10 – Automatic relation extraction

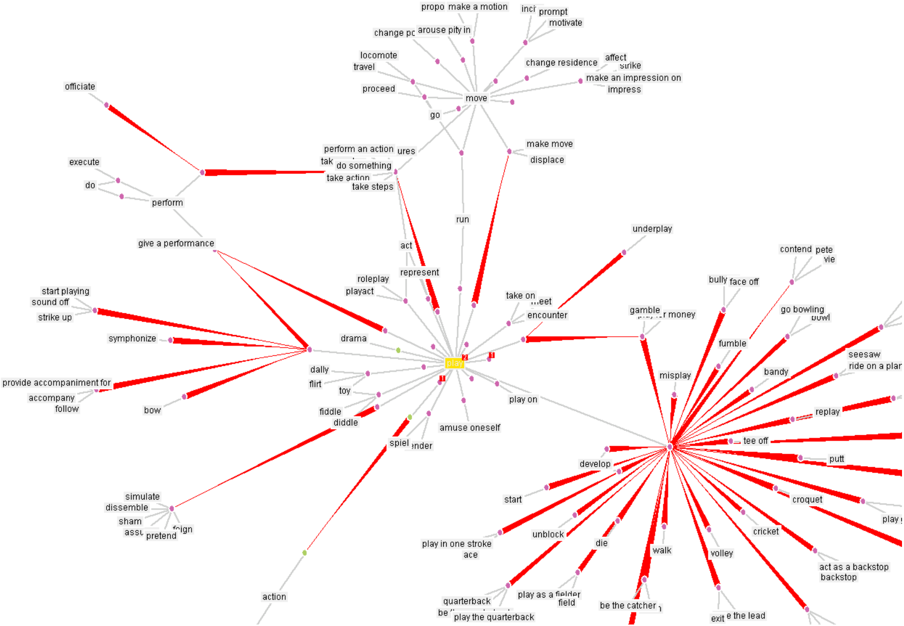
IA161 Advanced Techniques of Natural Language Processing

A. Rambousek

NLP Centre, FI MU, Brno

December 10, 2020

Automatic relation extraction



Semantic Networks

- network representing *relations between concepts*
- *knowledge graph*
- WordNet – lexical database of English
  - synsets, main relation hyponymy/hypernymy, meronymy, synonymy, antonymy...
  - Multilingual Wordnet network

Why would you do that?

- semantic analysis (house $\rightarrow$ home, music, MD?)
- query expansion (dog $\rightarrow$ poodle, terrier. . .)
- lexical substitution (match $\rightarrow$ game)
- machine translation
- question answering
- domain classification (lemon, apple, banana $\rightarrow$ fruit)
- summarization
- paraphrase

Why would you do that?

- semantic analysis (house $\rightarrow$ home, music, MD?)
- query expansion (dog $\rightarrow$ poodle, terrier...)
- lexical substitution (match $\rightarrow$ game)
- machine translation
- question answering
- domain classification (lemon, apple, banana $\rightarrow$ fruit)
- summarization
- paraphrase

Example

Human illuminates Document
AG[bird:1] VERB sezobnout SUBS[feed:1]

What do we need?

- morphological tags
- syntactic analysis (phrases)
- dataset (dictionary, corpus, Wikipedia...)

Pattern recognition

regular expression to match Part-of-Speech and text

Pattern recognition

regular expression to match Part-of-Speech and text

Example

NP {,} especially {NP, }* {or |and} NP

Pattern recognition

regular expression to match Part-of-Speech and text

Example

NP {,} especially {NP, }* {or |and} NP

...most *European countries*, especially *France*, *England*, and *Spain*.

European country >France

European country >England

European country >Spain

Pattern recognition

regular expression to match Part-of-Speech and text

Example

NP {,} especially {NP, }* {or |and} NP

...most *European countries*, especially *France*, *England*, and *Spain*.

European country >France

European country >England

European country >Spain

Example

e.g. {NP,}* {and |or} NP.

Pattern recognition

regular expression to match Part-of-Speech and text

Example

NP {,} especially {NP, }* {or |and} NP

...most *European countries*, especially *France*, *England*, and *Spain*.

European country >France

European country >England

European country >Spain

Example

e.g. {NP,}* {and |or} NP.

...e.g. apples, bananas, or pears.

