Skip N-grams and Ranking Functions for Predicting Script Events

Bram Jans   **Steven Bethard**   **Ivan Vulic**   **Marie Francine Moens**

SUMMARY

Aim:
Improving the following:
- Identification of events from narrative script
- Gathering statistics from the event chains
- Choose ranking functions for predicting script events

Corpus Used: Reuters Corpus, Volume 1
                Andrew Lang Fairy Tale Corpus

Procedure:
Model takes as input a partial script and produces as output a ranked list of events for that script.
I.   They **filtered out the non-narrative articles** out of the corpuses.
II.   **Identifying event chains from the script**
        As in the previous works of Chambers and Jurafsky in 2008 and 2009, and in Mcintyre and Lapata in 2009 and 2010, the following are used:
            Stanford Parser for identifying the dependency structure.
            OpenNLP coreference engine for identifying entities in each article.
        Finally event chains were constructed using all noun phrases associated with that entity and any subject/object dependencies with a verb were retrieved.
        Either <u>all, long or the longest</u> event chains can be used for training a model.
III.   **Gathering event chain statistics**
        <u>Regular bigrams, 1-skip bigrams and 2-skip bigrams</u> are three strategies for collecting bigrams. On these bigrams event pair counts and various conditional probability calculations are done to gather statistics.
IV.   **Ranking methods used**
        <u>Pointwise Mutual Information (PMI)</u> used by Chambers and Jurafsky
        <u>Ordered PMI</u>
        <u>Bigram probabilities</u> of language modelling
V.   **Evaluation Metrics**
        Given an event chain of n events, n cloze tests are run, each time removing one of the n events to form a partial script.
        Given a partial script as input, accurate event prediction will be one that ranks the missing event highly in its output guess list.
        Average rank and Recall@N are two such metrics. Better results are with lower average ranks and higher Recall@N values. N=50 is used in this research.

Results:

On Reuters corpus: all chains, 2-skip bigrams and bigram probabilities gave the best results.

On Fairy Tale corpus: long chains performed the best.
1-skip and 2-skip bigrams performed equally well.
Mixed results for bigram probabilities and PMI.

Main contributions:

- Used **Skip-grams** for the first time and showed that it is better than n-grams approach used previously.
- Proposed **bigram probabilities** method for ranking events in a given partial script.
- A new evaluation procedure **Recall@N** is also introduced.

References:

[1] Bram Jans, Steven Bethard, Ivan Vulić, and Marie-Francine Moens. Skip N-grams and Ranking Functions for Predicting Script Events. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012.
[2] http://en.wikipedia.org/wiki/Dependency_grammar Dependency Parsing
[3] http://en.wikipedia.org/wiki/Coreference Coreference resolution