EASY FUZZY TOOL FOR EMOTION RECOGNITION
Prototype from Voice Speech Analysis

Mahfuza Farooque and Susana Munoz-Hernandez
Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain

Keywords: Emotion Recognition, Fuzzy Reasoning Application, Voice Speech Analysis, Fuzzy Logic.

Abstract: In human beings relations it is very important dealing with emotions. Most people is able to deduce the emotion of one person just listening his/her speech. Voice speech characteristics can help us to identify people emotions. Emotion recognition is a very interesting field in modern science and technology but to automate it is not an easy task. Many researchers and engineers are working to recognize this prospective field but the difficulty is that emotions are not clear. They are not a crisp topic. In this paper we propose to use fuzzy reasoning for emotion recognition. We based our work in some previous studies about the specific characteristics of voice speech for each human emotion (speech rate, pitch average, intensity and voice quality). We provide a simple an useful prototype that implements emotion recognition using a fuzzy model. We have used RFuzzy (a fuzzy logic reasoner over a Prolog compiler) and we have obtained a simple and efficient prototype that is able to identify the emotion of a person from his/her voice speech characteristics. We are trying to recognize sadness, happiness, anger, excitement and plain emotion. We have made some experiments and we provide the results that are 90% successful in the identification of emotions. Our tool is constructive, so it can be used not only to identify emotions automatically but also to recognize the people that have an emotion through their different speeches. Our prototype analyzes an emotional speech and obtains the percentage of each emotion that is detected. So it can provide many constructive answers according to our queries demand. Our prototype is an easy tool for emotion recognition that can be modify and improved by adding new rules from speech and face analysis.

1 INTRODUCTION

Our approach[1] for emotion recognition uses Rfuzzy(Ceruelo et al.,2008) which is an advanced extension of Fuzzy Prolog(Guadarrama et al., 2004). The Rfuzzy shares with Fuzzy Prolog most of its nice expressive characteristics: Prolog-like syntax (based on using facts and clauses), use of any aggregation operator, flexibility of query syntax, constructiveness of the answers, etc. Logic Programming is traditionally used in Knowledge Representation and Reasoning. Using also a fuzzy extension of Logic Programming enrich the expressiveness of the approach. In this paper we present an expressive and simple tool to make human emotion recognition by Logic Programing language means Rfuzzy. The rules that can be used for emotion recognition can be as complex as we want. We can take into account voice, face and other characteristics to deduce the emotion of a person. In our prototype we have just used voice speech analysis.

[1]This work is partially supported by the project DESAFIOS - TIN 2006-15660-C02-02 from the Spanish Ministry of Education and Science, and by the project PROMESAS - S-0505/TIC/0407 from the Madrid Regional Government.

2 RECOGNITION EMOTION BY VOICE SPEECH

Nicu Sebe and his colleagues(Nicu Sebe and Huang,2004) represent in a table in their paper the relation between emotion and vocal affects relative to neutral speech.

On the base of that table we know that for anger, speech rate is slightly slower, pitch range is very much higher, speech intensity is higher. On the other hand for sad emotion, speech rate is slightly slower,

Farooque M. and Munoz Hernández S. (2009).
EASY FUZZY TOOL FOR EMOTION RECOGNITION - Prototype from Voice Speech Analysis.
In *Proceedings of the International Joint Conference on Computational Intelligence*, pages 85-88
DOI: 10.5220/0002321700850088

pitch average is also slower, intensity lower. Speech rate is faster or slower for happiness, pitch average is much higher and intensity is higher and so on for other emotions. Speech characteristics that are commonly used in emotion recognition can be grouped into three different categories. The first one includes frequency characteristics (for example a pitch and pitch-derived measures) which are related to voiced speech generation mechanism and vocal tract formation. The second group contains various energy descriptors that are related to speech production processes (such as mean or standard deviation of energy of an utterance). The third group comprises temporal features, which are related to behavioral speech production processes (such as utterance duration, pauses)(Jaroslaw Cichosz). It can be seen that this studies are using fuzzy concepts (slower, higher, faster, etc.) to characterize voice speech with respect to the different emotions. So it is strait forward to use a fuzzy approach to represent this model. Jawarkar (Jawarkar and Fiete, 2007) tried to emotion recognition by Fuzzy Min-Max using neural classifier. On the base of the previous works and observing different measurements we selected just the three most representative characteristics: speech intensity, time difference between voiced and unvoiced speech, and time duration of each voiced speech. These are the variables that we have used in our prototype to analyze voice speech in order to improve emotion recognition.

