NOVATEUR PUBLICATIONS
International Journal of Research Publications in Engineering and Technology [IJRPET]
ISSN: 2454-7875
VOLUME 3, ISSUE 8, Aug. -2017

IMPROVING PREDICTIONS USING QUALITATIVE PARAMETERS

PRASHANT MAHAJAN
Department of Computer Engineering, Atharva College of Engineering, Malad West, Mumbai, Mumbai University, India, prashantmahajan70@gmail.com

PRATIK DESHPANDE
Department of Computer Engineering, Atharva College of Engineering, Malad West, Mumbai, Mumbai University, India, deshpandepratikr@gmail.com

TEJAS NANAWARE
Department of Computer Engineering, Atharva College of Engineering, Malad West, Mumbai, Mumbai University, India, nanawaret@hotmail.com

PROF. MAHENDRA PATIL
HOD, Department of Computer Engineering, Atharva College of Engineering, Malad West, Mumbai, Mumbai University, India, mahendrapatil@atharvacoe.ac.in

ABSTRACT:

Selecting appropriate parameters while making any prediction model is a tedious task. Often, while constructing a prediction model, categorical variables are ignored. If we include more qualitative parameters for prediction, the observed results will have more accuracy.

Neural networks help in a proper learning methodology which utilizes the concept of machine learning. When prediction is to be made, the human behavioral patterns hamper the test results as it plays a crucial role in any decision making. Employing qualitative parameters in decision making, accurate conjectures are possible. Qualitative parameters are considered fuzzy in nature and neural networks, which is one of the major components of the soft computing, works very well with incomplete data.

In this paper we have discussed how qualitative parameters will help in improving, the prediction accuracy, and the decision making logic, to make predictive models more sustainable and robust.

KEYWORDS: Machine Learning; Prediction Models; Neural Networks; Behavioral Patterns; Qualitative Analysis.

1. INTRODUCTION:

Data collection for machine learning is an important aspect. The data that is being used for the training purposes helps in improving the results drastically. Often many times the qualitative parameters are ignored as they are difficult to judge, but they play a crucial role in the decision making. Decision making focuses on predicting the results inferring from the input parameters, which are, the variables in the dataset. Decision making is the base of machine learning. Before we proceed with the decision-making process, we should first observe the data and then look how the decision can be achieved. The decision process is carried out by using trees in machine learning. Human like decision making in machine learning is yet another field in itself. Decision-making task depends on the data and how the data is used, can it be trained again if incorrect, the amount of noise in the data and how efficiently the algorithm handles the data.

The variables or the parameters that help in the decision making can be simply classified into qualitative and quantitative parameters. The qualitative parameters are those that state the quality and are inferred from the data set like knowledge about some programming language cannot be judged directly like marks but can be inferred from the certifications and internships or projects made by using those languages. Quantitative parameters, on the other hand, state their quantity. For example, the number of programming languages known by a person is a quantitative parameter, but the knowledge about language becomes a qualitative parameter.

Qualitative parameters thus will affect the decision making to a large extent, but the quantitative parameters will help to examine the data and generate qualitative parameters. Quantitative parameters are simple to infer and produces decisions faster, but with the support of the qualitative parameters, accuracy can be improved. In many cases quantitative parameters become paramount for decisions like the subject marks can be stated as a quantitative parameter as more the marks, the better will be the student's performance. This in another scenario can become qualitative parameter as higher grades will indicate quality. The major limitation of the qualitative parameters is that they are difficult to deduce and furthermore, they should be represented as a form of probabilistic function (between 0 and 1) and not boolean form (only two values yes and no).

Qualitative parameters are difficult to include in the decision-making process. Some obstacles encountered during this process are:

1. A lot of qualitative parameters can be obtained, and these are all specific to the system. This variation is challenging to manage.
2. They are diversified in their applications and are difficult to integrate while building the learning model.
3. It is an effective way to consolidate qualitative and quantitative parameters but creating the learning hypothesis becomes tedious.
4. When qualitative parameters are inferred from some other variables, the probabilistic function should represent the quality accurately between Zero and One.

