Estimation of Well-being and Pattern Recognition of Causal Influence: A study through Big Data Analysis and Machine Learning

SUGATA SEN[1] ADITI CHAKRABORTY[2]

[1] *Associate Professor of Economics, Panskura Banamali College (Autonomous), sensugata@gmail.com*
[2] *Faculty, UG Department of Computer Science and Applications, Panskura Banamali College (Autonomous).*

Abstract— Using of Euclidean distance to measure the relative importance of causal variables can play an important role in the analysis of well-being. The use of Cartesian algebra along with machine learning can help to strengthen the traditional well-being measurement procedure. If the preferences of the domains are considered as a database, the frameworks developed by the big data can measure well-being for each individual.

I. INTRODUCTION

It is thought that well-being is synonymous to happiness. But actually well-being leads to human happiness. There are many measures of happiness and well-being. This work wants to deduce the level of well-being using a set of indicators so that the level of happiness can be properly estimated. To that respect this work wants to locate the causal variables affecting the composite level of well-being and the pattern of variation in causal variables to explain the variation in composite well-being. To that respect the help of Cartesian algebra and machine learning has been taken.

II. LITERATURE REVIEW

Sap et.al. (Sap, et al., 2016) have tried to define well-being through life satisfaction. To that respect they have used Social networking data to using the Cascading model. They found the message reply pattern is an indicator of well-being. Wilckens et.al. (Wilckens & Hall, February 2015) have used machine learning to estimate structural dependency of Subjective Well-being. Jaques et.al. (Jaques, Taylor, Azaria, Ghandeharioun, Sano, & Picard, 2015) have differentiated between happiness and unhappiness through using Gaussian mixture model by comparing

variety of machine learning and feature selection technique. They have also tried to evaluate the relationship between different components of well-being including happiness, health, energy, alertness and stress. Addabbo et. al. (Addabbo, Tommaso, & G.Facchinetti, 2004) have proposed in their paper the extent statistical and fuzzy techniques can be used to measure functioning with special reference to child well-being. For this purpose he makes a model to generate a structural equation. For application purpose of this model he used a data set based on a ISTAT. This data set contain the information of education of children, child care of relatives and parents and the socio-demographic structure of children. In our paper we have used a data set of indicator and our target is to find the causal effect of these indicators on well-being. He establishes that the statistical and fuzzy techniques can be applied to calculate child well-being. Presscott_Allen (Prescott_Allen, 1997) found in this book we know that measuring of communicating a society's well being and progress towards sustainability. Going to do these work here combining the various indicators and as the result he explains the condition of people and the eco system and the effects of people-ecosystem interactions. In this paper the condition of people are expressed as an index of human well-being and ecosystem are expressed as an index of well-being. Then the interaction of two points provides a reading of overall well-being and due to that progress toward sustainability.Juniper.et.al (Juniper, White, & Bellamy, 2016) proposed that the measurement of employee well-being will be defined by comparing Factor Analysis (FA) and Impact Analysis (IA), where collect the raw dataset of FA and IA. They found the score of impact analysis on the basis of role, sector,

level of different employee. The importance of these score and troublesome is known as well-being. This score is measured on five point scale where '1' ="not at all importance" and '5'=extremely importance .In our proposed work we will compare the performance of each individual on the basis of each indicator with respect to threshold value of Tj and determine on state and of state.

Hypothesis
- As an alternative to the traditional variables many new variables can influence human well-being.
- Euclidean distance can be used as an indicator of level of influence.

Modelling

This work finds the causal effects of independent indicators on the composite well-being and also finds the internal relationship between these indicators.

Let here n individuals are denoted by I={1,2,…..,n}.There are also d domains comprising m explanatory variables expressed as j={1,2,…..,m}. Thus an M=n*m matrix is formed where each cell explains the achievement of the ith individual on the jth variable. This performance η_{ij} as depicted in the said cells are enumerated through a threshold value T_j .Now another matrix B=n*m is formed m where each cell ρ_{ij} is either 1 or 0.Thus $\rho_{ij} = 1\ where\ \eta_{ij} \geq T_j$ otherwise $\rho_{ij} = 0$.

Now an n dimension column W vector is created where each element of that vector w_i is the well-being score of ith individual.

Thus $w_i = \sum_{j=1}^{m} \rho_{ij}$

So W={ w_i | $where\ w_i = \sum_{j=1}^{m} \rho_{ij}$ } .

To find the causal effects and the patterns of causation the techniques of Big Data Analysis, Data Mining and crowd sensing have been used in details here. Association rule of machine learning is used. For this purpose we have to calculate the value support, confidence, Lift. Here Support contains a set of fraction of indicators. Confidence contains probability of occurrence of well-being index given indicators. Lift represents the ratio of confidence to baseline probability of occurrence of well-being index. At last,

our target is to find that causal effect in different indicators on the composite well-being. Here association rules under data mining play active roles.

