Script Character Emotion Recognition

Authors:
Biao Sun A20475197
Jianhua Tu A20480216
Ye Yu A20478640

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. Leave two blank lines after the Abstract, then begin the main text. Look at previous CVPR abstracts to get a feel for style and length.

1. Introduction and problem description

The importance of script to film and television industry is self-evident. A good script is not only the basis of good word of mouth and traffic, but also can bring higher commercial returns. Script analysis is the first link in the production chain of film and television content, in which the emotion identification of script characters is a very important task, which is mainly to analyze and identify the emotions of each character involved in every dialogue and action description in the script from multiple dimensions. Compared with the usual news and commentary text sentiment analysis, it has its unique business characteristics and challenges.

This project will use a part of film scripts as training sets, and the data of the training sets has been manually labeled. We need to analyze and identify the emotions of each character involved in every dialogue and action description in the script scenes from multiple dimensions.

2. Description of the data used in the project

3. Our Work

This is an area for what we have done.
Preparation work(study papers, tools, libraries).
Build the environment.
Our baseline results.

3.1. Method

Our algorithm, including how to process the data, and turns to the problem of multi-label dichotomies of sentences

3.2. Experiments

3.3. How to count the accuracy

The score of the algorithm in this competition is calculated by the common root mean square error (RMSE), and the emotion values corresponding to the six emotions identified by "text content + character name" are counted:

$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{6}(y_{i,j}-x_{i,j})^2}{6n}}$

score = 1/(1 + RMSE)

Where $y_{i,j}$ is the predicted emotion value, $x_{i,j}$ are the marked emotion value, and n is the total number of test samples. The final ranking is based on score.

3.4. Baseline score

3.5. The flaws of the method in baseline

4. What remains to be done

5. Final copy

References

Field	Name	Type	Description
id	String	DataID	-
content	String	text content	Script dialogue or action description
character	String	role name	Characters mentioned in the text
emotion	String	Emotion recognition results (in order)	Love value, joy value, shock value, anger value, fear value, sorrow value

Table 1. Training Data