Hitarthi Bhatt et al. [6] designed an efficient algorithm that considers qualitative parameters, but the formula used by them is a boolean function that converts all the values to True or False (One and Zero) whereas it should have been probabilistic function that ranges the values between some values generally between zero and one.

Use of additional tests can be done to infer the qualitative parameters. It can also be done by formulating the parameters or combining two qualitative parameters resulting in a unique third parameter. The qualitative parameters decide the learning postulate and the decision-making process. This paper highlights different types of quantitative parameters currently being used for numerous prediction purposes. Later on, we've explained what qualitative parameters is and their importance in Machine Learning. Also, we have presented ways to integrate qualitative parameters with your current set of parameters. At last, we've given an example of student placement prediction using Neural Network which utilises qualitative parameters to improve decision making.

2. LITERATURE SURVEY:

The research on qualitative parameters for their usefulness is being done in various modelling systems for a couple of decades. In 1985, Kupiers published a research paper which realises best features amongst different qualitative algorithm and presents a new algorithm QSIM which is an efficient constraint-satisfaction algorithm [2] makes more appropriate qualitative distinctions since the { +, 0, - } semantics can collapse the distinction among the increasing, stable or decreasing oscillation which can be used in constructing and validation base of mechanism descriptions. Herman Aguinis et al [9] orchestrated a 30 year review, from 1969-1998, of all the articles published in Personnel Psychology (PP), Academy of Management Journal (AMJ), and Journal of Applied Psychology (JAP), and calculated the impact of qualitative or categorical variables, using multiple regression, and found out that the studies conducted in the late years had a greater effect size than those conducted in the previous years. Work done by Hocaoğlu attempts to clarify the differences and similarities between Traditional Simulation (TS) and Qualitative Simulation (QS) and bridge the gap in conceptual level [14]. Comparison between TS and QS is divided into two main headlines such as the perspective from Artificial Intelligence (AI) and simulation point of view. Rajguru et al. [10] present the performance analysis of various load balancing algorithms in the distributed system based on different qualitative as well as quantitative parameters. Their research shows that dynamic load balancing algorithms are always better than static as per as overload rejection,fault tolerant, response & waiting time is concerned.

In [12], a framework for building Gaussian process model which incorporates both qualitative and quantitative factors was proposed. These models were developed by constructing correlation functions with both types of factors. This Gaussian process model was illustrated with a real example for modelling the thermal distribution of a data centre involving various qualitative parameters like diffuser location, rack heat load non uniformity and return air vent location. Han et al. [3] introduce Baysian methodology for the prediction of computer experiments which involves both qualitative and qualitative inputs. Their approach involves the use of the Gaussian stochastic process model to deal with multiple qualitative data. In [15], a new modelling approach was proposed based on a hierarchical model structure to deal with systems consisting of qualitative and quantitative parameters. This integrated modelling for stochastic simulation and optimization can describe complex system more comprehensively.

2.1 USE OF QUALITATIVE PARAMETERS IN VARIOUS SYSTEMS:

Rutledge et al. [1] combined qualitative and quantitative computation in a ventilator therapy planner called as Vent Plan. It is a qualitative modelling system which consists of a belief network and a multi attribute value model work together to refine a patient-specific physiological model and settings of the ventilator (according to recommended treatment plans). Peter

NOVATEUR PUBLICATIONS
International Journal of Research Publications in Engineering and Technology [IJRPET]
ISSN: 2454-7875
VOLUME 3, ISSUE 8, Aug. -2017

Clark and Stan Matwin [4] designed a system that produces classification rules that are generated by the training set input by using numeric-simulators. The rules made using qualitative models perform twice better than those without qualitative models. Allison R. Fleming et al. [8] designed a system that evaluates characteristics influencing the placements to a larger data sample by conducting qualitative interviews, but limited a few characteristics like ethnicity, age and work experienced which hampers the results. In [13], different generalised methods for dynamic, complex mathematical models were proposed and compared which are widely used in all the fields to interpret the behaviour of complicated systems. The advantages and problems were discussed by applying Sheffield Dynamic Global Vegetation Model, proposed and made within the UK Centre for Terrestrial Carbon Dynamics.

