

Analysis of Hillary Clinton Emails

Debojyoti Dey (15511264), Nimisha Agarwal (15511267)

November 9, 2016

Overview

- 1 What we proposed?
- 2 What have we achieved?
- 3 How we achieved them?
- 4 References

Problem Statement

- Extracting important topics from the emails.
- Relation of US with different nations in Hillary's perspective.
- Persons linked to Clinton and Hillary's sentiment about them.

Topic Modelling

- Technique for discovering the abstract “topics” in a set of documents.
- Topic of a document is inferred from the words in it.
- Latent Semantic Analysis(LSA) - uses SVD on term-doc matrix.
- Latent Dirichlet allocation(LDA) - Bayesian Inference
 - A topic has probabilities of generating certain words.
 - A document is a mixture of various topics.
- We have used LDA model in gensim for this purpose.

Sentiment Analysis

- Process of identifying and categorizing texts on sentiments.
- For determining the person's attitude towards a particular thing.
- VADER - Lexicon and rule based approach.
 - Each word has some +ve or -ve score.
 - Rules to score the combination of lexicons.
- We have used NLTK-VADER library for this purpose.

Achievements

- Fulfilled objectives
 - 1 Extracting important topics from the e-mails.
 - 2 Relation of US with different nations in Hillary's perspective.
- Difficult to assess sentiment of Hillary about some person from emails between them.

Demo

References

 <https://www.kaggle.com/kaggle/hillary-clinton-emails>

 Hutto, Clayton J and Gilbert, Eric (2014)

Vader: A parsimonious rule-based model for sentiment analysis of social media text
Eighth International AAAI Conference on Weblogs and Social Media.

 [Wikipedia](#)

Latent Dirichlet allocation — Wikipedia The Free Encyclopedia

 [Wikipedia](#)

Latent semantic analysis — Wikipedia The Free Encyclopedia

Questions?