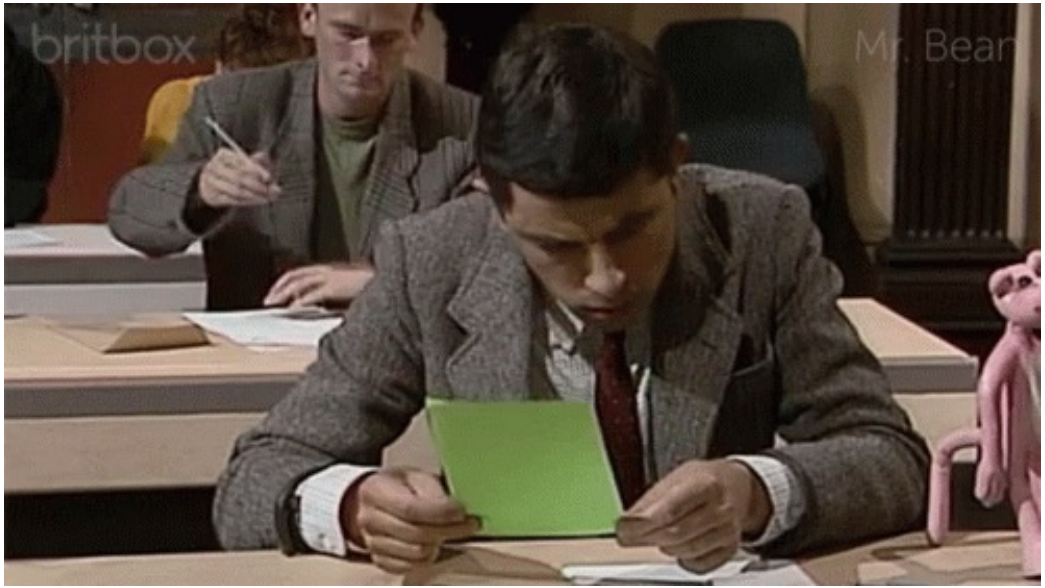
IR GROUP

Natural Language Processing

Technology
Arts Sciences
TH Köln

Dr. Narjes Nikzad
narjes.nikzad_khasmakhi@th-koeln.de

Bachelor Student Project



Bachelor Students' Project

Title: Conversational System with Sentiment Analysis

Objective: The objective of this project is to build a conversational system that can analyze the sentiment of user reviews. The system will use regular expressions for chatbot and logistic regression or Naive Bayes for sentiment analysis.

Bachelor Students' Project: Chatbot

Hi

Hi, how can I help you?

Analyze the sentiment of "it was tremendous movie"

Cool! Which texhnique would you like to apply?

Logistic Regression

The result is negative.

Project Detail

Key Steps:

- **Conversational System:** Design and implement a chatbot using regular expressions to recognize user intents.
- **Sentiment Analysis:** Train a logistic regression model to analyze the sentiment of user reviews.
- **Integration:** Integrate the sentiment analysis model into the conversational system to generate appropriate responses based on sentiment.

Project Detail

Regular expression

To establish the conversational part or chatbot section

Text Classification

To establish the sentiment analysis part

Sentiment Analysis

- Load dataset
- Create classifier
- Train your classifier
- Test on 50 samples (we provide in upcoming days)
- Evaluate using metrics

Conversational Part

- Create some pre-defined answers like ELIZA
- Use regular expression to find the user's query and provide the related answer

The submitted stuffs

- **Implementation Code**
- **Documentation Report**

- The last day of deadline we will download the submitted stuffs and it is **not possible to update** the content after the deadline.
- All files and linked should be included **before the deadline**.
- We **ignore submission after the deadline**.

The submitted stuffs

- **Implementation Code:**
  - Submit the **complete code implementation** of the conversational system and sentiment analysis.
  - **Include all relevant files** and **dependencies** necessary to run the code.
  - Ensure that the code is well-documented, organized, and easily understandable.

The submitted stuffs

- **Documentation Report :**
  - Prepare a comprehensive report documenting the project.
  - Submissions are limited to maximum 8 content pages.

The content of documentation report

- **Description of IMDB dataset**
  - Report some numbers like the number of reviews, the number of positive and negative ones…
  - Ploting the analysis can be considered as aditional features.
  - Preprocessing you performed
- **Description of the classifier process**
  - How did you train your classifier? (Write about the training process)
- **Evaluation**
  - Test your classfier with 50 samples we shared with you
  - Report the results of evaluation metrics on the test samples
  - What challenges did you ecnounter? How did you solve?

The content of documentation report

- **Regular expression**
  - Explain the procedure of building the conversational part (chatbot).
  - Discuss the regular expressions used.
  - What challenges did you face when you developed the conversational part .
- **Integration**
  - How did you integrate the sentiment analysis model into the conversational system?
  - How did you use the sentiment analysis results to generate appropriate responses or take relevant actions within the chatbot.
- **Write a discussion based on your findings**
  - Because of these reasons…. We could not find the pattern
  - Because of these reasons we could improve the performance

Add contribution table to your report

The regular expression	Training on IMDB dataset	Naive Bayes	Logistic Regression	Evaluation	Additional features (Creativity and Extensions)	Taks 5
Student 1	Student 1	Student2	Student 3	Student 1	Student 3	
Student 2				Student 2		
				Student 3		

An exmaple of how to show the contribution of memebers

Scoring items

- **Functionality** (**50**% of the score): Evaluate the functionality and correctness of their implementation, considering the following factors:
  1. Do the regular expressions effectively match the intended patterns in user input? (**20**%)
  2. Is the sentiment analysis model correctly integrated into the conversational system? (**20**%)
  3. Does the classifier correctly analyse the sentiment? (**10**%)

Scoring items

- **Evaluation of Sentiment Analysis Approach(es)** (**10**% of the score): Assess the performance of your sentiment analysis model(s).
  1. Consider how well the model(s) performs on the 50 samples provided by us and whether it demonstrates reasonable sentiment classification capabilities.

Scoring items

- **Creativity(Additional features)** (**25**% of the score): Evaluate the students' creativity and the complexity of their implementation. Consider factors such as:
  1. Did the students implement additional features or improvements?
  2. Did they handle edge cases and unexpected inputs effectively?
  3. Did they implement any additional techniques or optimizations to enhance the chatbot or sentiment analysis performance?

Scoring items

- **Report Structure and Presentation** (**15**% of the score):
  Reports have to be in English, your English skills will not be graded.
  1. The structure, organization, and clarity of the report document.
  2. Consider the coherence of ideas, adherence to the given format, and the quality of writing.
  3. Is there a contribution table (who did what)?

Creativity

- Add spell-checker
- Add recommendation facility to the chatbot
- Find a way to evaluate conversational part (maybe human evaluation or use current metrics)
- Error Analysis and Visualization
  - ✓ Perform an in-depth error analysis to understand the limitations and sources of errors.
  - ✓ Use visualizations, such as confusion matrices to illustrate the performance of the sentiment analysis model or the conversational system's response patterns.

Anything else?