Duluth at SemEval-2017 Task 6: Language Models in Humor Detection

Xinru Yan & Ted Pedersen
{yanxx418,tpederse}@d.umn.edu
https://xinru1414.github.io/HumorDetection-SemEval2017-Task6/
Department of Computer Science University of Minnesota Duluth

Introduction

SemEval-2017 Task 6 *#HashtagWars: Learning a Sense of Humor* aims to characterize humor from tweets submitted to a game show *@midnight* [1]. Duluth system completed the task using **Ngram Language Models (LMs)** [2].

Language Models

- **Ngram models**: predict the upcoming word from the previous N-1 words.
- **Markov** assumption: the probability (PR) of a word depends only on a small number of previous words. For trigrams:

$$P(w_n|w_1^{n-1}) \approx P(w_n|w_{n-2}, w_{n-1}) \quad (1)$$

- **Trigram LM**: use trigrams to compute the PR of a sequence of words:

$$P(w_1^n) \approx \prod_{k=1}^{n} P(w_k|w_{k-2}, w_{k-1}) \quad (2)$$

- We train LMs to assess the **similarity** of a tweet comparing to funny tweets, or **distinctiveness** of a tweet comparing to the **common language** (English news) to detect how humorous it is [3].

UMD
University of Minnesota Duluth
Driven to Discover

The Task

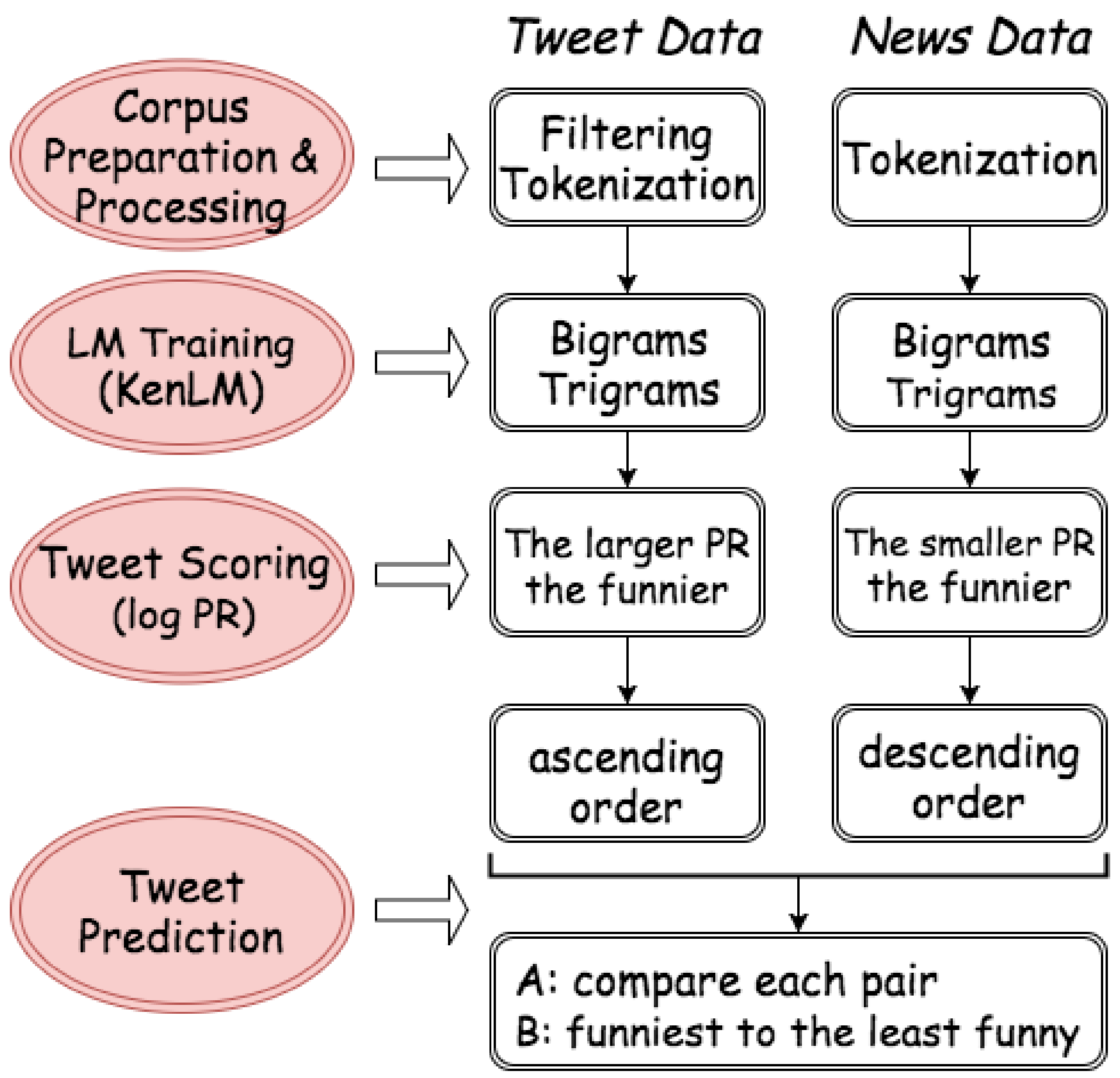
Tweets are in three baskets: top most funny tweet, next nine funny tweets and all remaining.

