International Journal of Engineering & Technology, 7 (3.36) (2018) 164-166

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Exploring Subjective Well-Being Factors with Support Vector Machine

Du Ni[1], Ming K. Lim[2,3]*, Zhi Xiao[1], Xiaodong Feng[1]

1School of Economics and Business Administration, Chongqing University, Chongqing 400044, PR China
2College of Mechanical Engineering, Chongqing University, Chongqing 400044, China
3Centre for Business in Society, Coventry University, Coventry, UK
**Corresponding author E-mail: ming.lim@cqu.edu.cn*

Abstract

National-level Subjective well-being (SWB) is known to be associated with six traditional factors, that is, GDP per capita, social support, healthy life expectancy, social freedom, generosity, and absence of corruption, but debates persist about the variability of these six factors. Whether the predicting in SWB is based only on these six factors or not? Are there any country-specific factors? Thus, we examined these two questions with the data sets from World Happiness Report and OECD database for U.S.. By control-ling the other factors except only one, we employed Support Vector Machine (SVM) to identify the weight of each factor without worrying about the limitations caused by the small size of the sample in years. We found that another three factors (protein con-sumption, fruit consumption and physician ratio among all employee) in addition to the six traditional factors can also affect na-tional-level SWB in U.S.; Moreover, the power of SVW in prediction is as high as 95.08%, which is much greater than that present-ed by the linear regression models (74.3%).

Keywords: *Use about five key words or phrases in alphabetical order, Separated by Semicolon.*

1. Introduction

Well-being (SWB) is an important issue in the United Nations Millennium Project[1]. Multiple factors affecting SWB have been continuously figured out since the 1970s. The happiness reports released by the Gallop World Company recommended the life ladder as a 0-10 scale questionnaire for the measurement of SWB [2]. The reports provided the policy makers with valuable reference in every country in the world. However, even these reports possessed high reliability and validity as the Gallop World Company claimed, the researchers had no consensus on the quantifying methods of the national-level SWB, and the arguments on the effective factors related to SWB have never been stopped. Thanks to the characteristics of SWB, just as its name suggested, SWB is subjective, dynamic, and might be affected by many possible factors. Thus, predicting the national-level SWB becomes one of the most difficult issues in the United Nations Millennium Project.

2. Literature review

The prediction for national-level SWB was first studied more than 20 years ago by Ed Diener in 1995 [3]. Diener first carried out an investigation on SWB in terms of several factors affecting national-level SWB like income growth, social comparison, equality, independence-interdependence, and cultural homogeneity over more than 55 countries. In 2000, Diener proposed to build a happiness index to quantify the national-level SWB specifically for U.S..[4]. Because the cultural and societal factors might lead to international differences, the methods to explore SWB were specified into two types in 2003, that is, the personal-level and national level data-based [5]. In 2012, Diener et al. re-claimed the national-level SWB based on the assumption that worldwide predictors for SWB such as social support and fulfillment of basic needs had been uncovered, and there might be large differences in SWB among different societies [6]. In the same year, the Organization of Economic Cooperation and Development (OECD) issued the guidelines for implementing national measures of SWB [7], which could combine the SWB at both the personal-level and national-level.

Of course, multiple factors besides those six mentioned above have been continuously examined. A negative effect of urban air pollution (in terms of nitrogen, sulphur, particulates) on SWB was identified by Welsch (2002) with a cross-sectional data from 54 countries [8]. Welsch also found in 2006 that a considerable monetary value was associated with improvements in air quality in Europe between 1990 and 1997 ($750 per capita per year for nitrogen dioxide and $1400 per capita per year for lead, on average) [9]. In addition to air pollution and monetary values, Costanzaa et al. (2007) disclosed that food structure, shelter or leisure were SWB related [10]; Clark and Oswald (1994, 1997) claimed that subject unemployment rate was higher than loss of income for predicting SWB at national level [11,12]; Lang et al. (2007) found that moderate alcohol consumption was more associated with better cognition than abstinence in older people [13]; McCann (2010) held that well-being and smoking prevalence were negatively interacted [14].