Based on above research, we proposed usage of qualitative parameters for the decision-making process in Machine Learning. The integration of qualitative parameters to the current system will help in achieving higher accuracy with better outcomes.

3. QUALITATIVE PARAMETERS IN DECISION MAKING:

Numerical, Categorical and Ordinal are the three flavors of data that we often come across. Numerical data is the data that has some quantitative measurement, for example, heights of the people, the time that a certain page takes to load. Statistical data is further dissected into two parts, Continuous data and Discrete data. Discrete data is integer based data which is, often, counts of individual events, like, the number of products purchased by a customer in a year, or the number of times heads was flipped when a coin was tossed. Continuous data is the type of data that can have an infinite number of possible values, this type of data has no defined set. For example, how much rain fell on a particular day or the time taken by the customer to log out from the app. Categorical data is qualitative data that has no inherent mathematical meaning. This type of data can be assigned a numeric value, but these numerical values will have no precise meaning. The examples of categorical data are Gender(Male, Female), Race, State of Residence. The last type of data is called as Ordinal Data. Ordinal data is a mixture of both Categorical and Numerical data. Ordinal data can be defined as the categorical data that has a mathematical meaning, for example, movie ratings between 1-5 is a type of an ordinal data as the numerical values can be compared and certain conclusions like a movie with a rating 3 are better than a movie with a rating 2.

Decision making is the basis of any machine learning algorithm. All the machine learning algorithms are focussed on predicting the values of dependent variables based on some other independent variables. The variables are broadly split into two categories, which are quantitative variables and qualitative variables. Quantitative variables are those variables that hold the Numerical data, may it be continuous or discrete. Qualitative variables are of a Categorical or Ordinal type. Machine learning algorithms are designed with the sole purpose of automating tasks and programming a machine to take decisions as the humans do. Using only quantitative parameters to take action in the machine learning world is an easier way to write these algorithms, but these algorithms somewhere compromise on the purpose that these algorithms had. As machine learning algorithms are designed to take decisions and predict things, the human way, the traditional, quantitative approach is not the appropriate way. People can make decisions on incomplete data and consider many qualitative parameters. Quantitative parameters are often clubbed with binary decisions which result in the loss of accuracy whereas qualitative parameters can significantly improve a decision-making process.

An effective way to collect qualitative data is taking unstructured interviews which will generate qualitative data by using open end questions. Interpreting qualitative data in itself is a big task. A lump of qualitative data is as good as a cluster of data that has no mathematical meaning. Such data has to be treated with care, as a lot of dependencies can be observed through them. Statisticians have to be notably creative and have to interpret this data in the right manner. Once the data is interpreted, and dependent variables and independent variables are separated, relative numerical values have to be assigned or computed depending on their type. Special care has to be taken for qualitative parameters that are categorical, for example, let us consider we are converting a qualitative parameter, Country of residence to numerical values, and we have a person living in country A and another living in country B, we cannot assign discrete random values like “1” to country A and “2” to country B as at no point in time we want to infer that country B is greater than country A. There special ways to treat categorical qualitative parameters, to avoid such a problem.

Qualitative parameters can be understood properly if they’re tried to understand the proper context. The context here in all ways is natural and not defined previously, and no context can be taken for granted. Inferences from qualitative parameters should be made by looking at the parameters from all the perspectives and understanding the story behind those parameters and why they have that value. These parameters are of no help to the person if the individual doesn't spend the time to understand them. The design of algorithms revolves around these parameters, and the algorithms will usually evolve with time. The inferences made out of the qualitative parameters is not a single output, it is subjective and will change if the same parameters are looked at with a different perspective.

