MEASURING STUDENT HAPPINESS LEVELS FROM FREE-FORM COURSE FEEDBACKS

Dr. Vivek Kumar Singh

Asst. Prof., Dept. of Computer Science, Banaras Hindu University, Varanasi – 221 005, India.

ABSTRACT

This paper presents an algorithmic design for opinion extractor from free-form course feedback texts. An unsupervised lexicon-based based algorithmic formulation is designed and implemented for identifying opinion polarity from free-form textual feedbacks written by students. The performance of the algorithm is evaluated on two different course feedback datasets obtained from ratemyprofessor website. The algorithm performance level observed is reasonably good. The algorithm produces easy to visualize graphical opinion summary from course feedbacks. This framework can be used in a course feedback system for automated interpretation of course feedbacks written by students.

INTRODUCTION

Feedbacks are responses expressed by user of a system in lieu of a services or product offered by system. Intention behind of feedback collection is usually to measure quality of service. Course feedbacks are a special kind of response information, in favor or against the course presentations or overall outcome of the course. Course feedbacks are important to measure the quality of a course and/ or experiences of students. Feedbacks also provide opportunity to the course coordinators to identify the weaknesses and improve course material. Feedbacks can be in various forms such as verbal, audio, binary responses, ratings, structured textual responses or unstructured free form text. The proposed system works with free-form textual reviews written by the students in feedback of a course. Free-form textual reviews are important because it allows the students to express themselves freely. Open ended textual responses provide freedom of expression to the student; a reviewer can express opinions better than binary or rating feedback systems. Reviews may contain sentiments of the students which are hidden but useful information for evaluation.

Open ended textual responses are hard to process by computer as it contain opinion expressed in natural language. The open ended text does not provide any fix scale where it requires to be quantified for evaluation. Unstructured textual responses have levels of

difficulties which ranges from ambiguity, polarity detection, quantifying extremeness of a positive or negative response, identifying the factor which is positive or which is negative. Sometimes a response can be much deeply evaluated which evaluates the variation of aspects in same or different scale. An open ended review may be multilingual, containing short cybernetic taxonomy, emoticons etc. A text response can be sarcastic, comic, and phrasal that makes it really hard to process. In addition spelling and grammatical errors makes it even difficult to deal with. Thus processing an open ended text response requires expertise in language processing and text analysis.

This system works on two different datasets collected from ratemyprofessor online course review system on chemistry and psychology and tried to solve the problem of sentiment analysis. The effort is towards tackling the problem at the document level sentiment quantification. An algorithm is devised which processes long reviews in various phases i.e. sentence segmentation, parts of speech based feature identification, lexical dictionary based lexical score assignment and finally sentiment score aggregation for a review.

SENTIMENT ANALYSIS

Sentiment analysis is natural language processing task to infer sentiment polarity of opinionated text. Mathematically it can be defined as a quintuple $< O_i, F_{ij}, S_{kij}, H_k, T_l >$ where, $O_i$ is the targeted object, $F_{ij}$ is a feature of the object $O_i$, $S_{kij}$ is the sentiment polarity of opinion of holder $k$ on $j^{th}$ feature of object $i$ at time $l$, and $T_l$ is the time when the opinion is expressed (Liu, 2009). The process involves identifying targeted object, which holds an opinion and then to identify the polarity of opinion, i.e. positive, negative or neutral.

Major methods of performing sentiment analysis task are either based on lexicon dictionaries which are usually unsupervised techniques and other major techniques are based of supervised kind where machine learning classifiers are devoted to perform sentiment classification task. Machine learning techniques require extensively annotated training data to learn the classes using a classifier function. Verbs and adjectives are holders of opinions. In this scheme, adverbs are seen as the polarity booster. Two SentiWordNet based approaches, SWN(AAC) and SWN(AAAVC) are employed in the system (Singh et al., 2013a, 2013b, 2013c). In SWN(AAC), ‘adverbs’ and ‘adjectives’ are used as features for identifying sentiment polarity. In SWN(AAAVC), ‘adverb+adjective’ and ‘adverb+verb’ patterns are used as features for sentiment polarity computation. It has been shown by Chesley et al. (2006) that verbs have a role in sentiment expression.

