USC CSCI-544 REPORT ON GROUP PROJECTS *

ThammeGowda Narayanaswamy
tnarayan@usc.edu
ID : 2074-6694-39

December 7, 2016

1 Text Summarization

1.1 What is the group's number?

47

1.2 What was the goal of the group's project?

To produce a coherent summary of text.

1.3 How did the group attempt to accomplish the goal?

The group used DUC 2001 dataset. The preprocessing steps followed are:

- Tokenization
- Case folding
- Stopwords removal
- stemming

There were 2 approaches for summarization:

1. Naive : The similarity score between the sentences was computed using Cosine Similarity measure and the overlap was determined.
2. Advanced: Graph based algorithm using page rank.

1.4 How did the group evaluate their work and what was the result?

The evaluation was done based on the precision, recall and F-scores. The results showed that the graph based approach has better performance than the naive approach.

2 Speaker detection in Sitcom (the big bang theory) corpus

2.1 What is the group's number?

15

2.2 What was the goal of the group's project?

To detect the speaker in dialog

*Produced for Fall 2016 session of CSCI 544, 'Applied NLP'

2.3 How did the group attempt to accomplish the goal?

They have used a dataset containing transcripts for 'The Big Bang Theory' Tv series. The features they experimented are :

- Bag of words
- Part of speech tags
- Receipts
- Topics
- Speaker Turns

They have use the following machine learning algorithms :

- CRF classifier from CRF++ suite
- Support vector machine Classifier
- Multi layer perceptron Classifier
- Maximum Entropy Classifier

2.4 How did the group evaluate their work and what was the result?

They have used the standard metrics such as precision, recall and F-score on an held out dataset. The top two performing algorithms were Max-entropy and Nave bayes with 33% and 31% respectively.

3 Comparison of Machine Learning classifiers in classification of Yelp reviews

3.1 What is the group's number?

3

3.2 What was the goal of the group's project?

To compare various machine learning classifiers for the task of classifying the yelp reviews. The classification they performed is a binary classification of sentiment analysis.

3.3 How did the group attempt to accomplish the goal?

They used the following features :

- Part of speech tokens
- Lemmatization
- Stemming

They have tried the following algorithms:

- Support Vector Machines
- Naive Bayes
- Decision Trees
- Perceptron
- RNN

3.4 How did the group evaluate their work and what was the result?

They have used the standard metrics such as precision, recall and F-score on an held out dataset. The top three performing algorithms were Perceptron, RNN and Nave bayes with precision and recall over 80%.

4 Book Genre Classification

4.1 What is the group's number?

8

4.2 What was the goal of the group's project?

To classify genre of the books.

4.3 How did the group attempt to accomplish the goal?

They have modeled this task as multi class classification task. They have tried the following algorithms:

- Maximum Entropy classifier
- Random forest classifier
- Support vector machines classifier

They have used TF-IDF for vector representation of the books.

4.4 How did the group evaluate their work and what was the result?

They have used standard metrics such as Precision, Recall and F-score for evaluating classifiers. Max Entropy and SVM Classifiers scored over 81% F-score.

5 Sentimental Analysis on Movie reviews.

5.1 What is the group's number?

36

5.2 What was the goal of the group's project?

To analyze sentiment of movie reviews.

5.3 How did the group attempt to accomplish the goal?

Sentiment analysis is a classification task. They tried to predict the target sentiment in a range between 0 and 4 where 0 being very negative and 4 being very positive. They have tried the following algorithms:

- Naive Bayes
- Random forest
- SVM

5.4 How did the group evaluate their work and what was the result?

They submitted their predictions to Kaggle. Kaggle had a standard way to evaluate the performance in terms of precision, recall and F-score. Their score was over 60%.

6 Restaurant Finder

6.1 What is the group's number?

28

6.2 What was the goal of the group's project?

To build restaurant finder bot with natural language interface.

6.3 How did the group attempt to accomplish the goal?

The group built an end to end system with the following components/modules:

- Speech Recogniser
- Natural Language Understanding
- Dialogue management
- Natural Language Generation
- Speech synthesis

6.4 How did the group evaluate their work and what was the result?

They evaluated the generated statements using two approaches:

- Key word matching
- Classification using Precision, recall and F-Score

The group had to perform manual evaluation for some parts of the system, the results were reasonably good.

7 Automated Essay Scoring System

7.1 What is the group's number?

57

7.2 What was the goal of the group's project?

To automatically assess and grade essay responses.

7.3 How did the group attempt to accomplish the goal?

Some of the preprocessing and feature transformations they have done are:

- Sentence parsing and word tokenization
- Add features for number of words and number of sentences.
- Added a feature for the ratio of word to sentence counts
- Added a feature for Spelling errors count
- Added part of speech tags.

This group have used simple bag of words and word2vec representation for mapping the text to vector space. The following machine learning algorithms were tried:

- Linear Regression
- Random Forest
- Support vector machines

7.4 How did the group evaluate their work and what was the result?

This group has compared their predictions with evaluation dataset on Kaggle. The cost function was mean squared error. The errors were:

Algorithm	Without Word2Vec	With Word2Vec
Linear Regression	41.97	13.39
SVM with RBF Kernel	77.71	45.66
Random Forest	47.86	23.13

Linear Regression with Word2Vec embedding performed best

8 Sarcasm Detection in Tweets

8.1 What is the group's number?

22

8.2 What was the goal of the group's project?

To detect sarcasm in tweets.

8.3 How did the group attempt to accomplish the goal?

The group used a twitter API client and collected data related to tags Sarcasm using related hashtags. They annotated tweets as YES/NO for the binary classification task. This group performed the following cleaning:

- Remove Hashtags
- Remove usernames and mentions
- Skip tweets with very little text
- Remove URLs in tweet

This group has tried the following algorithms:

- Perceptron
- Support Vector machines
- K Nearest Neighbors
- Max Entropy
- Naive Bayes

8.4 How did the group evaluate their work and what was the result?

They have used standard metrics such as Precision, Recall and F-score for evaluating the classifiers. The top two classifiers are MaxEntropy (accuracy:63%) and K Nearest Neighbors(accuracy: 61%).

9 Clause based Open Information Extraction.

9.1 What is the group's number?

7

9.2 What was the goal of the group's project?

To do open information extraction and the relationship between the clauses.

9.3 How did the group attempt to accomplish the goal?

They have used three different datasets for this project. Given a text sentence, first step is to compute the dependency graph for the sentence, then the second step is to determine set of clauses from the dependency graph, then finally to produce the set of coherent clauses from this set.

9.4 How did the group evaluate their work and what was the result?

This group did not have evaluation results at the time of group presentation, but they were working on it.

10 Headline Generation

10.1 What is the group's number?

25

10.2 What was the goal of the group's project?

To produce an headline or title for a passage of text.

10.3 How did the group attempt to accomplish the goal?

At the high level, their method had two steps, for each of those steps they had tried two techniques:

- Summarization
  - Using Text Rank
  - Using Lex Rank
- Summary to Headline
  - Hedge Trimmer
  - Keyword Extraction

10.4 How did the group evaluate their work and what was the result?

They have built a nice interface to collect the feedback from humans to evaluate the performance. The results were not great, however, considering the difficulty of this problem their system was doing the reasonable thing.