Emotion-specific Features for Classifying Emotions in Story Text

Harikrishna D M, K. Sreenivasa Rao
School of Information Technology
Indian Institute of Technology Kharagpur, Kharagpur - 721302, India
Email: harikrishna.dm@sit.iitkgp.ernet.in, ksrao@iitkgp.ac.in

***Abstract*—In this work, we are attempting emotion classification in view of synthesizing story speech. We are proposing emotion-specific text features (ESF) for classifying sentences from children stories into five different emotion categories: happy, sad, anger, fear and neutral. ESF is a five dimensional feature vector, where each dimension corresponds to weight of the sentence according to each emotion class. The dataset consists of 780 Hindi emotional sentences collected from children stories belonging to three genres: fable, folk-tale and legend. Part-of-speech (POS) and proposed ESF are used as features for emotion classification. Emotion classification performance is analysed using various combinations of features with three classifiers: Naive Bayes (NB), k-nearest neighbour (KNN) and support vector machine (SVM). The effectiveness of classifiers is analysed using precision, recall, F-measure and accuracy. The classification performance of 67.9% and 67.2% is achieved using POS and ESF respectively. The fusion of both features resulted an accuracy of 71.1%. Further, the importance of story genre information in emotion classification was observed from the experiments conducted on classifying emotions within story genre. An accuracy of 73.7% was observed after adding story genre information to the fusion of POS and ESF. SVM models outperformed other models in terms of classification accuracy.**

I. Introduction

The basic objective of this work is to classify the emotional sentences in the domain of children stories. The results of emotion classification will be used for deriving prosody modification rules, which is used for synthesizing story-style speech from the text using text-to-speech (TTS) systems. Synthesizing expressive speech involves embedding natural expressions into speech, according to the semantics present in the text. Story synthesis aims at synthesizing story-style speech from the text using TTS systems. To synthesize story speech, TTS systems can be developed using story speech corpus recorded from a professional story-teller. This approach requires a collection of hours of story speech corpus. Moreover, after collecting the story corpus, the entire corpus has to be manually annotated which is an expensive and laborious task. The limitations of this approach can be overcome by using rule-based story speech synthesis approach which aims at deriving the prosody modification rule-base specific to emotions. In order to derive the prosody modification rules specific to emotions, there is a need to identify the emotions present in the sentences of story text. In this work, we are attempting emotion classification in view of synthesizing story speech.

Generating an expressive, naturally sounding, story like speech from text is an extremely challenging task. In the domain of expressive TTS synthesis, expression dependent databases were recorded by a professional narrator to generate speech with different expressions in [1]. Statistical speech synthesis methods are popular in expressive speech synthesis and have shown promising results in synthesizing expressive speech [2]. In [3], a method to integrate the expression predictor and speech synthesizer to automatically generate expressive speech from text is presented.

978-1-5090-2361-5/16/$31.00 © 2016 IEEE

Emotion classification task can be considered as a multi-class classification problem. Given a sentence, we have to classify it into one of the predefined emotions. In this work, we have considered five emotions: *happy, sad, anger, fear* and *neutral*. Neutral is also considered as an emotion since the neutral sentences have story specific style that is different from neutral in normal speech. Emotion classification from the point of view of natural speech and human-computer dialogues is fairly extensive [4], [5]. In [6], a novel method for classifying news sentences into multiple emotion categories using an ensemble-based multi-label classification technique was proposed. In [7], a novel approach to construct an emotion lexicon was proposed, which was used to classify emotions in Chinese micro-blog text. In [8], emotion analysis on text corpus consisting of 22 children's fairy tales has been performed and the classification was performed with three classes, namely, positive, negative, and neutral emotion. As far as Indian languages are concerned, few works on sentiment analysis in Hindi [9] and Bengali [10] can be seen.