Now as per association rule we define Support ,Confidence and Lift to find the internal relationship of each individual with respect to there different indicators({I}) and also find the effect of each indicator on {W}.
Support ({j}$\to w_i$ } = (Indicator containing both j and w_i)/(Total number of indicator).
Thus, with the help of above mathematical expression we can be able to find the causal effects of independent indicators on the composite well-being.

Let S is the column vector where, S={s_i} and i=1,2,……..,n. Here $s_i = \sum_{j=1}^{m}$ on state/m. This S can be defined as the Support vector of individual well-being. $\sum_{i=1}^{n} s_i = \alpha$ can be termed as the overall level of well-being within the society and α=α/n is the average level of well-being within the society.

As we have m indicators of well-being we can get $\frac{m!}{(m-1)!}$ combinations of m-1 indicators. Let the set of these combinations is C with c combinations i.e, c= $\frac{m!}{(m-1)!}$.

Now we can get c column vectors where each of those vectors denoted as $\hat{S}c$ having n elements.

Now let us assume that H is the Euclidean distance between the column vector S and the adjusted column vectors $\hat{S}c$ .So, we have a set of c Euclidean distances denoted as D where D={$h_1,h_2,…..,h_c$ }. Thus, the elements of D are nothing but the Euclidean distances between S and $\hat{S}c$. Thus higher the importance of a column vectors, higher will be it's effect on the Support vectors. So withdrawl of that column vector will create a larger difference between the corresponding adjusted Support vectors and the original Support vectors. Thus higher the Euclidean distance between these two vectors higher will be the effect of that column vector or the indicator on the composite level of well-being. In this way we can generate a pattern among the causal indicators to

determine the partial influence of each of the causal variables.

To measure the extent of influence of each variable on the set of individuals this work has further taken the help of Euclidean distance and the coefficient of variation about Mean of each variable. To that respect a matrix Z is on from matrix M where each cell of Z is 0 if $\eta_{ij} \geq T_j$. Otherwise it is θ_{ij} where $\theta_{ij=} |T_j - \eta_{ij}|$.

Now, the left CV_j is the coefficient of variation of jth variable. This CV_j can be used as an indicator of the effort needed to achieve well-being to that variable.

CONCLUSION

Applying this methodology on the different individual ,we can be able to find the causal effect of different indicators on the composite well being index and the relative effort needed to achieve well-being on that variable. The above technique will show a new way for research to find the inter relationship between different indicators and well-being.

REFERENCES

[1] Addabbo, T., Tommaso, M. L., & G.Facchinetti. (2004). To what extent Fuzzy Set theory and structural equation modelling can measure functionings? An application to child well being. *Materiali di Discussione del Dipartimento di Economia Politica* , 468.

[2] Jaques, N., Taylor, S., Azaria, A., Ghandeharioun, A., Sano, A., & Picard, R. (2015). Predicting students' happiness from physiology,phone, mobility, and behavioral data. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 222-228). Institute of Electrical and Electronics Engineers (IEEE).

[3] Juniper, B., White, N., & Bellamy, P. (2016). Assessing employee well-being:is there another way? *International Journal of Workplace Health Management , 2* (3), 220-230.

[4] Addabbo, T., Tommaso, M. L., & G.Facchinetti. (2004). To what extent Fuzzy Set theory and structural equation modelling can measure functionings? An application to child well being. *Materiali di Discussione del Dipartimento di Economia Politica* , 468.

[5] Jaques, N., Taylor, S., Azaria, A., Ghandeharioun, A., Sano, A., & Picard, R. (2015). Predicting students' happiness from physiology,phone, mobility, and behavioral data. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 222-228). Institute of Electrical and Electronics Engineers (IEEE).

[6] Juniper, B., White, N., & Bellamy, P. (2016). Assessing employee well-being:is there another way? *International Journal of Workplace Health Management , 2* (3), 220-230.

[7] Prescott_Allen, R. (1997). *Measuring and communicating well-being and sustainable development.* Cambridge: IUCN.

[8] Sap, M., L.Kern, M., C.Eichstaedt, J., Kapelner, A., Agrawal, M., Blanco, E., et al. (2016). PREDICTING INDIVIDUAL WELL-BEING THROUGH THE LANGUAGE OF SOCIAL MEDIA. *Pacific Symposium on Biocomputing , 21*, 516-527.

[9] Wilckens, M., & Hall, M. (February 2015). Can Well-Being be Predicted?A Machine Learning Approach. *SSRN Electronic Journal* .