However, the previous studies reviewed had some gaps: (1) The correlation of SWB between specific countries were so diverse, for the correlation was mainly established on the means of the samples in these countries. But for the data specific to a certain country, the sample size was limited with the linear regression models not passing the Pearson correlation test. (2) Most of the models for predicting SWB were based on multiple linear regression (MLR). These models have different numbers of input variables, which would produce a much wider range of accuracy and

precision; (3) Obvious factors that might affect SWB have not been taken into consideration within one systematic framework, e.g. food, air pollution, working time. To fill in these gaps, we would apply Support Vector Machine (SVM), one of the machine learning methods, which requires no perquisite test for the samples size, to improve the accuracy of the data sets for prediction.

3. Methods

The traditional way for predicting national-level SWB was done first by a survey among multiple nations around the world, and then by the analysis of the correlated factors based on the world averages with linear regression models, and finally by adding up all the products of the corresponding difference and explanation power.

The basis of linear regression models is the least-squares problem, which is an optimization problem with no constraints (i.e., m = 0) and an objective which is a sum of squares of terms of the form $a_i^T x - b_i$:

$$\text{Minimize} \int 0(x) = \sum_{i=1}^{k} (a_i^T x - b_i)^2 \quad (1)$$

The least-squares problem may be limited by the sample size because the solution of a least-squares problem can be reduced to solving a set of linear equations, $(A^T\mathrm{A})\mathrm{x} = A^T b$, sometimes there will be no solutions for $(A^T\mathrm{A})^{-1}\mathrm{x} = A^T b$ when we have more features than the number of years involved.

Besides the problems mentioned above, we might encounter the following limitations in applying the traditional way for predicting the national-level SWB: (1) some related factors might be ignored. For example, in the happiness reports released by Gallop World Company, only 6 factors could explain a national-level SWB, which were far away from referring to all the factors involved [15]. As reported by Costanza et al. [10] and Cummins [16], the environment, air quality and the medical level might also increase the national-level SWB. Thus, a few researchers insisted that some kinds of factors would surely be ignored by the others. (2) If one factor was considered to be linearly correlated with the others, e.g. national-level SWB was correlated with GDP in one country, when you took another factor like social support into consideration in this country, the relationship between these factors might become non-linear; and (3) The time period recorded for the SWB of a specific nation was generally relatively short (about 10 years) and needed the help of world average level to ensure their correlations. The country-specification would be weakened in such a process. Therefore, in order to solve these limitations, this paper would employ a non-linear regression model with the SVM.

The SVM was a model proposed by Vapnik [17] and was processed based on statistical learning theory which seemed a promising approach to modelling multivariate data. The SVM is actually a learning system that uses a hypothesis space of linear functions in a high-dimensional feature space, trained with a learning algorithm from the optimization theory [18]. Unlike the traditional linear regression models that try to minimize the errors in the training data, the SVM attempts to minimize the upper bound on the generalization errors based on the principle of structural risk minimization [19], which has been found, in several cases, superior to the traditional linear regression models. The core formula of the SVM is as follows:

$$f(x) = \sum_{i=1}^{m} a_i y_i k(x, x_i) + b \quad (2)$$

where f(x) is the sum of kernel functions k(·,·).

The SVM only examines the linear relationship between f(x) and kernel function. There is no collinear requirement for the samples inside the kennel functions. This advantage of the SVM would permit us to include more features than the actually observed because the kennel functions could map the data into an arbitrary dimension without any prerequisite tests for the multicollinearity. Also, the SVM is trying to solve the problems with 'support vectors'[20], not all the data, which is suitable for those data sets with a small sample size (less than 30).

This paper, in a nutshell, would mainly focus on the six traditional key factors appeared in the World Happiness Report, 2018 [15]. These factors are GDP per capita, social support, healthy life expectancy, social freedom, generosity, and absence of corruption. Additionally, some apparently uncorrelated factors would be added to expand the width of the data sets: environmental pollution, medical level, food structure, and working hours from OECD database. With these factors included, we could first see which variable would make the biggest contribution to SWB, and then compare the power of the SVM in predicting with that of the traditional linear regression methods. To be specific, two experiments would be carried out in this paper.