In this work, we are proposing emotion-specific features for classification of emotions. We conducted experiments with part-of-speech (POS) and proposed emotion-specific features (ESF). We also conducted experiments on emotion classification within story genres and claimed that adding story genre information to the fusion of POS and ESF features improves the performance of emotion classification slightly. The rest of the paper is organised as follows. Section II gives an overview of the proposed framework for emotion classification task. Section III presents the experimental results and discussions. Finally, we conclude this paper by drawing conclusions in Section IV.

II. Proposed Framework for Emotion Classification

The main motivation for this work is to synthesize story speech from story text. The flow diagram for the proposed story speech synthesis framework is shown in Figure 1. Initially, the given story text is classified into one of the five emotion classes namely *happy, sad, anger, fear* and *neutral*.

The story speech is synthesized by story specific prosody rule-set using a TTS system. In this section, we are discussing about the datasets, feature selection and evaluation measures for the emotion classification task.

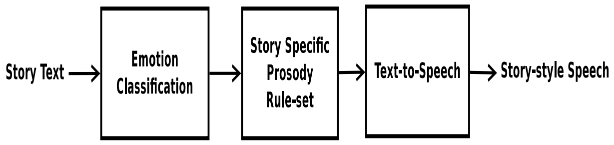

Fig. 1. Proposed framework to synthesize story-style speech from story text.

A. Database Preparation

The dataset consists of 780 Hindi emotional sentences collected from 90 children stories are used for experimentation. The 90 stories are composed of 30 stories of each genre: *fable, folk-tale* and *legend*. Details of emotional sentences considered for emotion classification task for Hindi are given in Table I. Even though most of the sentences in stories are neutral, we have considered 150 sentences of neutral emotion for this study. Emotions are hard to classify and are related to human subjectivity [11]. In order to classify emotions present in the sentence, we performed annotation task. Each annotator labels the sentences with one of the four emotions mentioned above based on mood, intensity and semantics of the sentence. A Sentence will be tagged as neutral if the annotator feels that there is an absence of happy, sad, anger or fear emotions. Four annotators were instructed regarding the annotation task. The measures of agreement among multi-annotators are discussed in [12]. Fleiss Kappa (κ) is a statistical measure of inter-annotator agreement and κ = 0.714 is observed which is considered to be substantial agreement.

TABLE I. DETAILS OF EMOTIONAL SENTENCES CONSIDERED FOR EMOTION CLASSIFICATION TASK

	Happy	Sad	Anger	Fear	Neutral	Total
Fable	26	77	21	22	50	196
Folk-tale	42	86	19	24	50	221
Legend	146	69	68	30	50	363
Total	214	232	108	76	150	**780**

B. Features

In this work, part-of-speech (POS) and the proposed emotion-specific features are explored for emotion classification. Details of these features are explained below.

- **POS features:** It is observed that most of the emotional words are nouns, adjectives, adverbs and intensifier emphasizes the emotional words. These observations motivated us to consider linguistic-based features (POS) for classifying emotions. POS features is a four dimensional feature vector comprising of count of nouns, adjectives, adverbs and intensifiers in a sentence.

  - # of nouns in a sentence
  - # of adjectives in a sentence
  - # of adverbs in a sentence
  - # of intensifiers in a sentence

- **Emotion-specific features:** We are proposing an approach to represent a sentence by emotion-specific features (ESF). ESF is a five dimensional feature vector, where each dimension corresponds to the weight of sentence according to each emotion class. The procedure to calculate the ESF for a sentence is given below.
  Emotion-specific feature vector (ESF) for a sentence is calculated as $< s_1, s_2, s_3, s_4, s_5 >$, where each value of s_i is sum of all the word's weights for each emotion class i. We have used weighting scheme to compute the weight of every word such that weight of $emotional\ words > non-emotional\ words$. For each word ($w_i$) of particular emotion class i, the weight of word is calculated as

  $$weight(w_i) = n_i \times log\frac{N}{N_0}$$

  where n_i is the number of times the word (w_i) occurs in sentences of particular emotion class i (e.g. happy), N_0 is the total number of sentences that the word (w_i) appears in other emotions (except happy), and N is total number of sentences of other emotions. In case if $N_0 = 0$, then N_0 will be set to 1.
  Each word is represented by an vector composed of five dimensions

