IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue, 7, March 2019 E-ISSN: 2347-2693

# A Survey on Data Analytics for Personification using Machine Learning

**Prema P. Gawade**

Dept. of Computer Engineering, Pune Institute of Computer Technology, Savitribai Phule Pune University, Pune, India

*Corresponding Author: prema.gawade12@gmail.com*

**Available online at: www.ijcseonline.org**

***Abstract* -** With proliferation of the internet based media application; like SMS, video messaging, has resulted in a surge of data sharing. Applications like WhatsApp, Facebook, attracted large number of users because of easy chat conversations. WhatsApp is also used in small-scale industries for business purpose where user can write message is any language, in short form and understanding the context in chat is very important. There is no standard text which can be conventional machine understandable. Categorization can form personification; but there is lack of support to interpretation the content in the text messages with direct and indirect meaning. Map reduce framework can be used for transliteration process with frequent term visualization. Also messages can be with intuitive background where safety features are required to focus. Hence in this paper a survey for personification of embedded media data through the confluence of embedded media data analytic and machine learning techniques is proposed.

***Keywords*—** Machine Learning, Personification, Safety

## I. INTRODUCTION

Now-a-days much known embedded device like mobile smart phone has applications with embedded front and rear camera and audio-video capabilities. An amount of time spent with such embedded media, storage and sharing by any common man is also increasing on large scale. The huge rate of arrival and the abundance of such embedded media data make the safety decision more challenging. In this challenging scenario personification algorithms may be seen as a hope of a major key to the solution.

There are number of issues hidden with embedded media data sharing, storage and control with regards to personification and safety over the embedded media data. The personification can be one of the techniques to establish an muti-dimensional control by considering personal expressions of policies related to the acceptability, privacy, protection and sharing of such embedded media content or data.

The content arriving through such embedded media data applications in personal device may be categorized in entertainment, political, private or personal and threat and general share forwarding like birthday or good morning messages etc. This categorization can form the basis of safety, privacy and projection. This survey is an effort to add a research opportunity if interrelating media data with personification and safety. Today, because of lack of [1,2,3] such support this paper proposes an survey of personification of embedded media data through the confluence of embedded media data analytic and machine learning techniques.

## II. RELATED WORK

Social application like WhatsApp provides different forms of communications like group chats, user to user communication and broadcast messages. As amount of data retrieved in communication is extremely high, it is better to use Machine Learning as it supports different applications [4]. Each user in WhatsApp associated with profile and central system provides information of user registration, authentication, message delay as well as location, name, set of files which will be helpful to find background information of user. WhatsApp Messenger Artifacts gives this information. Personification of messages and human activity [5,6] is possible by getting structure of chat database includes msgstore.db which has messages, chat list, sequence. This information helps to get background details of communicator and to get intention of messages.

Supervised learning is helpful in object recognition, spam detection, speech recognition. As data in WhatsApp communication is with any form, so for classification or linear regression this method can be used. In unsupervised learning training dataset for clustering consists X set of input vectors, while it can work for different tasks. Semi-supervised learning is hybrid of unsupervised and supervised, is helpful when labelled data are expensive on not common. In applications like robotics, inventory, finance management

 

reinforcement learning is mostly useful. With applications of all these, for personification and safety it is better to go for semi-supervised learning [7] and ANN algorithm [8,9]. There is availability of data, computing power, Algorithmic innovation in machine learning, it can be used in Socio-economic impacts, transparency, bias, accountability, new uses of data, security, safety, ethics, new ecosystems. Also in social multimedia applications for personification existing learning algorithms can be considered [10,11]. To train learning algorithms data of patterns is used resulting in increase of performance or outcomes. To generate data pattern on which success of machine learning is dependent, different methods can be used to represent like probabilistic models, deep networks, manifold learning, and auto encoders. With supervised and unsupervised optimization is remained to be done to understand the success and failure. There are certain actions combined with position, background, and intention, to track this and to generate several points' different methods can be used. The input streams can be from IP camera and that input can be used in open source tools like OpenCV to generate relevant data points. Intersection of embedded system and maintaining output data is possible in real time image processing using [8,12] OpenCV.

