noun hypernym hierarchy

1. With reference to Figure 1, what concept is the hypernym of `ungulate`? How many hyponyms does `carnivore` have? Give an example. Why do you think the word `cat` appears twice in the hierarchy?

2. What do you understand by *path length*? Give some examples of pairs of words which have a *path length* of 2. What limitations can you think of in using *path length* as a measure of semantic similarity?

3. How is information content for a WordNet concept computed from a sense-tagged corpus? How can information content for a WordNet concept be estimated from untagged data?

4. What is the *lowest common subsumer (LCS)* of `dog` and `big_cat`? What is the LCS of `mammal` and `reptile`? What is the LCS of `poodle` and `tabby`? Which of these three pairs would have the greatest similarity according to the *res* measure? What about if you used the *lin* measure? Or a measure based on *path length*?

5. What is the main experimental conclusion of the paper? Are you convinced?

References

Ted Pedersen. 2010. Information content measures of semantic similarity perform better without sense-tagged text. In *Proceedings of NAACL*.