  $$word = < d_1, d_2, d_3, d_4, d_5 >$$

  where the five dimensions corresponds to weights of the word specific to five emotion class: happy, sad, anger, fear and neutral respectively.
  For example, let us consider a sentence with three words i.e. $S = < w_1\ w_2\ w_3 >$. Each word is represented as a five dimension vector: $w_1 = < d_{11}, d_{12}, d_{13}, d_{14}, d_{15} >$, $w_2 = < d_{21}, d_{22}, d_{23}, d_{24}, d_{25} >$ and $w_3 = < d_{31}, d_{32}, d_{33}, d_{34}, d_{35} >$. Here each entry d_{ij} corresponds to weights of word (w_i) for j^{th} emotion classes i.e. happy, sad, anger, fear and neutral, respectively. Finally, ESF for the sentence is calculated as

  $$ESF = < s_1, s_2, s_3, s_4, s_5 >$$

  where $s_1 = (d_{11} + d_{21} + d_{31}),\ s_2 = (d_{12} + d_{22} + d_{32}),\ s_3 = (d_{13} + d_{23} + d_{33}),\ s_4 = (d_{14} + d_{24} + d_{34})\ and\ s_5 = (d_{15} + d_{25} + d_{35})$.

C. Evaluation

The performance of emotion classification are evaluated using three classifiers: naive Bayes (NB), k-nearest neighbour (KNN) and support vector machine (SVM). Classifier performance is evaluated using 10-fold cross validation. The performance of the classifier is evaluated using precision (P), recall (R), F-measure (F) and accuracy. Macro F1 measure is also used as a metric to assess the performance. Macro F1 computes a simple average of individual F-measures over classes.

$$P = \frac{No.\ of\ sentences\ correctly\ classified\ as\ class\ “x”}{No.\ of\ sentences\ classified\ as\ class\ “x”}$$

$$R = \frac{No.\ of\ sentences\ correctly\ classified\ as\ class\ \text{``}x\text{''}}{Actual\ No.\ of\ sentences\ of\ class\ \text{``}x\text{''}}$$

$$F = \frac{2 \times P \times R}{(P + R)}$$

$$Accuracy = \frac{No.\ of\ sentences\ correctly\ classified}{Total\ No.\ of\ sentences} \times 100$$

$$Macro\ F1 = \frac{\sum_{i \in C} F_i}{\mid C \mid}$$

where C is the set of predefined classes and F_i is the F-measure for the i^{th} class in C.

III. Experimental Setup and Results

In this section, we have attempted emotion classification using various features. Also, we tried classifying emotions within story genres and explored the use of story genre information for efficient classification of emotions. Details of the experimental setup and their results are discussed below.

A. Emotion Classification using POS and Emotion-specific Features

Various features such as POS, emotion-specific features (ESF) and fusion of both features are explored for emotion classification. Models are built using 780 sentences. Table II shows the result of emotion classification using various features. From the table, it can be noted that the performance of the models built using ESF features is less than models built using POS features which is still less than the fusion of POS and ESF features. The results show that both features can classify emotion significantly, and fusion of both features provides additional gain in classification performance. From the table, it can be concluded that $SVM > KNN > NB$, where $>$ indicates a better F-measure on the left-hand side than the one on the right-hand side for all emotion classes. *POS + ESF* features with *SVM* classifier gave the best performance for emotion classification.

B. Emotion Classification within Story Genre

After attempting emotion classification with various features, we tried to experiment with emotion classification within story genre. Since the emotional sentences are collected from children stories of different genres, the story genre information of the sentences were known in prior (refer Table I). Individual models are built with 196, 221 and 363 sentences of fable, folk-tale and legend respectively using POS features. Table III shows the results of emotion classification within story genre. From the table, it can be noted that there is a better performance of emotion classification within story genre. The highest F-measure of **0.769**, **0.89** and **0.954** are observed for happy, sad and anger emotions in fable, folk-tale and legend, respectively with SVM classifier.

