Ranking based emotion classification

Srinivas Parthasarathy,Soorya Manian

December 13, 2013

1 Introduction

Classification of emotional attributes remains one of the toughest problems in the field of multimodal signal processing. The nuances and variability of emotions make them extremely hard to classify. Emotions can either have a discrete label such as Happy,Angry,Neutral,Sad etc or on a continuous domain such as Valence,Arousal etc. Their definitions are explained here [1]. While discrete labels have been classified to an extent the challenge to classify continuous labels is still daunting. In this project we use the Ranking-SVM and the Ranking based on stochastic gradient descent to classify emotions

2 Database

The sustained emotionally colored machine-human interaction using nonverbal expression (SEMAINE) database is used in this project. The SEMAINE project is an audio visual database and aims to build a Sensitive Artificial Listener agents. The database has different users or 'speakers' interacting with operators simulating 'SAL' operators. The operators are designed in such a way so as to elicit particular emotions from the users. These operators are then used to build an artificial listener `http://www.semaine-project.eu/`. The emotions expressed by these users have been annotated by 6-8 raters using GTrace [2] a software used for annotating continuous emotions on the Arousal and Valence domains Gtrace helps raters annotate the arousal and valence on a continuous domain. The ratings vary between -1 and 1 where -1,1 . Fig. 1 and Fig.2 shows the data collection tool. correspond to minimum and maximum values. The raters continuously flow between these values based on what they observe There are 10 speakers in the database talking to 4 different operator characters. The audio-visual signals are captured with a camera capturing the frontal view the user and a close talk microphone. These signals are then annotated by different raters. For the project we will be using Arousal domain annotation from the raters for classfication

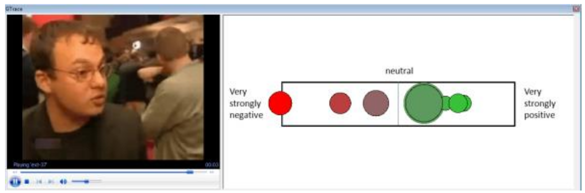

Figure 1: Gtrace. Emotional data annotation tool

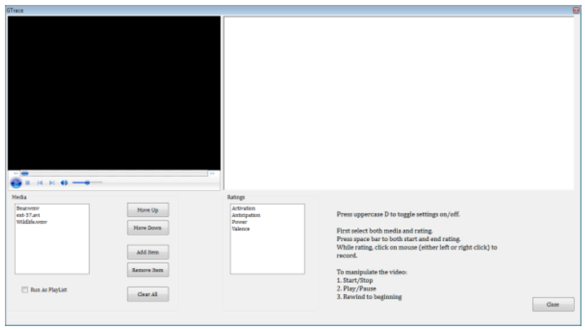

Figure 2: Gtrace. Emotional data annotation tool

3 Classfication

The challenge in classifying these values is the temporal change between different annotators. What one user may find to be more valent or more arousing may not be the same for another user , but the temporal information or the shift is almost always preserved [3]. Although users dont agree the absolute values they agree on relative the change with time. Figure 3 shows how different raters rate the sameclip. Clearly shows that the relative change is similar although absolute values are different. Thus this information was used as our labels and classification was done using a pairwise approach

3.1 Feature Construction

First the user speech was split into three second segments. Speech features for each segment was extracted using OpenSMILE [5]. The INTERSPEECH 2011 Emotion Challenge Featureset was used for extraction. The feature set contains 4368 emotion recognition speech features The labels were constructed as such - Given there are k annotators for the clips , we divide the data into three sets based on the percentage of consensus. If 66Thus we always have a balanced dataset. Features for the new labels are constructed by subtracting the features for the two segments and taking their sign alone. Thus if the features for the two segments , segment(A) and segment(B) are a and b respectively then, the features are sign(a-b) and the label will be 1 if segment(a)≥segment(b) or -1 if segment(a)≤segment(b) To balance the dataset the vice versa is also added. i.e sign(b-a) with the opposite label is also considered. This forms our dataset.

