

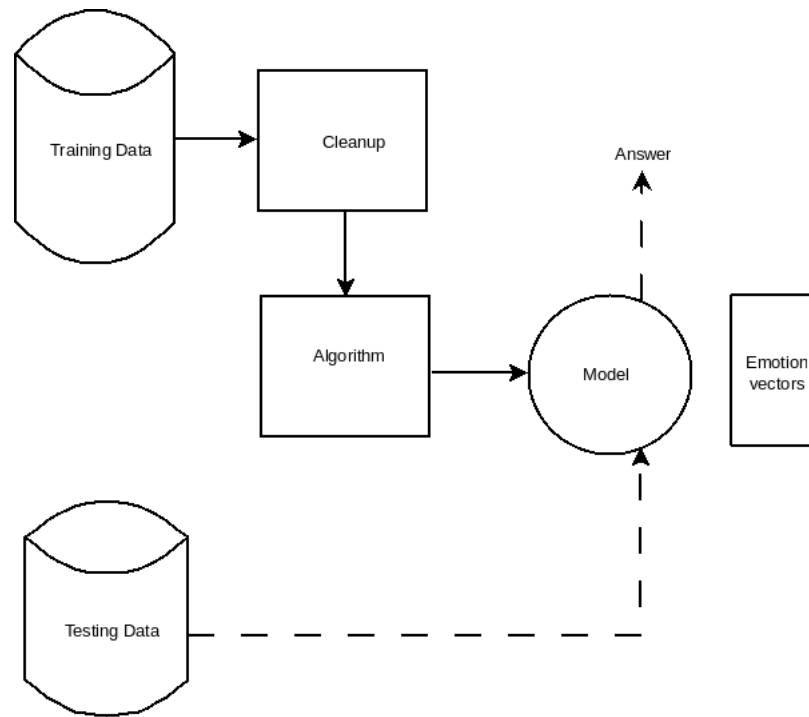
Emotion classification from text (chat)
Course Project, CS-725
Spring 2015-16

Ravi Shankar Mondal, 133059003
Saurabh Jain, 143050048
Srikrishna Khedkar, 143050055
Sushant Mahajan, 133059007

March 10, 2016

Feasibility

- We wish to create a system, which having been trained from sentences annotated with emotions portrayed, is capable of predicting the emotion for a new sentence.
- We will be closely following the paper Taner Danisman and Adil Alpkoçak titled **Feeler: Emotion Classification of Text Using Vector Space Model**.
- We will need to build a lexicon, which is basically an ordered set of all *good* words in the training data.
- For each sentence we will build a weight vector of all terms in the lexicon.
- The weights are calculated using the **tf-idf** technique.
- For each emotion class we take the mean of all vectors within it.
- We will be utilizing python for developing the application and be using **nltk** for text clean up, stemming etc.
- Estimated time to develop the project would be a month.



Dataset

- The paper mentions the ISEAR dataset, which is composed of around 15000 annotated sentences.

Scope

- At its core, the idea is to find median vectors corresponding to each emotion (joy, sadness, anger, ...) and then find the cosine similarity of the new vector composed from test input with the former. The highest similarity vector will be the winner.
- Other than implementing their proposed solution, we will also try to incorporate methods like bayesian classification and neural networks and analyse the results.
- If time permits, we will build some application around our system.