TABLE II. Performance measures for emotion classification using POS and Emotion-specific Features

Features	Emotions	NB			KNN			SVM		
		P	R	F	P	R	F	P	R	F
POS	Happy	0.64	0.738	0.685	0.727	0.696	0.711	0.776	0.729	**0.752**
	Sad	0.551	0.608	0.578	0.629	0.681	0.654	0.705	0.659	0.682
	Anger	0.388	0.176	0.242	0.6	0.472	0.528	0.587	0.565	0.575
	Fear	0.488	0.526	0.506	0.578	0.632	0.604	0.516	0.632	0.568
	Neutral	0.651	0.633	0.642	0.692	0.72	0.706	0.679	0.747	0.711
ESF	Happy	0.638	0.709	0.672	0.747	0.676	0.71	0.789	0.719	**0.752**
	Sad	0.564	0.582	0.573	0.658	0.663	0.66	0.738	0.646	0.689
	Anger	0.382	0.166	0.231	0.605	0.425	0.499	0.604	0.537	0.569
	Fear	0.513	0.5	0.506	0.558	0.565	0.561	0.517	0.592	0.552
	Neutral	0.541	0.666	0.597	0.577	0.766	0.658	0.587	0.78	0.67
POS + ESF	Happy	0.673	0.771	0.719	0.754	0.736	0.745	0.808	0.747	**0.776**
	Sad	0.616	0.61	0.613	0.666	0.689	0.677	0.761	0.698	0.728
	Anger	0.462	0.222	0.3	0.663	0.546	0.599	0.656	0.601	0.627
	Fear	0.494	0.552	0.521	0.617	0.657	0.636	0.573	0.671	0.618
	Neutral	0.572	0.633	0.601	0.679	0.733	0.705	0.641	0.773	0.701

TABLE III. Performance measures for emotion classification within story genre

Story Genre	Emotions	NB			KNN			SVM		
		P	R	F	P	R	F	P	R	F
Fable	Happy	0.75	0.346	0.474	0.444	0.154	0.229	0.769	0.769	**0.769**
	Sad	0.642	0.883	0.743	0.713	0.87	0.784	0.66	0.831	0.736
	Anger	0.375	0.143	0.207	0.882	0.714	0.789	0.615	0.381	0.471
	Fear	0.75	0.682	0.714	0.75	0.818	0.783	0.682	0.682	0.682
	Neutral	0.7	0.7	0.7	0.635	0.66	0.647	0.842	0.64	0.727
Folk-tale	Happy	0.778	0.833	0.805	0.848	0.929	0.886	0.737	0.667	0.7
	Sad	0.8	0.884	0.84	0.778	0.895	0.832	0.885	0.895	**0.89**
	Anger	0.636	0.368	0.467	0.5	0.421	0.457	0.824	0.737	0.778
	Fear	0.6	0.625	0.612	0.632	0.5	0.558	0.667	0.75	0.706
	Neutral	0.756	0.68	0.716	0.78	0.64	0.703	0.635	0.66	0.647
Legend	Happy	0.834	0.932	0.88	0.89	0.89	0.89	0.921	0.884	0.902
	Sad	0.76	0.826	0.792	0.811	0.87	0.839	0.836	0.884	0.859
	Anger	0.918	0.824	0.868	0.954	0.912	0.932	1	0.912	**0.954**
	Fear	0.92	0.767	0.836	0.92	0.767	0.836	0.867	0.867	0.867
	Neutral	0.769	0.6	0.674	0.698	0.74	0.718	0.69	0.8	0.741

C. Emotion Classification using Story Genre Information

Motivated by the results of emotion classification within story genre, we tried to explore whether using story genre information improves the performance of emotion classification or not. Apart from the POS and ESF features, we have used story genre (SG) information as a feature. SG is a single dimension feature that corresponds to the story genre of the particular sentence. Models are built using 780 sentences with various features. Table IV shows the comparison of performance of emotion classification using various features. The highest macro F1 measure and accuracy of **0.73** and **73.7%**, respectively was observed for $POS + ESF + SG$ features with *SVM* classifier. From the table, it can be noted that adding story genre information to the combination of POS and ESF features has improved the performance of emotion classification.

