

Refining & Clustering of Parameters for Predictive Analysis

How to filter factors for study!

¹Gagan Deep Bakhshi

¹Student

¹Computer Science and Engineering

¹CMR Technical Campus, Hyderabad, India

Abstract –Predictive Analysis is gaining popularity irrespective of the field of applicability. The vision for the future can be matched by tracing the present and the past of the system. Factors are chosen on which the focus is laid, and the analysis of those factors yields a bigger picture. But how these factors are to be selected and how efficient they should be is what this paper describes.

Index Terms – Parameters-Factors.

I. INTRODUCTION

Predictive Analysis is the process of analyzing the present state of the system as a whole with some reference to its past, and determining the future state the system will assume. This sounds very fascinating that we can determine the future state of a system, but how efficient is this process, and what factors it relies on? Even though we may focus on certain factors at some instance of time but who decides what those factors are, and what happens if the factors in focus are replaced by other factors. All these questions are very fascinating to answer and to analyze. This paper covers their description.

II. FACTOR BASED ANALYSIS

A factor if to be defined is the unit of analysis process. Obviously, the factors are process independent hence, factor analysis itself is a crucial process. What makes a factor? This question strictly depends on the parameters the system relies on. If I say about an analysis of a political system and decision making, then some of the possible factors can be the Age of the politicians, the educational background of the politician, the period of power for a party, the GDP of the nation, the tax related hiccups and many more. This example has been quoted so that one may realize that factors are an ocean in an analysis process but factors to focus are what collective form the fresh water.

III. SELECTING FACTORS FOR ANALYSIS

The selection of the factors for analysis is what decides the efficiency of the predictive analysis and the context of analysis.

The factors for factor selection can be:

- *The Relativity to the Context of Analysis
- *The Variation Rate
- *The Life of the Parameter in the System
- *The Parametric Conflict

For how to select the parameters we focus on the following example quoted:

EXAMPLE:

Let us consider the system of Weather Analysis. The possible set of parameters can be huge, but will include:

- *Temperature
- *Humidity
- *Clouds
- *Visibility
- *Sun

Now let us focus on how and when to select what parameters form this system.

If we are trying to predict for Rain let us suppose, then CLOUDS are of utmost concern, since dense black clouds signify rain, Next goes with the humidity, when the HUMIDITY is more there are more chances of raining rather than in a dry weather.

So our filtered category of factors includes:

- *Clouds
- *Humidity

We are not neglecting the other factors completely but we know that we need to focus on these factors. Let us create a matrix of focus for the parameters of selection as we described above:

MATRIX OF FOCUS

ANALYSIS	RELATIVITY(1-5)	VARIATION(1-5)	LIFE(1-5)	CONFLICT(1-5)
Temperature	3	1	5	1
Humidity	4	3	3	1
Visibility	1	3	3	1
Clouds	4	4	4	1
Sun	2	1	5	1

In the above matrix, we have given a score to the parameters based on the rule of selection we gave above. when we talk about the relativity, it is strictly with respect to the context of analysis which in our case is rain. Variation describes the degree of change in the factor, Life on the other hand is the period for which the factor holds valid in the system, and Conflict is the miscellaneous criteria which shows the degree of conflict with the other parameters i.e it changes with respect to the other parameters in the system under consideration (Focus Factors).

The more the score of the Factor we choose it for analysis. Since our factor chart above involves a score of 11 for Humidity and 13 for Clouds hence we choose these factors.

Note: The Scoring is Strictly based on the context and the analysis, the score of 1 we gave for CONFLICT for all is to give the essence of change of each factor, this score will change with respect to the actual measurements.

IV. CLUSTERING THE FACTORS

Clustering as we know is the process of grouping some objects such that the objects within the same cluster are closer in meaning to each other whereas unrelated to the objects of the other cluster. Now is the time to unleash the role of clustering in this process.

We may encounter certain anomalies in terms of the context of analysis and the variation of factors. Hence the degree of variation is something on which we can actually form the clusters. The parameters which can change drastically over a period of time and those which remain somewhat subtle can be formed into different clusters as shown:



The Sun and Temperature are relatively lesser changing.

The Humidity, Visibility, Cloudy nature is of more variable concern in the context.

This clustering helps in reanalyzing the factors for an analysis like the original.(Further Future Analysis)

We may further use these CLUSTER FORMATTING to categorize these parameters we need to focus. Like clustering for Relativity, Clustering for Life in the system etc.

Hence, Clustering helps in our concern of Factors.

V. SUMMARIZING THE ENTIRE RESEARCH AND APPLICABILITY

*Clustering has been used so as to separately group the parameters on the basis of some property.

*The Matrix of Focus is open for addition or removal of properties of focus.

*These criteria(s) can also successfully work for analyzing time series data.

*The Time Series data has the property of indefinite variation and unordered changes with respect to each measurements encountered over a time period.

Time series data can also be clustered with respect to a parameter of focus based on the above rules. If we are measuring the time series data in the form of Glacier melts or ice melts then we can focus on factors such as the geography of the area under focus, the pollution rate, the temperature variation, the human intervention etc.

Note: Time series data is not considered easy for analysis due to its irregular changes. But clustering has a solution for that as well,

As we can actually sub divide the period of changes for the parameters and compare them.

*Refining of factors is the selection of factors to focus from the available pool of factors.

*The research is oriented for a large pool of systems keeping in mind that predictive analysis is not an ignorant approach. A single point missed could lead to the entire result as a fallacy.

*This research hence gives an effective way for measuring and representing factors of analysis which will contribute in an efficient predictive analysis or analysis as a whole.

VI. ACKNOWLEDGMENT

The Entire Research work has been possible with the proper guidance of **Mrs. Suwarna Gothane** Associate Professor, CSE Dept ,CMR Technical Campus. Further I would like to thank the Head of the Department (HOD) of CSE **Prof. K. Srujan Raju** for his encouraging approach which has lead to this research paper being successful.

REFERENCES

[1] Data Mining –Concepts and Techniques- Jiawei Han & Micheline Kamber, Morgan Kaufmann publishers, Elsevier, 2nd Edition 2006