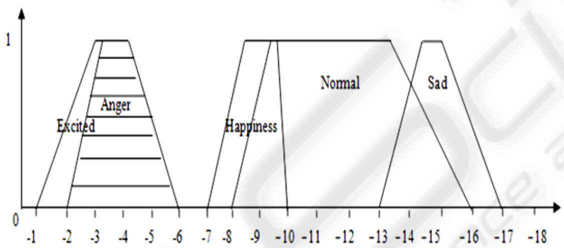

Figure 1: Fuzzy presentation of different emotions on the base of speech intensity.

3 METHODOLOGY

In our approach we try to identify emotion on the base of human speech. For this reason,we follow some systematic steps. Figure 2 shows the methodology that we have followed.

First of all we record some speeches of both female and male where the system used channel 2, 16 bit as bit depth and 44100hz for frequency. Then we synthesize voice record. To recognize emotional speech, we have taken average of speech intensity, the time difference between voiced and unvoiced speech and the duration of each word of a speech. On the base of these three results we characterize different speeches. We have got different values according to the emotional speech from the recorded speeches of different people. Fuzzy functions (as the ones represented at Figure 1) that can be defined for these three variables which we consider can be defined using the bounds shows at Table 1.

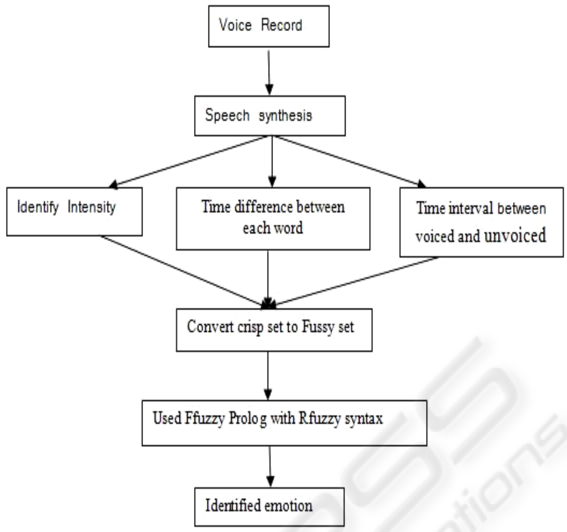

Figure 2: Methodology emotion recognition.

Table 1: Bound measurements according to different emotions.

Emotion	Intensity (db)	Difference (ms) between each word	Interval (ms) voiced and unvoiced
Sad	-12 to -16	70 to 105	80 to 350
Anger	-2 to -6	90 to 250	90 to 300
Normal	-8 to -16	40 to 60	100 to 200
Happy	-7 to -10	40 to 70	90 to 200
Excited	-1 to -6	90 to 200	95 to 200

From these crisp values we obtain fuzzy function definitions. In Figure 2 we show how it is done. After obtaining the truth value function of all emotional speeches,we concern on the interval between voiced and unvoiced speeches and time difference between each word. According to our experiment, we see that if people are sad then the time interval is more than other emotions and the time differences between each word is also long. Both measurements of time are measured by milliseconds (ms). On the other hand when people are anger and excited, then both time difference and time interval of speech are short comparing to other emotional speech. For normal speech we have seen that the value of three parameters is average. There is another important thing and that is the value of anger and excitement are very close but far from the values of other emotions.

4 EXPERIMENTAL RESULTS

In our experiment, first we record some speeches from different emotions of people. In this case we have taken records of cartoon characters from Internet where Bambi was in sad mood, Shaggy and the Beast were anger, Mickey was in happy mood, and Pocahontas was normal (in plane mood). According to these speeches, we calculate the crisp value of speech intensity, time differences between each word, time interval between voiced and unvoiced speech. Depending on these values, we converted the crisp values to fuzzy values to recognize different emotions.

We store the calculated crisp value of speech intensity of each person which we get from recorded voice.According to this record we see that the speech intensity of Bambi is 14db, Shaggy's is 12db, Beast's is 2db, Pocahontas' is 8db and Mickey's is 6db.

After more synthesis from the recorded speeches we store the crisp values of time differences between each word and the time interval between voiced and unvoiced speech which is measured in milliseconds(ms) according to each person. Here we see that the time difference between each word of Bambi is 250ms, Shaggy's is 43ms, Beast's is 70ms, Pocahontas's is 40ms, and Mickey's time difference between each word is 50ms.