For behaviour analysis based on analysis of features a novel motion descriptor can support. Optical flow can be improved with removal of errors and noises. Also fusing optical flow information with shape and trajectory information can effect to improve flow. Human automation interactive system [13] which uses support vector machine to recognize the behaviour with effectiveness in different algorithms gives scalability. Human intention expression can be applied to proactive collision avoidance to minimize need for sudden braking.

With existing survey a robot can predict human intention by effective teaching-learning prediction and conceptual model [14]. New technologies with business is connected to digital, virtual and physical network, because of which in business there is requirement of machine learning techniques to predict, observe and analyze human behaviour. Meta-heuristic algorithm aims to find near optimal solution at reasonable computing cost. Some meta-heuristics algorithms can be reviewed for intention personification in direct, indirect and hidden data with good, bad and ugly features. As WhatsApp is attracting social media, analysis of WhatsApp text messaging is evolved. Figure 1 shows number of WhatsApp users year wise where it increased rapidly. From WhatsApp communication data from different groups meanings of messages can be listed with data collection i.e. what intention behind message can be studied to implement with robotics as well [15,16].

WhatsApp user communication includes text messages, emoji's, images, audio, videos. To consider age-based distinction different age groups can be added with same database. Young adults may feel more social pressure to conform to norms about language set by society. Preliminary analysis required in-depth considering letters, diacritics, punctuation, spacing, capitalization.

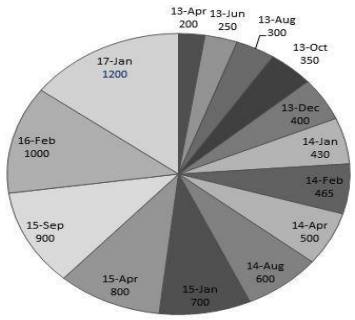


Figure 1. Number of WhatsApp users year wise

Hence survey for .Meta-heuristic algorithm aims to find near optimal solution at reasonable computing cost. Some meta-heuristics algorithms can be reviewed for intention.

## III. METHODOLOGY

Learning database, embedded media data contents such as text messages, images and video frames are taken as a input focus because these media contents are extensively used by different apps like WhatsApp, Facebook, Twitter etc. A messaging chat of WhatsApp data can be converted to simple English statements. The meaning of text message can be selected from the per-determined classes Viz; direct meaning, indirect meaning and hidden meaning with intention [17,18]. The data analytic can be proposed for developing distribution logic in convergence with pattern analysis to improve the success rate. For this method it is necessary to go with following stages:

*A. Data Acquisition:*

In data acquisition some WhatsApp groups data can be collected with attributes like sending time, sender name, data with chat contents.

*B. Text Pre-processing:*

Messages chat consists of upper, lower case, punctuations for which text preprocessing technique can be applied. In preprocessing removing punctuation, converting to lowercase, Tokenization are main steps. Original key values can be mapped with different values to create WhatsApp chat corpus model. The mapping database can be created from collection of WhatsApp messages.

 

*C. Map reduce framework for visualization*.

Map reducer can unite value pairs with key pair and it creates sorted list. Figure 2 shows map reduction gives semantic matching because of which visualization of data helps people understand meaning of text message.

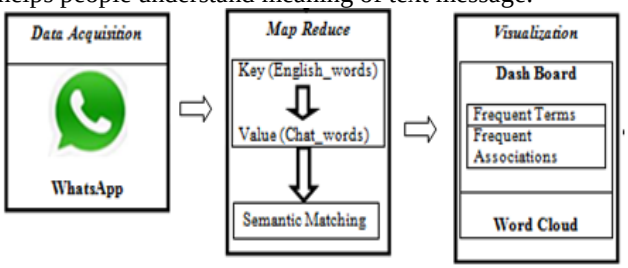


Figure 2. Text Visualization in WhatsApp messages.

Map reduce is a process of Map and Reduce wherein Map function applies to all value pair and Reduce function comprises list of values to create shorter list. More frequent words can be plotted in word cloud and minimum frequent words can be dropped. WhatsApp chat corpus model can be plotted as shown in table 1. Each key, value pair has unique id. If map function maps key value pair then data is replaced by key where English words are keys and WhatsApp messages are values.