3.1. Experiment 1

We would replicate the analysis of the linear regression methods according to World Happiness Report, 2018, by defining the predicted changes in happiness based on the six traditional factors:

Step 1. To take periodical averages (2006-14 and 2015-17, respectively) of the six factors from the data sets;

Step 2. To explore the differences between the two periods for each specific factor;

Step 3. To multiply the differences with corresponding coefficients for a target factor;

Step 4. To take the summation of the products from the previous steps. The resulted summation was the predicted change in a ladder due to changes in the six factors. According to the introduction of World Happiness Report, 2018, the data were obtained by the means correlation of 137 nations. In addition, we calculated the results of the US with the trained SVM model for the happiness by adding the width of data sets with environmental factors. By so doing, we could cut down the sample size and examine the uncorrelated factors of SWB at the same time. To make a comparison with the US., we employed the results of UK.

3.2. Experiment 2

There were six steps involved in this experiment:

Step 1: to gain the trained SVM on training samples with the true values for the 30 factors (from 2006-2014) as the input, with the life-ladder point (measurement for SWB) for U.S. (from 2015-2016) as the output;

Step 2: to change the first factor (percentage of population exposed to more than 10 micro-grams/m3) by adding the standardised values (the minimum and maximum values between 0 and 100) by 10, while the values of the other 29 factors were kept unchanged. Then, we got the new simulated data sets for the first factor;

Step 3: to put the new simulated data set into the SVM to get the predicted life ladder points;

Step 4: to subtract the simulated predicted life-ladder points by the absolute values of the true life-ladder points as its corresponding identified weight;

Step 5: to repeat the above-mentioned steps for the other 29 factors;

Step 6: to draw the weight plot for all the 30 factors of each country investigated. All paragraphs must be justified alignment. With justified alignment, both sides of the paragraph are straight.

4. Results

First, we compared the predicted values of SWB offered by the SVM (as shown in Figure 1), the y-axis indicates the life ladder values. We can see that the predicted values by the SVM are much closer to the true values than those offered by the linear regression models.

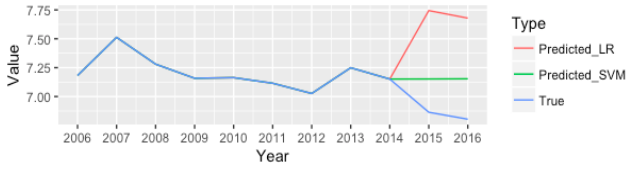

Fig. 1: Predicted values and the true values

Actually, according to World Happiness Report, 2018, the prediction power of their model was as high as 74.3%, while that of the SVM was 95.08%. This justified the great superiority of the SVM to the traditional methods in predicting SWB.

Then, the results obtained were to be tested in a broader view in Figure 2. By adding the data of UK into the data sets, as indicated in Figure 2, it could be proved that different countries might have different effective factors on SWB.

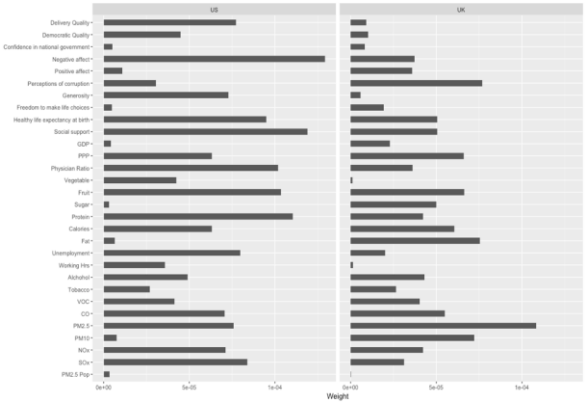

Fig. 2: Weights of different factors

We first chose the top-three of the most weighted factors for each country tested in addition to the 6 traditional factors, and then did a Pearson correlation test among these factors. The top-three factors included protein consumption (r = 0.65, P <0.05), fruit consumption (r = 0.63, P <0.05), and physician ratio among all employed (r = 0.63, P <0.05). The results obtained suggested that the factors selected by the SVM could reveal the linear relationships between some human-behavior factors with SWB for the country tested. These factors could not be disclosed by the traditional linear regression models on the whole world level.

5. Conclusions

To our best knowledge, this paper is the first to apply the SVM to the study of SWB. With the machine learning methods, specifically the SVM, we could make up the factors ignored in World Gallup Reports, that is, besides the six traditional factors on SWB, we found the other three factors being also effective on SWB: fruit consumption, physician ration among all employee, and the protein consumption. There are also two contributions of this work for the studies on SWB: (1) The national-level SWB can be predicted with a higher accuracy by the country-specific data if taking the differences of each country into consideration; (2) More variables can be added into the model for prediction to simulate the true situation of the world. Hopefully, the contributions made in this paper should have practical implications to the application of machine learning methods in the studies of human behaviours, and to the governmental authorities in making corresponding policies.