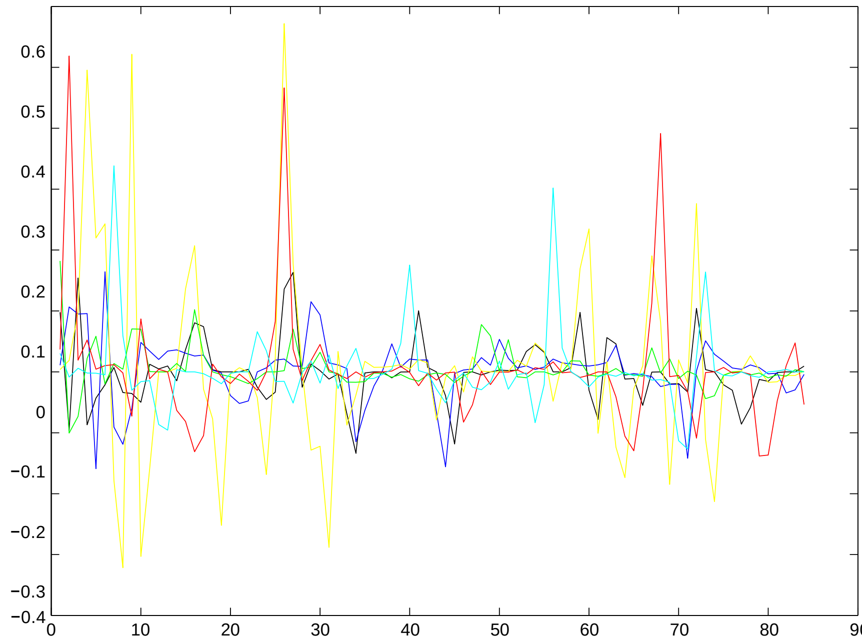

Figure 3: Ratings by 6 different annotators for the same clip

3.2 Classfication Algorithm

Two types of pairwise ranking algorithms was used for our classfication. The general learning to rank approach and the two algorithms are described below.

3.2.1 Learning To Rank

Learning to Rank is a machine learning technique in which the models are learned by supervised or semi supervised learning algorithms. The goal is to develop a ranking model from a training data. There are many approaches for developing this rank model .Pointwise approach, Pairwise approach and Listwise approach. The Training data set will typically contain list of items which will have limited order between items specified in the list. This order is implemented by giving a numerical order or binary judgment and in our case it is either 1 or -1.Learning to Rank has many applications in the field of Information retrieval, Data Mining, Natural Language Processing. Typical applications include document search, document summarization, collaborative filtering, machine translation and question answering.In this report a pairwise approach is used for emotion classification.

3.2.1.1 Rank SVM:

RankSVM is a pairwise approach for developing a Ranking model.There are many pairwise approaches proposed such as RankBoost,RankNet,,LamdaRank,GBRank.These approaches differ only in the way they are trained and the loss function used.The RankSVM develops a ranking model by minimizing a reguralized margin based pairwise loss.Web page ranking can be considered for explaining the RankSVM.The concept of links in webpage ranking was introduced to make search engines resistant against automatically generated web pages based on the content.The number of links the particular document measures reflects its importance or higher rank.The training data for the web page ranking consists of number of queries,for each query a set of documents ,for each(query,document)pair a feature vector $x_i \in R^d$ i=1,…,..n and the relevance judgments of these documents to the query.A set of preference pairs P can be constructed from the judgements for the given query.If (i, j)€P then i is preferred over j where i and j correspond to two audio segments.The RankSVM uses this objective as the minimizing function [6]

$$\frac{1}{2} ||w||^2 + C \sum_{(i,j) \in P} l(w^T x_i - w^T x_j)$$

where l is a suitable loss function such as $l(t) = \max(0,1-t)^2$.

