Learning Orientation and Pattern Analysis using Sentimental Analysis with Report Generation

Rishabh Singh¹, Gaurav Rai², Rinkesh Gorasia³, K. Meenakshi⁴
Department of Computer Science
SRM Institute of Science and Technology, VDP, Chennai, India

Abstract:
Virtual Psychosis is an easier, more accessible way to gaining access to standardized psychometric tests and analysis by engaging with a virtual interface. The idea of virtual psychosis has been existent for some time now, but it's dynamics and long-term consequences have not been studied thoroughly. The field of psychosis constitutes to various methodologies dealing with the mental illness ranging from normal to severe conditions. Current solutions deals with a specific region while our solution is trying to challenge and cover the whole spectrum. Keeping in mind the complexities of the proposed solution, we would be harnessing the power of Artificial Intelligence and Machine Learning to arrive at solid conclusions. Our solution is trying to find the fine balance between generalisation and personalisation respectively. Generalisation deals with presuming the approximate mental state of the subject. Personalisation implies the recommendation of apt and accurate of learning orientation which will best fit the subject.

Keywords: psychosis, classification, recommendation, analysis, sentimental analysis, Bayes Theorem, TD-IDF.

I. INTRODUCTION

With the increase in the focus on learning and eduction, a huge section of the society remains left with this privilege due to lack of understanding of how one adapts to learn and grow with their own pace and way of learning. Given that, sentimental analysis and opinioninion can be a great tool for understanding learning patterns of the specially abled people and help the give a better chance at learning. The content is totally personalised according to their data extracted and how they have interacted with the software. This gives them the wholesome experience of learning and not making feel the left out. This can be also a great insights of how a particular students learns and interested in and what way one can be taught.

II. OBJECTIVE

The objective of this paper is to give out analysis and know learning patterns of specially abled people and give them curated content using that data to help them learn better and grow.

III. LITERATURE SURVEY

Title
Fault Detection and Diagnosis

Description
Fault detection can be classified based on two models. One model is the detection by signal where periodic signals, stochastic signals, non-stationary signals are also included. The other is a model based approach. Here, the parameter of estimation is employed for fault detection whereas the signal-based detection model methods are used for imbalance or bearing faults, knocking etc. In signal based detection the measured signal show oscillations that are either harmonic or stochastic in nature or both. Sensors like electric current, positions, force, pressure can result in a variety of higher frequencies than the usual dynamic frequencies.

There are two methods of fault diagnosis:
Classification method and inference method. The classification method works when we have a structural knowledge between the symptom and the fault. Few of the techniques are Bayes classifier which is based on the statistical distribution of the symptom. The other are decision tree which are based on the series of questions to be answered. The geometric classifier are based on membership of a class of the data point from the point of reference. Another method includes polynomial classifier which is the functional approximation. In model based approach includes measuring input signals and output signals. The detection methods generates residuals and parameter and estimates. These are also called features.

Title
Sentimental analysis and Opinion Mining.

Description
In sentimental analysis it deals with identifying and classifying opinions/sentiments from a given source. In this growing age of social media where it is creating a huge amount of sentiments and opinions among the people about various things they see on the internet there is rich data of these sentiments or opinions in form of likes, tweets, status updates, blogs etc. Here machine learning approach is used for analysing sentiments from the text. All the sentiments can be classified into three different categories. Positive, negative or neutral. The source data is extracted, the classification model or algorithm is applied on the collected data. The features are extracted via N-gram modelling technique. The sentiments/opinions are categorised either in Positive, negative or neutral.
Title
Cloud Computing and its configuration

Description
Cloud computing in simple terms can be defined as the computation of software, processing the data, accessing it and storing them that does not require the end-user the knowledge or experience of the infrastructure or architecture and configuration of the server or which delivers there services. It is the fastest way for delivering on demand services over the internet. It can also be seen as a by-product of ease of access to remote computing.

The important characteristics of cloud computing are:
Agility which improves the user ability to rapidly scale the architecture.
API also know as Application Programming Interface which enable machines to communicate with the cloud in the same way like humans and computers. Device and Location independence give users access to use web browsers regardless of their location or device. Cost with the use of public clouds the cost is greatly reduces. The infrastructure is typically provided by a 3rd party vendor. Scalability with dynamic and on-demand service the scalability can be easily done without having to engineer for peak loads, performance or anything else. Security With cloud due to centralisation of data the security improves a lot.

Advantages of Cloud Computing
Management becomes easy.
Services are uninterrupted
Scalable Infrastructure
Reduction in the cost

IV. EXISTING SYSTEM
A definitive question/answer model based platform that covers all the major use-cases and makes sure that the subject finds the closest one to their conditions, following which, the detailed analysis would take place, resulting a set of advices and instruction that has to be implemented by the user positively. It is the responsibility of the platform to steer clear of any negative nuances or provocation syllable, in order to avoid negative progression of the conversation, resulting in invalid conclusion or fatal consequences. The challenges faced are Accuracy of the system is user-dependant. Question/Answer model is a rigid and immutable model. Covers most of the use-cases but not all of them. System is not smart, as it is context-free and works on a rigid model. System does not learn from its mistakes

V. PROPOSTED SYSTEM
With a task to detect various emotions and mental condition, out model breaks the task into fundamental sub emotions and detects each one of them using a standardised test, which is then analysed to keep the result as error free as possible, Once the results are generated a prediction on the mental state is made, depending upon which learning orientation plan is recommended which trains itself to provide a better personalised content for the subject.