If proper qualitative parameters are selected for designing the model and are used correctly, then a better accuracy rate can be guaranteed, as the designer spends a lot of time studying the input data and gets an in-depth knowledge of the problem they are tackling. Relations that are often missed while making these models are addressed when such parameters are introduced. Qualitative parameters have a vital role as they can be used to infer possible relations, outcomes, causes, effects, which are dynamic. These are the parameters that help the model to make decisions even if they're ambiguous. Sometimes, they can even draw two different relations or outcomes which are an accurate reflection of society. These parameters help us gain insight that only quantitative parameters cannot, all of this while devising a proper solution.

Qualitative parameters are useful in predicting the results with high accuracy but a lot of time and cost is associated to do a research based on these parameters. If the data set is not large enough, then there is a problem of validity, and hence the designer cannot rely on the data or the inferences made by these parameters. While interpreting the qualitative parameters, the designer is the one who plays the major role, and that is the reason why some other person cannot draw the same studies or same set of inferences. A lot of generalizations have to be made when dealing with qualitative parameters as there are some outlier cases which cannot be treated with equal importance, as the context which the designer is approaching may be very broad. The required time for extracting valuable information and making logical inferences from such data is very high and is not feasible if the deadline is short. Also, the designer needs to have expert knowledge in that field, and must be very observant and should not miss even a single detail.

Devising a model by considering an adequate amount of both types of variables, qualitative and quantitative can yield greater accuracy and better performance. Quantitative data can be achieved directly by performing some experiments or calculating values, and qualitative data and parameters can be achieved by field study and market study of the problem you are addressing.

3.1 USE OF QUALITATIVE PARAMETERS IN PLACEMENT PREDICTION SYSTEM:

Let us consider an example of a placement prediction system. This system predicts whether the student is going to be placed in a company after completing their education. We can tackle this problem in multiple ways. We can treat this problem as a binary problem, that is, we can predict 1 for a candidate being placed and 0 for a candidate not being placed. The other approach to this problem would be treating placement as a probability, that is, a number between 0 to 1 which would make more sense and an accurate representation of the scenario.

To make this system accurate, we must consider both qualitative as well as quantitative parameters. The quantitative parameters, in this case, would be the marks of the students throughout the course, the number of internships they have done, the credits they received after completing the entire degree, the number of languages they know, etc. But as we list down these parameters we can see that the list will be contained, and the problem would not be addressed with the intensity it has to be. While hiring a student, there are many more parameters that have to be considered.

So, for a better decision making, we now have to introduce qualitative parameters. Parameters like confidence, passion, determination, communication skills etc. can be of great help to determine whether the student will be placed or not. We cannot directly quantify these parameters as these parameters cannot be associated with a direct integer value. But, by designing an open end question set and making the students to answer the set, we can quantify these values and assign them a weight between 0 to 1.

After the parameters have been listed out, and the weights to the parameters have been assigned, we must model the solution by using the right machine learning algorithm that serves best to tackle this problem. We want to treat placement as a probability which lies between 0 to 1, and we have introduced qualitative variables. Therefore, we need to select an algorithm which deals with the fuzzy nature of our qualitative variables and the inconsistency of data and take the appropriate decision and make an accurate prediction. We have to eliminate the hard-computing models as those models work well with only quantitative variables.

A soft computing approach which works well with the qualitative parameters would be neural networks, as they are designed on the architecture of the human brain and take decisions, even if the data is incomplete or fuzzy. A neural network approach for this will also learn from the wrong predictions and re-learn and would increase its accuracy over time.

Van Heerden et al. [7] designed a system for placement prediction in the University of Pretoria Medical School, using artificial neural networks, which had 99 input parameters out of which 80 parameters were qualitative. They tested their system on some students and found out that for those students where all the parameters were available the prediction of the system was almost 100% accurate, and a 90% accuracy was obtained when only qualitative parameters were used. This shows that qualitative parameters used with the right algorithm can make decisions as good as humans if not better.

4. CONCLUSION AND FUTURE WORK:

Decision making is one of the root logics of machine learning, which is used to predict the possible outcome by making inferences from a given data set. Hence, input data plays a prominent role for successful predictions. In this paper, we've highlighted the importance of qualitative parameters for prediction purposes. Necessary methods are also explained to integrate the two types of parameters, qualitative and quantitative, to achieve a better outcome. These hybrid models can improve decision making concerning both the execution process, and the performance.