Algorithmic Steps for SentiWordNet based Adverb+Adjective Approach

**SWN(AAC)**

For each sentence, extract adv+adj combines.
For each extracted adv+adj combine do:
- If adj score=0, ignore it.
- If adv is affirmative, then
  - If score(adj)>0
    - $f_{SAAC}(adv, adj) = \min(1, \text{score(adj)} + s_f \times \text{score(adv)})$
  - If score(adj)<0


Algorithmic Steps for SentiWordNet based Adverb/ Adjective + Adverb Verb Combine Approach SWN(AAAVC)

For each sentence, extract adv+adj and adv+verb combines.

1. For each extracted adv+adj combine do:
   - If adj score=0, ignore it.
   - If adv is affirmative, then
     - If score(adj)>0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{adj}) = \min(1, \text{score(adj)} + sf \times \text{score(adv)}) \)
     - If score(adj)<0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{adj}) = \min(1, \text{score(adj)} - sf \times \text{score(adv)}) \)
   - If adv is negative, then
     - If score(adj)>0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{adj}) = \max(-1, \text{score(adj)} + sf \times \text{score(adv)}) \)
     - If score(adj)<0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{adj}) = \max(-1, \text{score(adj)} - sf \times \text{score(adv)}) \)

2. For each extracted adv+verb combine do:
   - If verb score=0, ignore it.
   - If adv is affirmative, then
     - If score(verb)>0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{verb}) = \min(1, \text{score(verb)} + sf \times \text{score(adv)}) \)
     - If score(verb)<0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{verb}) = \min(1, \text{score(verb)} - sf \times \text{score(adv)}) \)
   - If adv is negative, then
     - If score(verb)>0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{verb}) = \max(-1, \text{score(verb)} + sf \times \text{score(adv)}) \)
     - If score(verb)<0
       - \( f_{\text{AAAAC}}(\text{adv}, \text{verb}) = \max(-1, \text{score(verb)} - sf \times \text{score(adv)}) \)

3. \( f_{\text{AAAAC}}(\text{sentence}) = f(\text{adv}, \text{adj}) + \text{weightage_factor} \times f(\text{adv}, \text{verb}) \)

DATASET AND IMPLEMENTATION

3.1 Dataset

Two Datasets obtained from ratemyprofessor website comprising of 757 student feedback reviews for 3 Professors. Data for Andrew MacFarlane, referred to as Dataset1, comprises of 20 positive and 78 negative feedback reviews. Data for Joseph Morrissey, referred to as Dataset 2, comprises of 603 positive and 56 negative feedback reviews.
Table 1: Details of Dataset Used

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Course</th>
<th>No. of Reviews</th>
<th>Avg. Length (in Words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew MacFarlane</td>
<td>Chemistry</td>
<td>98</td>
<td>30</td>
</tr>
<tr>
<td>Joseph Morrissey</td>
<td>Psychology</td>
<td>659</td>
<td>35</td>
</tr>
</tbody>
</table>

3.2 Java Implementation

The algorithms are implemented in java language. A review is described as the source document in the picture. Then using Stanford parser the document is spitted into the sentences. Every sentence is then annotated with Part of speech tagger. Based on the methods respective features were selected. Using lexical dictionary and WordNet Synset sentiment scores were calculated. And for whole document the sentiment polarities are aggregated. Aggregated score is used as the sentiment polarity. Figure 1 presents the pictorial representation of the above said process.

RESULTS

The experiments result in document-level sentiment analysis results on two datasets using two different implementations (SWN(AAC) and SWN(AAAVC)). The table 2 presents the Accuracy F-Measure and Entropy values with two implementations. The result shows that for Dataset 1 SWN(AAC) achieve better accuracy than SWN(AAAVC) whereas for Dataset 2 SWN(AAAVC) achieve best accuracy. The table 3 presents percentage of feedback reviews labeled as ‘positive’ or ‘negative’ by different methods. The table 3 results shows that Dataset1 is more ‘negative’ which means Professor Andrew MacFarlane (Dataset1) more disliked by the student, whereas Joseph Morrissey (Dataset2) is more liked by the students. The table 4 presents the total number of ‘positive’ or ‘negative’ assigned by these two methods. The figure 1 presents sentiment polarity strength for Dataset1.