VI. COMPARISON OF EXISTING SYSTEM WITH PROPOSED SYSTEM
With a task to detect various emotions and mental condition, out model breaks the task into fundamental sub emotions and detects each one of them using a standardised test, which is then analysed to keep the result as error free as possible, Once the results are generated a prediction on the mental state is made, depending upon which learning orientation plan is recommended which trains itself to provide a better personalised content for the subject.

VII. SYSTEM ARCHITECHTURE
System network architecture for data detection is illustrated in Fig.1.1.

VIII. ALGORITHMS USED
Gaussian Naive Bayes Algorithm: The Gaussian naive based algorithm [6] is very effective when the dataset is small and high accuracy is needed. It works on conditional probability, wherein each feature is given equal weightage and each feature is treated as an independent entity. In Psychosis state detection, Gaussian naive based fit well as less amount of data with equal weight age provision is needed. Which results in good accuracy. Each mental state is given equal importance and in dependence. When the classification is done, the hypothesis (h) can be the class to assign new data (d).

The Bayes Theorem can be applied as follows:
\[ P(h|d) = \frac{P(d|h)P(h)}{P(d)} \]
P(h|d) : Probability of hypothesis h with data d. Also known as Posterior probability.

TF-IDF to detect the various conditions and classify for each one of them.
P(d|h) : Probability of data d given that hypothesis h is true.
P(h) : Probability of hypothesis of h being true irrespective of data d
P(d) : Probability of data d irrespective of hypothesis (h).

The Naive Bayes can have two types of probabilities which are class probability and conditional probability.
The class probability simply gets the frequency of instances that belong to each class divided by total number of instances.
P(Class 1) = count(Class 1) / (count(Class 0) + count(class 1))

The conditional probability are frequency of each value for a given class value divided by the frequency of instances of that class.

In Gaussian Naive Bayes probability of new x values are calculated using the PDF (Gaussian Probability Distribution Function).

To make predictions these parameters are put into the Gaussian PDF with a new variable input and in return it provides the estimate probability of new value for that class.

**TF-IDF Algorithm**

The TF-IDF Algorithm [7] is used to find the relevance of the word in the given document. Hence it used to detect the emotions depending upon the occurrence of few presumed words. It is used here to score the those words depending upon the usage along with the context. TF-IDF will give low rank the commonly used words to avoid the biased or improper results in the final output. The output is in the range from 0 to 1, more the occurrence of the word, more will be its value in the previously mentioned range. The higher the score is, the rarer term is. The smaller the score, the more common it is.

Once we have transformed words into a set of numbers, the TD-IDF score can be fed to algorithm such as Naive Bayes algorithm which can greatly improve the results of more basic methods. The TD-IDF can be used in various ways such as Information retrieval and keyword extraction which can be very useful in an application like ours. Extracting out particular keywords and finding it frequency and feeding it to the Naive Bayes algorithm can give out some useful information and results.

To put it in a mathematical form, the TD-IDF score for the word t in a document d from the Document set D is:

\[ \text{tf idf}(t,d,D) = \text{tf}(t,d) \cdot \text{idf}(t,D) \]

Where:
\[ \text{tf}(t, D) = \log(1 + \text{freq}(t, d)) \]

**IX. TECHNICAL MODULES**

- Pre-Processor
- Cloud Configuration
- Classification Engine
- Recommendation Engine
- Fault-Tolerance

**A. Pre - Processor**

Users are greeted with pleasantries to lighten up the mood or environment. Human-like conversation pans out and asks users about how they are, how they feel, if they feel conflicted etc to gauge as much information as possible for the recommendation engine. Agility of both the algorithms used, depends on the context driven from the initial conversations. Hence, context collection takes place.

**B. Cloud Configuration**

The collected conversation is sent to our cloud servers. The inferred context is sent to our context store, inside the cloud servers. The informations has to be sent to the classification and recommendation system respectively through cloud state management.

**C. Classification Engine**

The classification engine deals with the collected conversation between our intelligent fault-tolerant systems and the user itself. Naive bayes is used for internal classification inside each mental health problem. The most probable mental illness with the highest gross definitive score is selected and classification is
ensured for at-par levels using TF-IDF classification algorithms. On successful verification of classification, the probable mental cause of the subject is selected and sent to the recommendation engine along with metadata payload through internal cloud web hooks.

D. Recommendation Engine
The most probable cause along with its meta-data is received by the recommendation engine. The recommendation engine pulls the relevant context concerning the meta-data from the context store. The context and metadata are horizontally matched and are given a standardised score according to the algorithm. The recommendation engine then uses this score to order the priority level of advices and instructions which are then sent back to the user.

X. CONCLUSION
The amalgamation of fault tolerant systems with machine learning algorithms and sentimental analysis helps in understanding the opinions, sentiments when exposed to different things and this gives out patterns on how they can learn in a better way. This patterns or data can be used to give out curated content to the individuals so that it becomes personalised learning for them and also it can be used as analytics for teachers and parents to know how is one able to learn and how to proceed with them. Hence sentimental analysis and fault tolerance becomes a great tool for report generation and learning orientation.

XI. REFERENCES


