

Predictive Methodology for Child Behavior from Children Stories

V.V. Sruthy, Amrutha Saju and A.G. Hari Narayanan
Department of Computer Science and IT, School of Arts and Sciences,
Amrita University, Kochi, India

Abstract: Story reading brightens our child's imagination and it develops curiosity in children. It helps them to develop brain. The graphics and rhyming style could help the children to understand the language patterns. They develop a difference between the real and make believe around the world and it helps to create their own ideas. Story reading and storytelling can emotionally impact children which help them to understand the change of different emotions according to the change of situations which is expressed through words. Seven basic categories of emotion (joy, fear, anger, disgust, surprise, sadness and neutral) of child are used here. The degree of these emotions depends upon the basic character of child. In this study, we applied SVM with the help of Hadoop MapReduce frame work environment to find out the child behavior according to the stories they read or listen. As per our analysis research, our system provides an efficient mechanism to select good stories for our child as well as developing an application in smartphone which does the same.

Key words: Child behavior, children stories, Hadoop/MapReduce, SVM, character, imagination

INTRODUCTION

A child is the biggest asset of a parent. No one in this world is like him. It is because the behavior and personality are made up of the combination of his own natural characters inherited from parents and external influence. So, minor change in the environment can affect him in the way he behaves and his thoughts. This is why some children cry if they are away from their parents. They like to get pampered all the time. Otherwise they feel unsafe and insecure and act differently from their nature. Of course, all children react to the physical environment differently. The basic behavior of child includes happy, violent, lazy, bullying, lying, short tempered, arguing, fighting and disrespect, self-esteem, yelling, stealing, playful and serious. These behaviors are pointing into the basic six emotions by Ekman (1971).

The child's mind gets an emotional impact with the stories they read as well as hear (e.g., expressive behavior, actions, goals, motives, appraisals, lazy, experimental feeling). This may or may not change over time as child grows and it depends according to the environment where the child grows. Emotional development is an effect of social experience including the cultural context (Calix *et al.*, 2010; Ghazi *et al.*, 2010).

This is a model to predict the behavior of children according to the stories they read. The behavior of children depends according to the environment they are in and the basic characteristic of the child (Alm, 2008).

The factors like age, gender, attitude, IQ level, etc. have an impact in child's behavior. Mostly the children of age from 4-10 like to listen and read stories. They are the fresh blooms though they have a clear mind there is definitely a chance that they highly motivated by the stories they hear both in positive and negative way. Som, it is very necessary to take caution while selecting stories for our children because the future of the nation is in the hands of our children. We hope that this model will help the parents and those people who interact with children to select good stories for their children.

Literature review

Background study: Predictive analysis is a special branch of data mining. Data mining is the methodology of extracting useful data from a huge database. Prediction can be of anything. Here, we are concentrating on the prediction based on text. Text mining is the branch of data mining referred to as mining quality information from the text. It is basically done with the help of pattern matching or statistical learning.

With the help of emotional analysis we can infer the emotions from text, thereby we can bring sensitivity to analytics. The lower level model uses a unique combination of language features such as word, phrases and overall sentiments with some machine learning algorithms. As of now the emotion analysis of children stories is done with context information which is used to clarify the meaning of some sentence or word (Zhang *et al.*, 2014).

Text based emotion recognition is often done with YouTube comments to develop interesting applications on mobile phones by exploiting the emotional states of users from the comments and this is classified according to the emotion using point wise mutual information measure (Yasminaa *et al.*, 2016).

All the above researchers have been successful in analyzing the emotion of text from context information and from YouTube comments. In this study, we use the predictive analysis technique in Hadoop/MapReduce environment to predict and classify the behavior of children from the stories they read (Saravanakumar *et al.*, 2015).

Proposed system: Based on unsupervised machine learning algorithm on (Agrawal and An, 2012) which is a normalized form here. The emotion category is same as of Ekman. It is represented by expressive words to find out the emotion which expressed in stories; we first classify its sentences which has emotional words (Calix *et al.*, 2010; Ghazi *et al.*, 2010). Implementation is done with the help of unsupervised machine learning algorithm in supervised machine learning model SVM.

For each word first computed the probabilities with each emotion category. PMI (Point wise Mutual Information) value computed which is nothing but what we find out using probability of a single word to belong to a particular emotion category is the value of the normalized form of the PMI (Point wise Mutual Information), between this word and the representative words of the emotion category (Yasminaa *et al.*, 2016). The probability of the whole sentence is obtained by finding the average of probability of the obtained number of classified words. The data set that we use to compute the different PMIs is built by examining the sentences which contain the emotional words. Emotion of child compared with the theme of the story (which is derived from the words used in the story related to emotions). It is normalized using the PMI between these words and the representative words of the emotion category. These calculations are done inside the data warehouse.

We assume the proposed system will give a satisfying result compared to the previous works. To get an efficient result we selected 2 types of stories, i.e., first one contains the affective words which corresponds to the target emotions and the second type is a neutral one. At last the probable result of the study is discussed (Fig. 1).