Table 2: Performance Result on Different Dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>Datasets</th>
<th>Dataset1</th>
<th>Dataset2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWN(AAC)</td>
<td>Accuracy</td>
<td>67.35%</td>
<td>73.30%</td>
</tr>
<tr>
<td></td>
<td>F-Measure</td>
<td>0.701</td>
<td>0.789</td>
</tr>
<tr>
<td></td>
<td>Entropy</td>
<td>0.207</td>
<td>0.114</td>
</tr>
<tr>
<td>SWN(AAAVC)</td>
<td>Accuracy</td>
<td>59.18%</td>
<td>76.93%</td>
</tr>
<tr>
<td></td>
<td>F-Measure</td>
<td>0.631</td>
<td>0.814</td>
</tr>
<tr>
<td></td>
<td>Entropy</td>
<td>0.212</td>
<td>0.116</td>
</tr>
</tbody>
</table>

Table 2 presents the comparison of both methods in terms of accuracy, F-measure and Entropy. For dataset1 accuracy of SWN(AAC) is 67.35% and SWN(AAAVC) is 50.18%.
The F-measure values are .701 and .631 respectively. This shows that SWN(AAC) obtains greater accuracy and F-measure than SWN(AAAVC) for dataset1. Entropy has no significant difference as it is .207 for SWN(AAC) and .212 for SWN(AAAVC).

For dataset2 accuracy of SWN(AAC) is 73.30% and SWN(AAAVC) is 76.93%. The F-measure values are .789 and .841 for the two datasets, respectively. SWN(AAAVC) shows greater accuracy and F-measure than SWN(AAC) for dataset2. Entropy has no significant difference as it is .114 for SWN(AAC) and .116 for SWN(AAAVC).

**Table 3: Category wise Accuracy Percentage Assigned by Two Approaches**

<table>
<thead>
<tr>
<th>Methods</th>
<th>Datasets</th>
<th>Dataset1</th>
<th>Dataset2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWN(AAC)</td>
<td>Positive</td>
<td>60.0%</td>
<td>73.79%</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>69.23%</td>
<td>67.85%</td>
</tr>
<tr>
<td>SWN(AAAVC)</td>
<td>Positive</td>
<td>65.0%</td>
<td>78.77%</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>57.69%</td>
<td>57.14%</td>
</tr>
</tbody>
</table>

Table 3 shows total percentage of positive and negative labels for both datasets. Table 3 presents numbers of positive and negative examples in datasets.

**Table 4: Total Number of Positive and Negative Labels Assigned by Two Approaches**

<table>
<thead>
<tr>
<th>Methods</th>
<th>Datasets</th>
<th>Dataset1</th>
<th>Dataset2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWN(AAC)</td>
<td>Positive</td>
<td>12/20</td>
<td>445/603</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>54/78</td>
<td>38/56</td>
</tr>
<tr>
<td>SWN(AAAVC)</td>
<td>Positive</td>
<td>13/20</td>
<td>475/603</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>45/78</td>
<td>32/56</td>
</tr>
</tbody>
</table>

The figure 2 presents sentiment polarity strength values for Dataset 1 and the figure 3 presents sentiment polarity strength values for Dataset 2.

**Figure 2: Sentiment Polarity Strength for Dataset 1**
CONCLUSION

The Paper presents an algorithmic design and implementation of unsupervised lexicon-based framework for measuring happiness of students from the unstructured free-form textual reviews of course feedback texts. Two algorithms which combine linguistic preprocessing and unsupervised Lexical Synsets for quantifying happiness are evaluated. The sentiment polarities are presented graphically, with polarity strength of such feedbacks ranging on the strength scale from -4 to 4. Results are identified with 67.35% accuracy by SWN(AAC) and 50.18% accuracy with SWN(AAAVC) for dataset1, 73.30% accuracy with SWN(AAC) and 76.93% accuracy with SWN(AAAVC) on dataset2. The results prove usefulness of the approach. This framework can be combined in any platform for such happiness polarity detection tasks in an affective and easy manner.

REFERENCES