Free Paraphrases of Noun Compounds

SemEval-2013
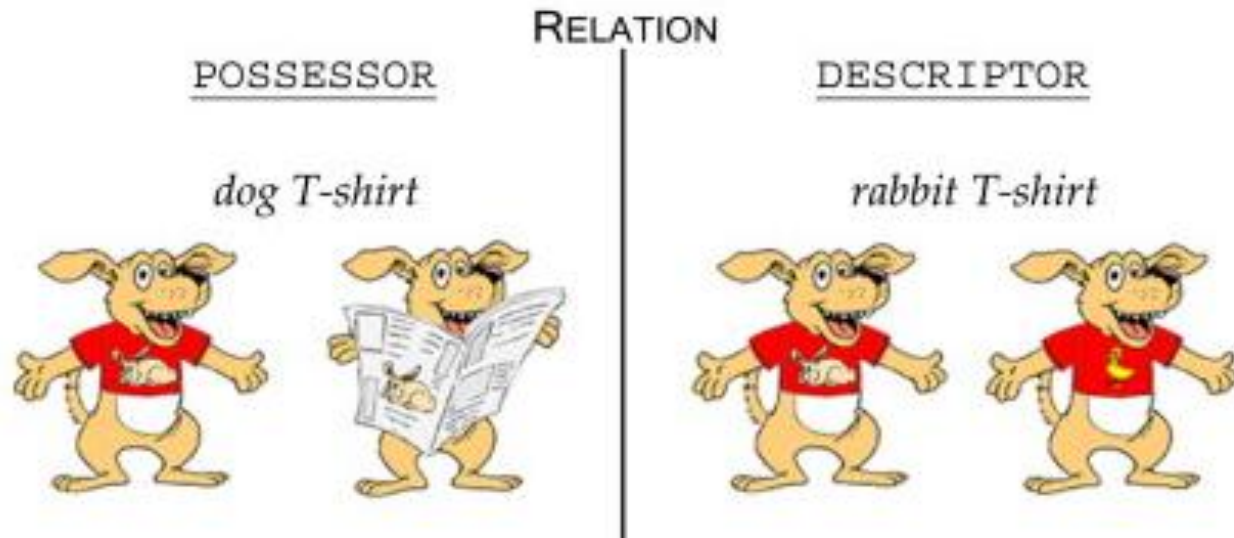
Vaibhav, Niraj Kumar
Advisor: Dr. Amitabha Mukerjee

CS365: Artificial Intelligence
Department of Computer Science and Engineering

IIT Kanpur, India

Introduction

- A noun compound is a noun that is made of two or more nouns.

- Our aim is to produce a list of paraphrases for a two word noun compound. (Semeval2013: Task 4)

- Example:

  malaria mosquito is a noun compound which can be paraphrased as mosquito that carries malaria, or that causes malaria.

Motivation

From [03.ibm.com]

