UNSUPERVISED HINDI SONGS' EMOTION RECOGNITION BASED ON LYRICS

APOORVA CHOUDHARI | 8181755808 | apoorvac@usc.edu
ARCHAN DUTTA | 8427371202 | archandu@usc.edu
SONIYA SADALKAR | 6771634935 | sadalkar@usc.edu
SUNIL MURALIDHARA | 1120242213 | sunilmur@usc.edu

INTRODUCTION

In the field of Music Information Retrieval, emotion detection of a song is a challenging and intriguing problem. The rise of recommendation systems in e-commerce spread to web-based music databases. This resulted in personalized music recommendations based on similar songs, emotion being one of the similarity criterions. This project will combine unsupervised learning with music information retrieval to achieve emotion detection of Hindi songs based on lyrics. The clusters so obtained can be used not just to classify new songs but also tap into some interesting inter-cluster and intra-cluster relationships.

1. Challenges:

- The corpus to classify Hindi songs is unavailable. Therefore, we will develop a corpus using unsupervised learning.
- Unsupervised Learning to create the corpus is much harder than its supervised counterpart.
- The emotion of a song is subjective.
- Processing Hindi characters.

2. What existing work has been done:

No such work is found for Hindi songs. Similar work has been done for Chinese songs.

3. How your contribution relates to that:

The output of the project would be a corpus of Hindi songs based on lyrics, grouped by emotions. There is no such publicly available corpus for Hindi songs. We will create metrics for identifying emotion of a Hindi word (concept similar to ANEW - Affective Norms for English Words). The metric would be a catalog of Hindi words and a value corresponding to a set of emotions. Such a catalog could also serve as a normative emotional rating for future research in problems involving Hindi language even in other domains.

METHODS

1.Materials:

1. Source : http://smriti.com/hindi-songs
    a. The songs are categorized by title, movie, year and actors. The lyrics are available in Devanagari script.
2. Annotation : No annotations available in the source
3. Data Size: approximately 10,000 songs

2. Procedure:

To cluster the songs, lyrics will be considered as features. Lyric based features could take any form such as bag-of-words, lexical features and/or orthographic features. The catalog previously mentioned will be referred during clustering. Since, a song could fall into multiple clusters, Fuzzy clustering will be applied.

3. Evaluation:

1. Compare the clusters obtained with the classes developed from a sample set of songs using existing classification methods (Example : Naive Bayes, Perceptron)
2. Performance (Precision and Recall) : F1 measure
3. Asymptotic analysis of clustering algorithm

REFERENCES CITED

1. A similar paper to classify Chinese songs :
https://pdfs.semanticscholar.org/4388/195556af596fd83050635e2310a7748d8166.pdf

2. Filtering social tags for songs based on lyrics :
http://scholarlyrepository.miami.edu/cgi/viewcontent.cgi?article=1274&context=oa_theses

DIVISION OF LABOR BETWEEN THE TEAMMATES

All the team members will be actively involved in all the phases of the project.

1. Annotation: Apoorva, Archan, Soniya, Sunil
2. Catalogue Development: Apoorva, Archan, Soniya, Sunil
3. Clustering: Archan, Sunil
4. Evaluation: Apoorva, Soniya
5. Documentation: Apoorva, Archan, Soniya, Sunil

WORD COUNT FOR THE DOCUMENT : 498