Entity Linking - Part II

Pawan Goyal

CSE, IIT Kharagpur

Week 10, Lecture 2

Keyphraseness and Commonness: Always the best decision?

Depth-first search

From Wikipedia, the free encyclopedia

Depth-first search (**DFS**) is an algorithm for traversing or searching a tree, tree structure, or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

Formally, DFS is an uninformed search that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hadn't finished exploring. In a non-recursive implementation, all freshly expanded nodes are added to a LIFO stack for exploration.

sense	commonness	relatedness
Tree	92.82%	15.97%
Tree (graph theory)	2.94%	59.91%
Tree (data structure)	**2.57%**	**63.26%**
Tree (set theory)	0.15%	34.04%
Phylogenetic tree	0.07%	20.33%
Christmas tree	0.07%	0.0%
Binary tree	0.04%	62.43%
Family tree	0.04%	16.31%
...		

Keyphraseness and Commonness: Always the best decision?

Depth-first search

From Wikipedia, the free encyclopedia

Depth-first search (**DFS**) is an algorithm for traversing or searching a tree, tree structure or graph. One starts at the root (selecting some node as the root in the graph case) and explores as far as possible along each branch before backtracking.

Formally, DFS is an uninformed search that progresses by expanding the first child node of the search tree that appears and thus going deeper and deeper until a goal node is found, or until it hits a node that has no children. Then the search backtracks, returning to the most recent node it hadn't finished exploring. In a non-recursive implementation, all freshly expanded nodes are added to a LIFO stack for exploration.

sense	commonness	relatedness
Tree	92.82%	15.97%
Tree (graph theory)	2.94%	59.91%
Tree (data structure)	**2.57%**	**63.26%**
Tree (set theory)	0.15%	34.04%
Phylogenetic tree	0.07%	20.33%
Christmas tree	0.07%	0.0%
Binary tree	0.04%	62.43%
Family tree	0.04%	16.31%
...		

Using Relatedness: Basic Idea

- In a sufficiently long text, one finds terms that do not require disambiguation at all.
- Use every unambiguous link in the document as context to disambiguate ambiguous ones.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.
- Comparison of articles is facilitated by the Wikipedia Link-based measure, which measures the semantic similarity of two Wikipedia pages by comparing their incoming and outgoing links.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.
- Comparison of articles is facilitated by the Wikipedia Link-based measure, which measures the semantic similarity of two Wikipedia pages by comparing their incoming and outgoing links.
- The relatedness of a candidate sense is the weighted average of its relatedness to each context article.

Computing Relatedness

- Each candidate sense and context term is represented by a single Wikipedia article.
- Thus the problem is reduced to selecting the sense article that has most in common with all of the context articles.
- Comparison of articles is facilitated by the Wikipedia Link-based measure, which measures the semantic similarity of two Wikipedia pages by comparing their incoming and outgoing links.
- The relatedness of a candidate sense is the weighted average of its relatedness to each context article.

How to give different weights to the context terms?

Weighting the Context Terms

Weighting the Context Terms

- **link probability:** Use the ones that are almost always used as a link within the articles where they are found, and always link to the same destination

Weighting the Context Terms

- **link probability:** Use the ones that are almost always used as a link within the articles where they are found, and always link to the same destination
- **relatedness:** We can determine how closely a term relates to the central document by calculating its average semantic relatedness to all other context terms

Weighting the Context Terms

- **link probability:** Use the ones that are almost always used as a link within the articles where they are found, and always link to the same destination
- **relatedness:** We can determine how closely a term relates to the central document by calculating its average semantic relatedness to all other context terms

These two variables - link probability and relatedness - are averaged to provide a weight for each context.

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold.

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*
- All the remaining phrases are disambiguated using the approach mentioned earlier.

Can we improve mention detection with this approach?

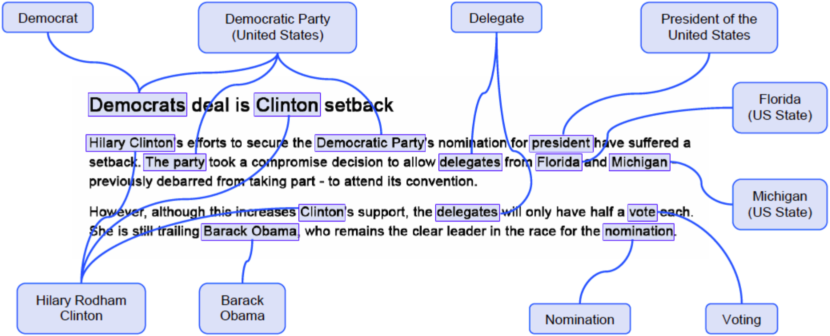
- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*
- All the remaining phrases are disambiguated using the approach mentioned earlier.
- This results in a set of associations between terms in the document and the Wikipedia articles that describe them.

Can we improve mention detection with this approach?

- The link detection process starts by gathering all n-grams in the document, and retaining those whose probability exceeds a very low threshold. *Is it the best method?*
- All the remaining phrases are disambiguated using the approach mentioned earlier.
- This results in a set of associations between terms in the document and the Wikipedia articles that describe them.

Can you use this to learn – which concepts should be linked?

Example


Democrat
Democratic Party (United States)
Delegate
President of the United States
Florida (US State)
Michigan (US State)
Democrats deal is Clinton setback
Hilary Clinton's efforts to secure the Democratic Party's nomination for president have suffered a setback. The party took a compromise decision to allow delegates from Florida and Michigan previously debarred from taking part - to attend its convention.
However, although this increases Clinton's support, the delegates will only have half a vote each. She is still trailing Barack Obama, who remains the clear leader in the race for the nomination.
Hilary Rodham Clinton
Barack Obama
Nomination
Voting

The Learning Problem: Which topics should be linked?

- The automatically identified Wikipedia articles provide training instances for a classifier.

The Learning Problem: Which topics should be linked?

- The automatically identified Wikipedia articles provide training instances for a classifier.
- Positive examples are the articles that were manually linked to, while negative ones are those that were not.

The Learning Problem: Which topics should be linked?

- The automatically identified Wikipedia articles provide training instances for a classifier.
- Positive examples are the articles that were manually linked to, while negative ones are those that were not.
- Features of these articles – and the places where they were mentioned – are used to inform the classifier about which topics should and should not be linked.

What are the features?

- **Link Probability:** Average as well as maximum of link probability of the link locations – (e.g. Hillary Clinton and Clinton)
- **Relatedness:** Topics which relate to the central thread of the document are more likely to be linked
- **Disambiguation Confidence:** The confidence score of the classifier for disambiguation
- **Generality:** Defined as the minimum depth at which it is located in Wikipedia's category tree. More useful for the readers to provide links for specific topics.
- **Location and Spread:** Where are these mentioned? First occurrence, last occurrence and the spread.

References

- Mihalcea, Rada, and Andras Csomai. “Wikify!: linking documents to encyclopedic knowledge.” Proceedings of the sixteenth ACM conference on information and knowledge management. ACM, 2007.
- Milne, David, and Ian H. Witten. “Learning to link with wikipedia.” Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008.