To maintain database a document matrix with sender information, number of terms, documents, maximum term length and weighting off term frequency added with correlation limits for each term in range from zero to one.

Table 1: Data Collection – Text Messages

| Sr. No. | Key | Value |
|---|---|---|
| 1 | gm | Good morning |
| 2 | OMG | O My GOD |
| 3 | are | r |
| 4 | able | abl |
| 5 | about | abt |
| 6 | above | abv |
| 7 | after | aftr |
| 8 | afternoon | aftrnun |
| 9 | always | alwys |
| 10 | among | amg |

A classification and training as per message using Semi-Supervised analytic data proposition algorithm for selected classes of embedded media data and personification algorithm on selection meanings can be used for verifying the outcomes proposed.

Classification in accordance with me, private, protected, others are for direct, indirect, hidden meanings. This follow-me technology can be designed using Artificial Neural Network (ANN) algorithm [8,9] as different hidden layers are to consider.

## IV. RESULTS AND DISCUSSION

Analysis of algorithms must be data-driven is a general analysis. To prove or disprove any assumption we need to check for nature of data and previous studies with questionnaires. For data-driven approach significant differences must be studied evidences in different groups of data. To check and confirm this assumption we can clarify such differences in fact. The embedded media data contents such as text messages, images and video frames are taken as a input focus because these media contents are extensively used by different apps like WhatsApp, Facebook, Twitter etc.

Other than text messages there is use of emoji's, images, stickers, videos in communication where everyone should understand the meaning of text message so that safety can be achieved. Also because of unwanted annoying messages there can be wastage of memory space which can be properly utilized if personification is done properly. Table 2 shows the data collection format for emoji's and whether that is to be used as positive or negative intention. While communication details stored in msgstore.db can be helpful to get background details of user. Data collection can be done like as shown in table for better personification and safety.

Table 2: Data Collection – Emoji Messages

| Smiley | Positive Classifier | Negative Classifier |
|---|---|---|
| Grinning face | √ | ¬T |
| Bimming face with smiling eyes | √ | ¬T |
| Face with tears of joy | √ | ¬T |
| Rolling on the floor laughing | √ | ¬T |
| Grinning face with big eyes | √ | ¬T |
| Frowing face | ¬T | √ |
| Pouting face | ¬T | √ |
| Angry face | ¬T | √ |

Classification of Smiley in WhatsApp messages can be done in two classifiers. Positive classifier is favorable to personification and Negative classifier is in negative perception. Also follow-me technology can be helpful to classify in accordance with gender and age. After collection of these records a semi-supervised machine learning algorithm can be applied to get the better perception.

 

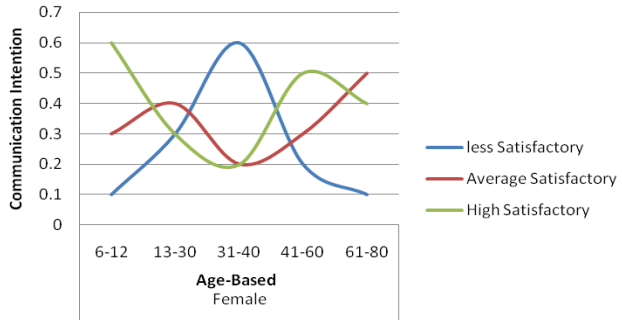


Figure. 3 Communication intentions by Female Group

As shown in above Figure 3 dataset of WhatsApp group analyzed for female age groups 6-12, 13-30, 31-40, 41-60, 61-80 where it shows that 6-12 age group female communicates with others without any intention as it shows high satisfactory. Like this data collection can be done for male and with different age groups with different types of communication to achieve with personification and safety.

## V. CONCLUSION AND FUTURE SCOPE

This paper gives survey about communication where personification and safety required more and research can be conducted for the same. Analytic classification method can be developed for embedded media data ambiguities, with use of Artificial Neural Network. A result of this paper can contribute in safety, privacy, management and data content management using selected personification intentions. The verification of socio-technological significance attainment proper selection of method can improve performance when operated with assessment of intention personification.