As future work, different ways to integrate qualitative and quantitative models for prediction purposes will be a great task that can be promoted in detail and verified. Also, with the help of psychological studies novel ways for interaction between different qualitative parameters can be found out which will result in better outcomes. This concept of hybrid modelling system can be used in various applications concerning decision making.

5. ACKNOWLEDGEMENT:
The authors thank Prof. Mahendra Patil for his support and guidance that has deeply influenced the quality of this manuscript.

REFERENCES:

1) Rutledge, G., Thomsen, G., Beinlich, I., Farr, B., Sheiner, L., & Fagan, L. (1989, November). *Combining qualitative and quantitative computation in a ventilator therapy planner*. In Proceedings of the Annual Symposium on Computer Application in Medical Care (p. 315). American Medical Informatics Association.
2) Kuipers, B. (1985, August). *The Limits of Qualitative Simulation*. In IJCAI (pp. 128-136).
3) Han, G., Santner, T. J., Notz, W. I., & Bartel, D. L. (2009). *Prediction for computer experiments having quantitative and qualitative input variables. Techno metrics*, 51(3), 278-288.
4) Clark, P., & Matwin, S. (1993). *Using qualitative models to guide inductive learning*. In Proceedings of the 1993 international conference on machine learning (pp. 49-56).
5) Simpson, G. S., Lamb, F. E., Finch, J. H., & Dinnie, N. C. (2000, January). *The application of probabilistic and qualitative methods to asset management decision making. In SPE Asia Pacific Conference on Integrated Modelling for Asset Management.* Society of Petroleum Engineers.
6) Bhatt, H., Mehta, S., & D'mello, L. R. (2015). *Use of ID3 Decision Tree Algorithm for Placement Prediction*. International Journal of Computer Science and Information Technologies, 4785-4789.
7) Van Heerden, B., Aldrich, C., & Du Plessis, A. (2008). *Predicting student performance using artificial neural network analysis. Medical education*, 42(5), 516-517.
8) Fleming, A. R., Phillips, B. N., Kaseroff, A., & Huck, G. E. (2014). *A qualitative study of job placement provider decisions in vocational rehabilitation*. Rehabilitation Counseling Bulletin, 58(1), 7-19.
9) Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). *Effect size and power in assessing moderating effects of categorical variables using multiple regressions: a 30-year review*.
10) Rajguru, A. A., & Apte, S. S. (2012). *A comparative performance analysis of load balancing algorithms in distributed system using qualitative parameters.* International Journal of Recent Technology and Engineering, 1(3).
11) Qian, P. Z., & Wu, C. J. (2008). *Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. Techno metrics*, 50(2), 192-204.
12) Qian, P. Z. G., Wu, H., & Wu, C. J. (2008). *Gaussian process models for computer experiments with qualitative and quantitative factors. Techno metrics*, 50(3), 383-396.
13) Conti, S., & O'Hagan, A. (2010). *Bayesian emulation of complex multi-output and dynamic computer models.* Journal of statistical planning and inference, 140(3), 640-651.
14) Hocaoğlu, M. F. A *Comparison between Qualitative Simulation and Traditional Simulation: Bridging the Conceptual Gap.*
15) Yan, X., Zhou, Y., Wen, Y., & Chai, X. (2013). *Qualitative and quantitative integrated modeling for stochastic simulation and optimization.* Journal of Applied Mathematics, 2013.

16) Simpson, G.S., Finch J.H. and Lamb F.E.: 1999.: "*Risk, uncertainty and relations between "strategic" and "within strategy" decision making in the upstream oil and gas industry"* In: Proceedings of the 4th international conference on the dynamics of strategy, University of Surrey (1999) 335.
17) Lamb, F. E., Simpson, G. S., & Finch, J. H. (1999). *Methods for evaluating the worth of reserves in the upstream oil and gas industry*. Geopolitics of Energy, 4(22), 2-7.