related terms

Example

NP such as {NP, }* {and |or} NP

Example

NP such as {NP, }* {and |or} NP

common *domestic animals* such as the *ferret* and the *fancy rat*

domestic animal >ferret

domestic animal >(fancy) rat

Example

NP such as {NP, }* {and |or} NP

common *domestic animals* such as the *ferret* and the *fancy rat*

domestic animal >ferret

domestic animal >(fancy) rat

in areas with a long history of *mining* such as *South-west England*

mining >South-west England

Example

NP such as {NP, }* {and |or} NP

common *domestic animals* such as the *ferret* and the *fancy rat*

domestic animal >ferret

domestic animal >(fancy) rat

in areas with a long history of *mining* such as *South-west England*

mining >South-west England

in *areas* (with a long history of mining) such as *South-west England*

area >South-west England

- remove stopwords
- detect optional adjunct phrases
- detect named entities

No.	Pattern	Number of occurrences	Number of relevant occurrences	Intermediary precision (%)
1.	other than	168	164	97.6
2.	especially	120	90	75
3.	principally	11	6	54.5
4.	usually	18	14	77.8
5.	such as	2470	1950	78.9
6.	in particular	78	48	61.5
7.	e(.)g(.)	280	216	77.1
8.	become	780	510	66.7
9.	another	92	72	78.3
10.	notably	76	42	55.3
11.	particularly	130	80	61.5
12.	except	13	4	30.8
13.	called	270	220	81.5
14.	like	1600	1300	81.3
15.	including	670	430	64.2

Corpus query

- special case of pattern recognition, CQL query
- bigger data at hand, less options

Corpus query

- special case of pattern recognition, CQL query
- bigger data at hand, less options

Example

je/jsou

```
2:[k="k1"&c="c1"] ([lc=","] [k="k1"])*
([lc="a"|lc="i"|lc="nebo"|lc="či"] [k="k1"])?
[lemma_lc="být"&tag="k5eAaImIp3.*"&lc!="ne.*"]
([k="k1"&c="c[1246]"] [k="k2"]{0,2})?
1:[k="k1"&c="c[1246]"]
```

experiment on domain dictionary: precision 40 %, when limited to dictionary terms 52 %

Multilingual translation

using translation equivalents from multilingual dictionary to provide synonyms

Multilingual translation

using translation equivalents from multilingual dictionary to provide synonyms

Example

stůl = table
table = stůl, stolek
stůl = stolek

Synonym transitivity

- expanding relations based on existing relations (transitive closure)

Example

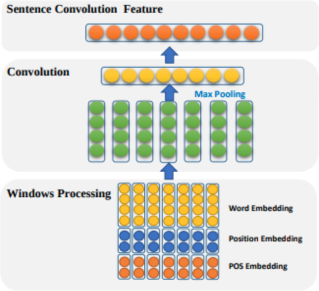
city = town, town = municipality
$\Rightarrow$ city = municipality

Distributional approach

- vector space model
- word-context frequency matrix
- clustering
- similar context $\neq$ synonym
- e.g. Sketch Engine thesaurus

Neural networks

- word embeddings
- position embeddings – relative distance between words
- part of speech embeddings – tag PoS for each word
- WordNet information may help
- combine properties to get relations between entities in sentence



Relation	Representation of Word Attention Weight
Instrument-Agency	The author of a keygen uses a disassembler to look at the raw assembly code
Message-Topic	The Pulitzer Committee issues an official citation explaining the reasons for the award
Cause-Effect	The burst has been caused by water hammer pressure
Instrument-Agency	Even commercial networks have moved into high-definition broadcast
Component-Whole	The girl showed a photo of apple tree blossom on a fruit tree in the Central Valley
Member-Collection	They tried an assault of their own an hour later, with two columns of sixteen tanks backed by a battalion of Panzer grenadiers

TOEFL test evaluation

- evaluation by solving TOEFL synonym test
- Choose synonym for *fabricate*.
  - construct, alter, select, demonstrate
- build synonym set for each word
- detect overlap
- success rate 88 %

SemEval

- various tasks evaluating computational semantic analysis systems
- human annotators provide *gold standards*
- NLP systems are evaluated
- tasks include Word Sense Disambiguation, Machine Translation, Information Extraction, Learning Semantic Relations...

References I

Barbu, V. (2008).
Hyponymy patterns: Semi-automatic extraction, evaluation and inter-lingual comparison.
In *Text, Speech and Dialogue*, pages 37–44.

Grefenstette, G. (2015).
Inriasac: Simple hypernym extraction methods.
arXiv preprint arXiv:1502.01271.

Hearst, M. A. (1998).
Automated discovery of wordnet relations.
WordNet: an electronic lexical database, pages 131–153.

References II

Lefever, E., Van de Kauter, M., and Hoste, V. (2014).
Evaluation of automatic hypernym extraction from technical corpora in english and dutch.
In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, pages 490–497.

Li, Q., Li, L., Wang, W., Li, Q., and Zhong, J. (2020).
A comprehensive exploration of semantic relation extraction via pre-trained cnns.
Knowledge-Based Systems, page 105488.

Sang, E. T. K. and Hofmann, K. (2009).
Lexical patterns or dependency patterns: which is better for hypernym extraction?
In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 174–182. Association for Computational Linguistics.

References III

Schropp, G., Lefever, E., and Hoste, V. (2013).
A combined pattern-based and distributional approach for automatic hypernym detection in dutch.
In *RANLP*, pages 593–600.

Shen, Y. and Huang, X.-J. (2016).
Attention-based convolutional neural network for semantic relation extraction.
In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2526–2536.

Wang, T. and Hirst, G. (2012).
Exploring patterns in dictionary definitions for synonym extraction.
Natural Language Engineering, 18(03):313–342.