It can be observed that performance of content classification algorithm can be increased in comparison with existing heuristics. This method allows us to select function i.e. hypothesis which has properties personification and safety can be benefited in future with different socio-technological applications.

## REFERENCES

[1] M. Krishnakumar,"*Future Apparel Buying Intention : Mediating Effect of Past Apparel Buying Behaviour and Experience*.", SAGE publications, Vol: **19**, issue: **3**, pp. **737-755 2018.**

[2] Jun Liu,et.al.,"*Skeleton Based Human Action Recognition with Global Context Aware Attention LSTM Networks*", IEEE Trans., Vol: **27**, Issue: **4**, pp. **1586 – 1599**, **2018.**

[3] Moses L. Gadebe and Okuthe P. Kogeda, "*Personification of Bag.of.features Dataset for Real Time Activity Recognition*", IEEE, **2016.**

[4] Francesco Musumeci, et.al.,"*A Survey on Application of Machine Learning Techniques in Optical Networks*", April **2018.**

[5] Rizky Septiani, et.al.,*"Factors that Affecting Behavioral Intention in Online Transportation Service*",Elsevier, Vol. **124**, pp. **504-512**, **2017**.

[6] 18Stefanos Doltsinis, et.al.,"*A Symbiotic Human ML Approach for Production Ramp-up*",IEEE Trans. Vol. **48**, No. **3**, pp. 229-240, **2017**.

[7] Aravind Kota Gopalakrishna et.al.,"*Evaluating ML algorithms for applications with humans in the loop*" IEEE 14th International Conf. on Networking, Sensing and Control (ICNSC) **2017.**

[8] Namit Juneja, Rajesh Kumar,"*Generating Analytic Insights on Human behavior using Image Processing*",I2C2, pp.**1-7**, **2017**.

[9] Mingqi Lv et.al , "*BiView SemiSupervised Learning Based Semantic Human Activity Recognition Using Accelerometers*", IEEE Trans. on Mobile Computing, Vol: **17**, Issue: **9**, pp.**1991-2001**, **2018**.

[10] Zhanxiang Feng et.al.,"*Learning View Specific Deep Networks for Person ReIdentification*", IEEE Tras. on Image Processing, pp. 99, March **2018**.

[11] Xiantong Zhen, et.al.,"*Supervised Local Descriptor Learning for Human Action Recognition*", IEEE Trans., Vol: **28**, Issue:**9**, pp.**2035 – 2047**, **2017.**

[12] Aasia Ali, et.al.,"*Content Based Image Retrieval using Feature Extraction with Machine Learning*",IEEE, ICICCS, Madurai, **2017**.

[13] Krasimir Tonchev et.al.,"*Digitizing human behavior in business model innovation*", IEEE Conf. on wireless, Cape Town, South Africa, **2018**

[14] Xinbo Yu et.al., "*Neural control for constrained human-robot interaction with human motion intention estimation and impedance learning*", IEEE CAC, **2018**

[15] S. Rajagopalan, USA, "*Product Personification: Parag Model to Successful Product Development*", jmvsc. Vol. **6**, No. **1**, pp.**1-12**, March **2015**.

[16] Zhanxiang Feng, et.al.,"*Learning View-Specific Deep Networks for Person ReIdentification*",IEEE Trans., pp. **1-12**, **2018**.

[17] A. Shahroudy et.al, "*Deep multimodal feature analysis for action recognition in rgb+ d videos*," IEEE Tran. On Patt. Analysis and Machine Intelligence, April **2017**.

[18] Julio Cesar dos Reis, "*Recognizing Intentions in Free Text Messages: Studies with Portuguese Language*", IEEE 26th ICET : Infrastructure for Collaborative Enterprises (WETICE), pp.**302-307**, **2017**

## Authors Profile

*Mrs. Prema Gawade (Sahane),* pursed Bachelor of Computer Engineering from University of Pune, Pune in year 2004, and in 2013 Master of Computer Engineering from Pune University. She is currently pursuing Ph.D from Savitribai Phule Pune University. She is a member of ISTE and CSI since 2013. Total 20 research papers published in international journals. Her research work in Artificial Intelligence, Machine Learning focuses Computational Intelligence based education. She has total 11 years experience with 6 years industrial and 5 years in teaching.

 