- Subtask A: Pairwise Comparison – a system should predict which tweet is funnier for every possible combination of tweet pairs from a given hashtag file.
- Subtask B: Semi-Ranking – a system should produce a ranking of tweets from funniest to the least funny for a specific hashtag file.

Examples from #BreakUpIn5Words (Trigram LM, News Data)

Tweet	*@midnight*	Duluth
It's not you, it's meth.	funniest	funny
Hey, can we NOT talk?	funny	funny
You need your own Netflix	funny	not funny
Figured I'd try being happy.	not funny	funny
You're a Mac, I'm PC	not funny	not funny

Method


Tweet Data / *News Data*

Corpus Preparation & Processing ⇨ Filtering Tokenization / Tokenization

LM Training (KenLM) ⇨ Bigrams Trigrams / Bigrams Trigrams

Tweet Scoring (log PR) ⇨ The larger PR the funnier / The smaller PR the funnier

ascending order / descending order

Tweet Prediction ⇨ A: compare each pair B: funniest to the least funny

Dataset

- **Tweet Data**: provided by the task, 106 hashtag files, about 21,580 tokens.
- **News Data**: We used 6.2 GB English news, about 2 million tokens [1].

Results

Dataset	Ngram	Accuracy (A)	Distance (B)
news	**3**	**0.627** (4th)	**0.872** (1st)
news	2	0.624	0.853
tweet	**3**	**0.397** (8th)	**0.967** (8th)
tweet	2	0.406	0.944

Discussion & Future Work

- Duluth relied on bigram and trigram LMs since tweets are **short** and **concise**
- Bigram LMs performed slightly better than trigram LMs:
  –> *Unigram* and *character* level LMs
- The **type** and the **quantity** of the corpora is what really matters:
  –> *more tweet data, less news data*
- Duluth did extremely well on #BreakUpIn5Words by using trigram LM trained on news data and performed the worst on #RuinAChristmasMovie:
  - Language in #BreakUpIn5Words is the least similar to news compared to other hashtags thus represented better;
  - LMs do not have external knowledge such as movie titles.
- Traditional Ngram models do not account for long distance dependencies and creative use of language (OOV).
  –>*Deep learning method*:
  - LSTMs: long term dependencies
  - Character-based CNNs: unknown words

References

[1] Peter Potash, Alexey Romanov, and Anna Rumshisky.
SemEval-2017 Task 6: #HashtagWars: learning a sense of humor.
In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, BC, August 2017.

[2] Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H. Clark, and Philipp Koehn.
Scalable modified Kneser-Ney language model estimation.
In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 690–696, Sofia, Bulgaria, August 2013.

[3] Cristian Danescu-Niculescu-Mizil, Justin Cheng, and Lillian Kleinberg, Jonand Lee.
You had me at hello: How phrasing affects memorability.
In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 892–901. Association for Computational Linguistics, 2012.

[a] http://www.statmt.org/wmt11/featured-translation-task.html