Acknowledgement

This work was supported by the National Natural Science Foundation of China [grant numbers 71671019]; The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Sachs, Jeffrey D., and John W, McArthur, "The millennium project: a plan for meeting the millennium development goals", *The Lancet*, Vol-365-9456, (2005), pp.347-353.

[2] Helliwell, John F., Richard Layard, and Jeffrey Sachs. World happiness report [2012]. (2012).

[3] Diener, E., Diener, M. and Diener, C., "Factors predicting the subjective well-being of nations", *Journal of Personality and Social Psychology*, Vol 69-5, (1995), pp.851-864.

[4] Diener, E., "Subjective well-being: The science of happiness and a proposal for a national index", *American Psychologist*, Vol-55-1, (2000), pp.34-43.

[5] Diener, E., Oishi, S. and Lucas, R., "Personality, Culture, and Subjective Well-Being: Emotional and Cognitive Evaluations of Life", *Annual Review of Psychology*, Vol-54-1, (2003), pp.403-425.

[6] Diener, E., Oishi, S. and Lucas, R., "National accounts of subjective well-being", *American Psychologist*, Vol-70-3, (2012), pp.234-242.

[7] OECD-Total. (2013). Quarterly National Accounts, 2013(3), pp.310-311.

[8] Welsch, H., "Preferences over Prosperity and Pollution: Environmental Valuation based on Happiness Surveys", *Kyklos*, Vol-55-4, (2012), pp.473-494.

[9] Welsch, H., "Environment and happiness: Valuation of air pollution using life satisfaction data", *Ecological economics*,Vol-58-4, (2006), pp.801-813.

[10] Costanza, R., Fisher, B., Ali, S., Beer, C., Bond, L., Boumans, R., Danigelis, N., Dickinson, J., Elliott, C., Farley, J., Gayer, D., Glenn, L., Hudspeth, T., Mahoney, D., McCahill, L., McIntosh, B., Reed, B., Rizvi, S., Rizzo, D., Simpatico, T. and Snapp, R. "Quality of life: An approach integrating opportunities, human needs, and subjective well-being", *Ecological Economics*, Vol-61-2.3, (2007), pp.267-276.

[11] Clark, A. E., & Oswald, A. J., "Unhappiness and unemployment" *The Economic Journal*, Vol-104-424, (1994), pp.648-659.

[12] Oswald, A. J. "Happiness and economic performance", *The economic journal*, Vol-107-445, (1997), pp.1815-1831.

[13] Lang, I., Wallace, R. B., Huppert, F. A., & Melzer, D., "Moderate alcohol consumption in older adults is associated with better cognition and well-being than abstinence", *Age and ageing*, Vol-36-3, (2007), pp.256-261.

[14] McCann, S. "Subjective well-being, personality, demographic variables, and American state differences in smoking prevalence", *Nicotine & Tobacco Research*, Vol-12-9, (2010), pp.895-904.

[15] Sachs, Jeffrey D., Richard Layard, and John F. Helliwell. World Happiness Report 2018. No. id: 12761. 2018.

[16] Cummins, R. A., "Personal income and subjective well-being: A review", *Journal of Happiness Studies*,Vol-1-2, (2000), pp.133-158.

[17] Vapnik, V., "The nature of statistical learning theory", *Springer science & business media*, (2013), Vol-114-434, (1994), pp.645-669.

[18] Chen-Chia Chuang, Shun-Feng Su, Jin-Tsong Jeng and Chih-Ching Hsiao., "Robust support vector regression networks for function approximation with outliers" *IEEE Transactions on Neural Networks*, Vol-13-6, (2002), pp.1322-1330.

[19] Hong, W. C., Dong, Y., Chen, L. Y., & Wei, S. Y., "SVR with hybrid chaotic genetic algorithms for tourism demand forecasting" *Applied Soft Computing*, Vol-11-2, (2011), pp.1881-1890.

[20] Syam, N., & Sharma, A. "Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice", *Industrial Marketing Management*, Vol-6-9, (2018), pp.135-146.