3.2.1.2 Ranking using logistic loss (logistic regression ranking):

In this pairwise approach ,the original distribution of training examples D is expanded into set P of candidate pairs and learning proceeds over a set of pairwise example vectors. The set of candidate pairs implied from the dataset D is the set of example pairs (a,y_a,q) , (b,y_b,q) drawn from all examples in D where $y_a \neq y_b$ and $q_a = q_b$. When $y_a > y_b$,then a is preferred over b.
With P defined we find w by optimizing a pairwise objective function[7] :

$$\mathrm{Min}_{w \in R^m} L(w,p) + \frac{\lambda}{2} ||w||^2_2$$

The Loss function L (w,p) is defined over pairwise difference vectors from P :

$$L(w,p) = \frac{1}{|P|} \sum_{((a,ya,q),(b,yb,q)) \in P} l(t(\mathrm{ya} - \mathrm{yb}), f(w, a - b))$$

Logistic Loss:

The Logistic Loss is commonly applied in the linear classifier Logistic Regression.This Classification can also be used to predict real valued probability scores.The logistic loss function for y € [0,1] and y′ € [0,1] is

$$l(y, y') = y \log y' + (1-y) \log (1-y)'.$$

The Predicted function is

$$f(w, x) = \frac{1}{1+e^{-(w,i)}}$$

The Transformation function used is

$$t(y) = \mathrm{sign}(y)$$

Algorithm for Combined Ranking

Given:tradeoff parameter α,regularization parameter λ, training data D, iterations t

1: $w_\circ \leftarrow Ø$
2: for i=1 to t do
3: $((a,y_a,q)$, $(b,y_b,q)) \leftarrow$ RandomCandidatePair(P)
4: $x \leftarrow (a-b)$
5: $y \leftarrow t(y_a-y_b)$
6: $\eta_i = \frac{1}{\eta i}$
7: $w_i \leftarrow$ StochasticGradientStep(w_{i-1},x,y,λ, η_i)
8: end for
9: return w_i

For the Rank-SVM algorithm the SVM classifier, with a gaussian kernel, in Weka was used. For the Logistic Loss Ranking algorithm a MATLAB function was used. This is the same as the logistic regression in Weka

4 Results

For both algorithms three types of consensus was taken 66Further windows were placed on the time axis. Only clips that are between these windows was compared. ie clips within 5,10,15 (3 second) windows were used. A 10-fold cross validation was done for both algorithms Results in the form of accuracy of labels correctly recognized for the RANK-SVM algorithm is shown in table 1 and the Rank-Logistic in table 2 As expected when there is more consensus there is a better performance. As the window size increases the performance drops as the signals are not reliable anymore The RANK-SVM algorithm generally outperforms the Logistic Loss Ranking algorithm. This is due the fact that speech features are generally non-linear

Table 1: Classification results Rank-SVM

Seconds of Correalation ↓ \ Consensus percentage →	66%	80%	100%
5 (3 second windows)	67.0294	81.7073	87.069
10 (3 second window)	67.5097	75.4937	84.3173
15 (3 second window)	69.3506	72.6095	84.2105

Table 2: Classification results Rank-Logistic

Seconds of Correalation ↓ \ Consensus percentage →	66%	80%	100%
5 (3 second windows)	55.77	56.53	73.64
10 (3 second window)	54.68	62.07	64.81
15 (3 second window)	51.76	60.67	59.92

5 Appendix

The preprocessed dataset with speech features extracted from OpenSMILE is avaialble at `http://www.utdallas.edu/~sxp120931/` The feature extraction as described above was done earlier. This project focusses only on the classfication algorithms. Refer to Readme for details .

References

[1] Kensinger, Elizabeth A. ”Remembering emotional experiences: The contribution of valence and arousal.” Reviews in the Neurosciences 15.4 (2004): 241-252.

[2] Cowie, Roddy, and Martin Sawey. ”GTrace-General trace program from Queens, Belfast.” (2011).

[3] Metallinou, Angeliki, and Shrikanth Narayanan. ”Annotation and Processing of Continuous Emotional Attributes: Challenges and Opportunities.”

[4] Joachims, Thorsten. ”Optimizing search engines using clickthrough data.” Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002.

[5] Eyben, Florian, Martin Wllmer, and Bjrn Schuller. ”Opensmile: the munich versatile and fast open-source audio feature extractor.” Proceedings of the international conference on Multimedia. ACM, 2010.

[6] Chapelle, Olivier, and S. Sathiya Keerthi. ”Efficient algorithms for ranking with SVMs.” Information Retrieval 13.3 (2010): 201-215.

[7] Sculley, D. ”Combined regression and ranking.” Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010.