**What are the deliverables you plan to submit by the end of this project? Provide concrete items that are measurable, quantifiable, etc.**

We intend to submit one movie review sentiment analysis model, which could identify one certain sentence into three possible sentiments. The three emotional tendencies are positive(POS), middle(MID) and negative(NEG), where positive means reviewers recommend that movie, middle means reviewers have neutral emotion or no obvious sentiment and negative means reviewers have bad experience when watching that movie.

To evaluate the accuracy of the Sentiment model, we would like to use precision factor which comes from the following expression:

Precision = (#correctly recognised sentences)/(#all input sentences)

**Which tentative dataset do you use in your project? Show the basic statistics of the dataset.**

The dataset comes from the famous movie town — [movie.douban.com](http://movie.douban.com), which have million movie lovers from mainland China. The whole dataset contains 20 different types of movies that are classified by official Douban website. Each type include around 1000 popular movies, where one movie has at least 2000 reviews. The reviews are divided into three categories following the scored stars when reviewers made these reviews. We assume 4-5 stars is positive attitude, 3 stars is middle attitude and 1-2 stars is negative emotion. Now we have approximately 500 reviews in three classes of emotions to train our model and another 500 hottest reviews plus 100 newest reviews for testing. We developed our unique crawler to get all the reviews from website and build the special dataset ourselves.

The example file:

File name: The Shawshank Redemption (肖申克的救赎)

File content:

|  |  |  |  |
| --- | --- | --- | --- |
| 肖申克的救赎 | **1** | **2** | **…** |
| **POS** | 超级喜欢，不看的话人生不圆满 | 经典的电影 | … |
| **MID** | 好看但这么高的分有点假了吧 | 在我的心目中，它一直都是最被高估的电影 | … |
| **NEG** | 真的不喜欢，不好看，没感觉 | 大众经典我从不感冒，为什么？我欣赏水平不行？ | … |
| **HOT** | 不需要女主角的好电影 | 关于希望最强有力的注释 | … |
| **NEW** | 超级震撼 | 看完后劲很大的一部经典！关于希望的另一种诠释！强推！ | … |