- Noun compounds are a major part of any language.
- Correct interpretation of noun compound is important for semantic analysis of any text.
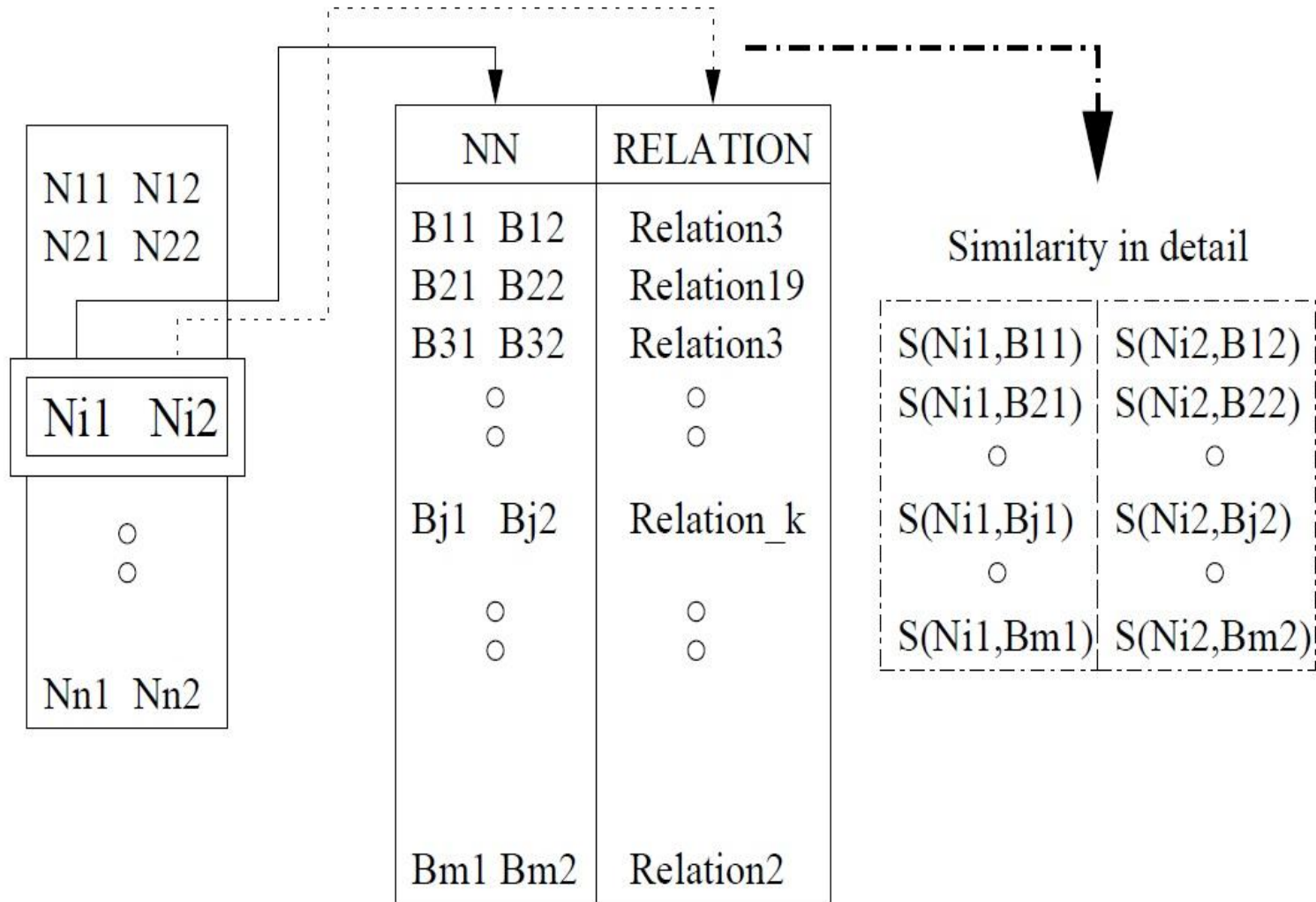
Past Works

From [ars.els-cdn.com]

- Previously, most of the work was done to interpret the noun compounds by classifying them into different relations.
- Girju et al. (2005) suggested 21 classes of abstract relations (top-down)
- Our aim is to generate paraphrases for any noun compound.
- Recently, Lilit Derbinyan (2010) used British National Corpus for verb phrase extraction followed by validation (bottom-up). (Very slow because of huge corpus size)

Algorithm

NN	RELATION
B11 B12	Relation3
B21 B22	Relation19
B31 B32	Relation3
○	○
○	○
Bj1 Bj2	Relation_k
○	○
○	○
Bm1 Bm2	Relation2

N11 N12
N21 N22

Ni1   Ni2

○
○

Nn1 Nn2

Similarity in detail

S(Ni1,B11)	S(Ni2,B12)
S(Ni1,B21)	S(Ni2,B22)
○	○
S(Ni1,Bj1)	S(Ni2,Bj2)
○	○
S(Ni1,Bm1)	S(Ni2,Bm2)

From [2]

Our Approach

Similarity:

The algorithm gives the most similar noun compound from the training set. This similarity calculation is done using WS4J, a Java API for WordNet Similarity.
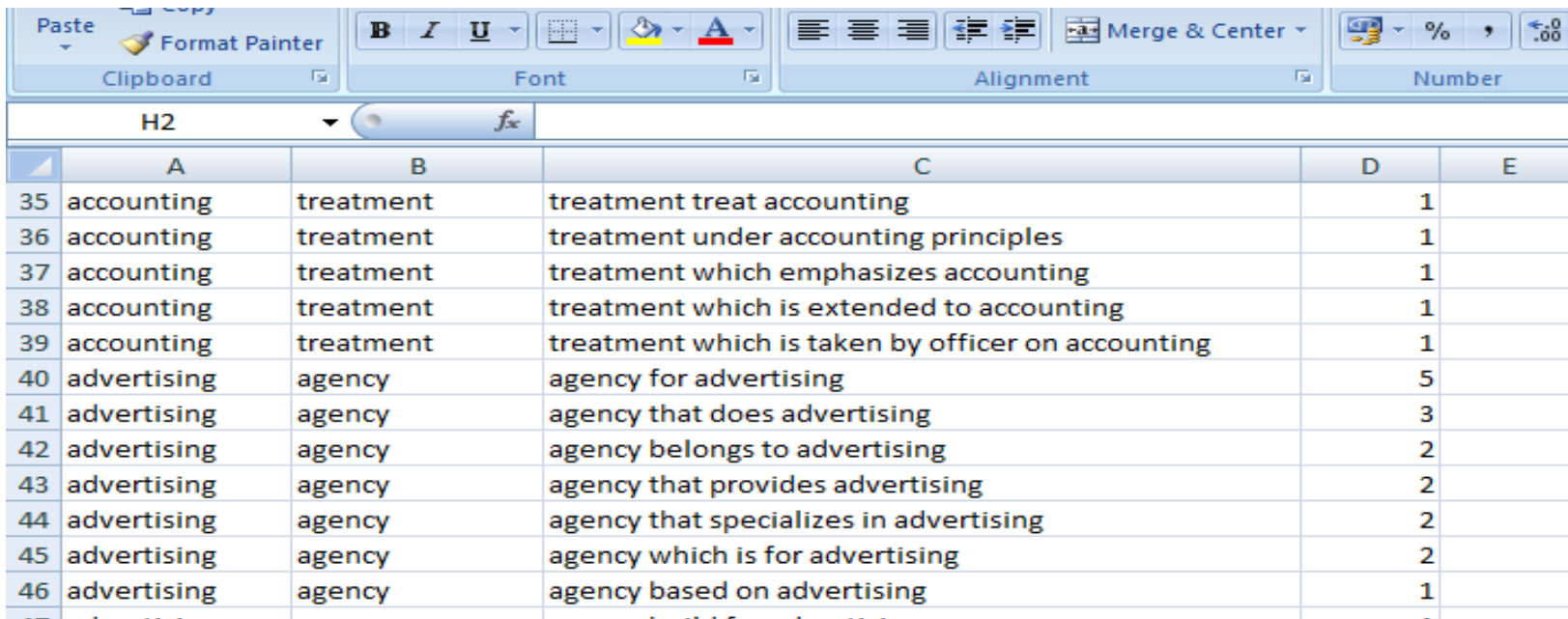
Validation:

Web based validation using Faroo API, which is a free web search API.

Datasets

- Semeval 2013: Task 4 - Training Dataset

- Semeval 2013: Task 4 - Test Input

- Fellbaum, C. (1998). WordNet: An electronic lexical database, MIT press Cambridge, MA.[A]

Results

- According to the scorer run on the output(for 58 test noun compounds)-
  - Isomorphic Score: 0.14
  - Non Isomorphic Score: 0.23

- Paraphrases for some of the test inputs-

application	software	application to make software series	172
application	software	application about software	118
application	software	application assigned in software class	73
application	software	application for software	70
application	software	application that creates software	63
application	software	application planned as a software	61
application	software	application for the software class	60
application	software	application related to software	60

business	unit	unit of business	159
business	unit	unit for the business	91
business	unit	unit based on business	78
business	unit	unit handling business	77
business	unit	unit that handles business	77

bike	ride	ride on bike	53
bike	ride	ride required for bike	53
bike	ride	ride in a bike	52
bike	ride	ride derived from the bike	51
bike	ride	ride coming from bike	49
bike	ride	ride from the bike	49
bike	ride	ride from bike	49

alarm	bell	bell designed for alarm purposes	98
alarm	bell	bell is used in alarm of its efficiency	75
alarm	bell	bell which is meant for alarm	64
alarm	bell	bell that lead to alarm	61
alarm	bell	bell utilized for alarm process	54
alarm	bell	bell used for alarm	53
alarm	bell	bell developed for alarm	51

Conclusion

We found that the similarity calculation based on the length of the **shortest path between two synsets** (Leacock & Chodorow, 1998) produces paraphrases which are most meaningful against other similarity algorithms.

For further accuracy, we can take into account the significance of semantic classes for calculating similar noun compounds.

Possible Improvements

Large corpus (like BNC) can be used to increase the accuracy of the paraphrases.

A powerful web search API (like Google, Yahoo web search) can be used for more precision.

References

- Butnariu, C., Kim, S., Nakov, P., Ó Séaghdha, D., Szpakowicz, S. & Veale, T. (2009). SemEval-2010 Task 9: *The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. In Proceedings of the SemEval-2010 Workshop.*[1]

- Su Nam Kim and Timothy Baldwin (2005). *Automatic Interpretation of Noun Compounds using wordNet similarity.*[2]

- Lilit Darbinyan supervised by Stephen Pulman (2010). *Interpretation of noun compounds*, Oxford University.[3]