In the same way we store the time interval between voiced and unvoiced speech of Bambi is 250ms,interval of Shaggy's speeches is 90ms, Beast is 80ms,Pocahontas is 150ms,and Mickey's time interval of voiced and unvoiced speech is 90ms.

Our prototype let us to ask any kind of query. As for example we are able to ask who is in a particular emotion, how much intense in someone's emotion, if a person has a specific emotion or not at all, etc. Suppose if we want to know about who is sad, we can write that query as:

```
?- sad(X,1).
```

Here X represent the person's name and 1 is the truth value of sadness of X. So, it is asking for the possibles values of X that provide a truth value of sadness of 100%. We get the result according to our database as:

```
X = bambi
```

If we want to know who is not sad, then we need to write our query as:

```
?- sad(X,0).
```

where the truth value 0 represents the equivalent of 'not at all'. So, we are asking for the people that is not sad at all. We get three results in this case:

```
X = beast ;
X = pocahontas ;
X = mickey
```

Beside these an interesting advantage of our approach is that we can qualify the query. We can constraint the truth value. For example, asking for the people that is "very" sad, with a sadness over the 70% for example. The query for this consult is:

```
?- sad(X,Y), Y > 0.7.
```

where the truth value Y represents how sad a person and X represent the name of the person. We will get the answer:

```
X = bambi,
Y = 1
```

On the same way if we interest to make a negative query to know who is a little sad but not with absence of sadness we can constraint the truth value as:

```
sad(X,Y), Y > 0.4, Y < 0.0 .
```

Then our answer will be:

```
X = shaggy,
Y = 0.2
```

We can check direct statements. For example, we can check that Mickey is not sad at all.In this case answer is affirmative because it is true.

```
?- sad(mickey,0).
Yes
```

We can query the database to consult the state of sadness of the people of the database:

```
?- sad(X,Y).
```

We obtain truth value of sadness emotion for all people included in the database:

```
X = bambi, Y = 1 ;
X = shaggy, Y = 0.3 ;
X = beast, Y = 0 ;
X = pocahontas, Y = 0 ;
X = mickey, Y = 0
```

We can say that our approach let us to model speech characteristics in an easy and crisp way and also let as represent fuzzy functions related to the perception of this characteristics for different emotions. So, we do not only provide emotion recognition but also we let the user to make any kind of expressive queries receiving constructive answers.

5 CONCLUSIONS

We recognize human emotion using fuzzy logic and fuzzy inference which allows to define default truth values and default conditioned truth values which helps to recognition emotion efficiently. So finding emotion according to speech, it is better than other techniques and policies. Extensions added to Fuzzy Prolog by Rfuzzy syntax are: Types, default truth values (conditioned or not), assignment of truth values to individuals by means of facts, functions or rules, and assignment of credibility to the rules. In this paper we recognized the specific emotional person and also their emotional level. We started our work with a very limited sample: we measure only three variables to recognize speech emotion but as future work we plan to take every possible voice measurable aspect not only consider the voice speech but also considering the face analysis and other context condition. That is, we will try to integrate both speech and facial expression using Fuzzy Logic to achieve more efficiently emotion recognition. We think that our approach presents a simple tool based on Rfuzzy syntax that can model perfectly the data of a recognition problem and that provides a huge expressivity to represent the fuzziness that is implicit in these kind of problems.

ACKNOWLEDGEMENTS

Authors would like to thank the suggestions and initial help of Suman Saha from the Programming Languages Laboratory of Hanyang University at Ansan Gyeonggi, Korea.

REFERENCES

Guadarrama S., Munoz-Hernandez S., and Vaucheret C. (2004). Fuzzy Prolog: A new approach using soft constraints propagation. Fuzzy Sets and Systems, FSS,GS-14, volume 144,1, pages 127-150, ISSN 0165-0114.

Jaroslaw Cichosz, K. S. Emotion recognition in speech signal using emotion extracting binary decision trees.

Nicu Sebe, I. C. and Huang, T. S. Multimodal emotion recognition. In *WSPC*.

Jawarkar, N. P. and Fiete (2007). Emotion recognition using prosody features and a fuzzy min-max neural classifier. In *IETE Technical Review Vol 24, No 5*.

Victor Pablos Ceruelo, Susana Munoz-Hernandez, Hannes Strasse. Rfuzzy Framework. In *Proceedings of the Workshop on Logic Programming Environmments, WLPE 2008. 15 pages*.