TABLE IV. COMPARISON OF PERFORMANCE OF EMOTION CLASSIFICATION USING VARIOUS FEATURES

Features	**Macro F1**			**Accuracy**		
	NB	**KNN**	**SVM**	**NB**	**KNN**	**SVM**
POS	0.566	0.657	0.681	58	65.8	67.9
ESF	0.516	0.618	0.646	56.5	64.5	67.2
POS + ESF	0.551	0.672	0.69	60.2	68.7	71.1
POS + ESF + SG	0.57	0.69	**0.73**	61.5	70.3	**73.7**

IV. CONCLUSIONS

In this work, we have attempted emotion classification in view of synthesizing story speech in Hindi. We classified sentences from children stories into five different emotion categories: *happy, sad, anger, fear* and *neutral*. The dataset considered in this study consists of 780 Hindi emotional sentences collected from children stories of three genres: *fable, folk-tale* and *legend*. We have proposed emotion-specific features for classifying emotions. POS, ESF and fusion of POS and ESF are used as features for emotion classification. Also, we have attempted emotion classification within story genres. *NB, KNN* and *SVM* classifiers are used for analysing the performance of emotion classification. Experiments with different feature combinations reveal that *POS + ESF + SG* features with SVM classifier achieved the highest accuracy of **73.7%**. A little improvement in the overall emotion classification performance has been observed after adding story genre information. In case of classifiers, the performance of the models built using SVM outperformed other models in terms of classification accuracy.

ACKNOWLEDGMENT

We are thankful to the annotators who helped in the annotation task. The authors would like to thank Department of Information Technology, Govt. of India for supporting this research work, *Development of Text-to-Speech synthesis for Indian Languages Phase II*, Ref. no. 11(7)/2011HCC(TDIL).

REFERENCES

[1] J. F. Pitrelli, R. Bakis, E. M. Eide, R. Fernandez, W. Hamza, M. Picheny *et al.*, "The IBM expressive text-to-speech synthesis system for American English," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1099–1108, 2006.

[2] H. Zen, K. Tokuda, and A. W. Black, "Statistical parametric speech synthesis," *Speech Communication*, vol. 51, no. 11, pp. 1039–1064, 2009.

[3] L. Chen, M. J. Gales, N. Braunschweiler, M. Akamine, and K. Knill, "Integrated expression prediction and speech synthesis from text," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 2, pp. 323–335, 2014.

[4] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, vol. 40, no. 1, pp. 227–256, 2003.

[5] D. J. Litman and K. Forbes-Riley, "Predicting student emotions in computer-human tutoring dialogues," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. pp. 351–359, 2004.

[6] P. K. Bhowmick, A. Basu, and P. Mitra, "Reader perspective emotion analysis in text through ensemble based multi-label classification framework," *Computer and Information Science*, vol. 2, no. 4, pp. 64–74, 2009.

[7] C. Li, H. Wu, and Q. Jin, "Emotion classification of chinese microblog text via fusion of bow and evector feature representations," in *Natural Language Processing and Chinese Computing*. pp. 217–228, 2014.

[8] C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. pp. 579–586, 2005.

[9] P. Arora, "Sentiment analysis for Hindi language," Ph.D. dissertation, International Institute of Information Technology Hyderabad, 2013.

[10] D. Das and S. Bandyopadhyay, "Labeling emotion in Bengali blog corpus - A fine grained tagging at sentence level," in *Proceedings of the 8th Workshop on Asian Language Resources*, pp. 47–55, 2010.

[11] M. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are they different? affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Transactions on Affective Computing*, vol. 5, no. 2, pp. 101–111, 2014.

[12] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.