Hadoop: Hadoop is an open-source distributed data processing platform from Apache. It can be used as a data organizer and an analytic tool. The main components of Hadoop is MapReduce and Hadoop distributed file system.

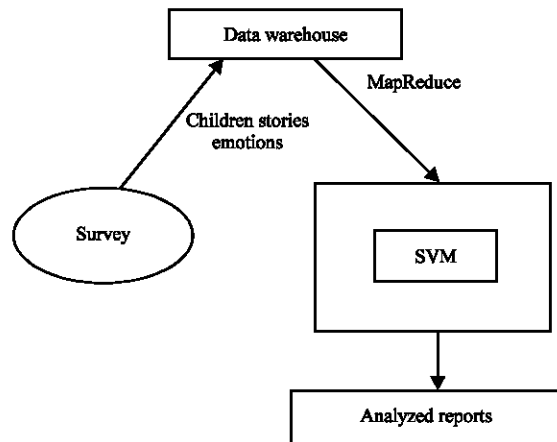


Fig. 1: Architecture of the system

The Map () function which resides in master node divides the input data or task into smaller ones and it is distributed in worker nodes, it will process the smaller task and give the results back to the master node. It is processed parallel. The Reduce () function collects the results from the Map function and combines them to produce an aggregated result.

SVM-Support Vector Machine: Support vector machine is a supervised learning model which is used in machine learning. It is mainly associated with classification and regression analysis of data. Given training data belongs to two categories. The SVM training algorithm will assign new examples to either of the two categories and make it a non-probabilistic binary linear classifier.

MATERIALS AND METHODS

The development of the system consists of two main phases first, we created the data set to train our system using children stories. We collected the stories of different kinds of emotions expressed to increase the performance of the system. The second phase is to run the unsupervised machine learning algorithm inside data warehouse to classify text entries.

Story-based emotion detection algorithm: The basics of the algorithm we propose are as follows:

Algorithm 1; Story-based emotion detection:

- Step 1: The sentences in the story is classified with the help of PMI which is an unsupervised machine learning algorithm and computes the relation between the Words (w) and the targeted Emotion (e). The Representative words (r) of targeted emotion category (eg: fright, terror, scare for fear)
- Step 2: Find the set sentences (Sw) which contains at least one occurrence of w
- Step 3: Find the sentences (Sr) which contains one occurrence of targeted Emotions (e)

Step 4: Find the N value (Number of lines inside our story which contains emotional words)

Step 5: Normalized PMI Calculation using Eq. 1:

$$\text{NPMI}(sw, sr) = \frac{\text{PMI}(sw, sr)}{\log(\text{card}(sw \cap (U_{i=1}^n sr))/N)} \quad (1)$$

RESULTS AND DISCUSSION

Analyzed report: After the analysis of stories and their emotions with the child basic characteristics data went through the Hadoop, the final results will be the story either creates a good impact on the child or a bad impact. This will help the parents, teachers and researchers to select good stories for the children.

CONCLUSION

In this study, we presented our story-based emotion detection system, based on an unsupervised machine learning algorithm and this is passed on to a supervised learning model using children stories as data set. We hope our system can achieves average precision and average accuracy (93.25 and 69.28%) after comparing with previous systems results close variation by using SVM as machine learning algorithms.

REFERENCES

Agrawal, A. and A. An, 2012. Unsupervised emotion detection from text using semantic and syntactic relations. Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, Vol. 1, December 4-7, 2012, Macau, China, pp: 346-353.

Alm, E.C.O., 2008. Affect in text and speech. Ph.D Thesis, University of Illinois at Urbana-Champaign, Illinois, USA.

Calix, R.A., S.A. Mallepudi, B. Chen and G.M. Knapp, 2010. Emotion recognition in text for 3-D facial expression rendering. IEEE. Trans. Multimedia, 12: 544-551.

Ekman, P., 1971. Universals and Cultural Differences in Facial Expressions of Emotion. In: Nebraska Symposium on Motivation, Cole, J.K. (Ed.). University of Nebraska Press, Lincoln, NE., USA., ISBN-13: 978-0803256194, pp: 207-283.

Ghazi, D., D. Inkpen and S. Szpakowicz, 2010. Hierarchical versus flat classification of emotions in text. Proceedings of the HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 2010, Stroudsburg, PA, USA., pp: 140-146.

Saravanakumar, N.M., T. Eswari, P. Sampath and S. Lavanya, 2015. Predictive methodology for diabetic data analysis in big data. Procedia Comput. Sci., 50: 203-208.

Yasminaa, D., M. Hajar and A.M. Hassanaa, 2016. Using YouTube comments for text-based emotion recognition. Procedia Comput. Sci., 83: 292-299.

Zhang, Z., M. Dong and S.S. Ge, 2014. Emotion analysis of children's stories with context information. Proceedings of the Annual Summit and Conference on Asia-Pacific Signal and Information Processing Association (APSIPA'14), December 9-12, 2014, IEEE, Siem Reap, Cambodia, ISBN:978-6-1636-1823-8, pp: 1-7.