Analysis of Hospitalized Pathways Using Locally Weighted Learning

Yong Gyu Jung, Jae Hong Lee, Young Jin Choi

Abstract: ***In recent years, data mining and treadmill machines have become an issue in various fields. Data mining is the systematic and automatic discovery of metabolic rules and patterns within large-scale stored data. Among data analysis algorithms in data mining and machine learning, there are bifurcation / clustering algorithms. In this study, we use the LWL algorithm among classification algorithms. Identifying these various admission pathways and finding out the weight of various hospitalization pathways can develop toward the response and quality of medical care. Therefore, it is classified into LWL algorithm to predict various hospitalization routes and specific gravity.***

Index Terms: ***LWL, linear regression, locally weighted learning, machine learning and hospitalization pathways***

I. INTRODUCTION

In recent years, data mining and treadmill machines have become an issue in various fields. Data mining is the systematic and automatic discovery of metabolic rules and patterns within large-scale stored data. In other words, it is called knowledge-discovery in databases (KDD). Data mining is mainly used for exploring predicting future information to support decision making or for and modeling useful relationships that have not been previously known in vast amounts of data through processes such as data refinement and summary, pattern and rule discovery, and extraction. Data mining is used in a variety of fields and is used in military, information, security, medical, business, etc. Machine Mining is a technology that predicts the future by analyzing big data. Collecting and analyzing data to predict the future is similar to big data analysis, but there is a difference in that the computer itself can collect and confirm vast amounts of data. Among the data analysis algorithms in data mining and machine learning, there is a classification / clustering algorithm. Medicine and biomedical sciences have become data-intensive fields, which, at the same time, enable the application of data-driven approaches and require sophisticated data analysis and data mining methods. Biomedical informatics provides a proper interdisciplinary context to integrate data and knowledge when processing available information, with the aim of giving effective decision-making support in clinics and translational research.

In this study, we use the LWL algorithm among the classification algorithms. LWL (Locally weighted learning) is simple but appealing, both intuitively and statistically. And it has been around since the turn of the century. When you want to predict what is going to happen in the future, you simply reach into a database of all your previous experiences, grab some similar experiences, combine them and use the combination to make a prediction, do a regression, or many other more sophisticated operations. We like this approach to learning, especially for learning process dynamics or robot dynamics, because it is very flexible (low bias) so provided we have plenty of data we will eventually get an accurate model.

II. RELATED RESEARCH

A. LWL (Locally weighted learning)

LWL is a class of function approximation techniques, where a prediction is done by using an approximated local model around the current point of interest. The goal of function approximation and regression is to find the underlying relationship between input and output. In a supervised learning problem training data, where each input is associated to one output, is used to create a model that predicts values which come close to the true function. All of these models use complete training data to derive global function. However, a disadvantage of global methods is that sometimes no parameter values can provide a sufficient good approximation. Also the computational costs are also very high in such cases. An alternative to global function approximation is Locally Weighted Learning. LWL methods are non-parametric and the current prediction is done by local functions which are using only a subset of the data. The basic idea behind LWL is that instead of building a global model for the whole function space, for each point of interest a local model is created based on near neighboring data of the query point.

As bellow example, Locally Weighted Learning containing in the upper graphic the set of data points (x,y) (blue dots), query point (green line), local linear model (red line) and prediction (black dot).

The graphic in the middle shows the activation area of the model. The corresponding weighting kernel (receptive field) is shown in the bottom graphic.

B. Memory-Based Locally Weighted Regression

Locally Weighted Regression (LWR) is the classic approach to solve the function approximation problem locally [2]. It is also called Memory-Based Learning, because all training data is kept in memory to calculate the prediction. The single steps of LWR are outlined in algorithm

Revised Manuscript Received on December 22, 2018.

Yong Gyu Jung, Dept. of Medical IT, Eulji University, Korea.
Jae Hong Lee, CEO, Ubivelox Co, Ltd., Korea.
Young Jin Choi, Dept. of Healthcare Management, Eulji University Korea.(E-mail: yuzin@eulji.ac.kr)

Table. 1 Process of Locally weighted learning

Algorithm 1 Memory-Based Locally Weighted Regression

Given:

- query point x_q
- n training points $\{x_i, y_i\}$

Prediction:

- Build matrix $X = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_n)^T$ where $\hat{x}_i = [x_i^T \quad 1]^T$
- Build vector $y = (y_1, y_2, ..., y_n)^T$
- Compute diagonal weight matrix W:

$$w_{i,i} = \exp\left(-\frac{1}{2}(x_i - x_q)^T D(x_i - x_q)\right)$$

- Calculate Regression coefficient:

$$\beta_q = (X^T W X)^{-1} X^T W y$$

- Predict

$$\hat{y}_q = [x_q^T \quad 1]\beta_q$$

C. *Locally Weighted Projection Regression*

Locally Weighted Projection Regression (LWPR) is a purely incremental LWL method. It was developed to solve two major problems that exists with memory-based methods like LWR. One of them are the cost intensive computations of LWR with high dimensional data that increase quadraticaly. This makes the algorithm unusable for tasks that need many predictions in small time steps. Another problem of LWR is that the matrix inversion in algorithm 1 for obtaining the regression coefficient cannot handle redundant input dimensions and can become singular. Instead of throwing away the model after each prediction like in LWR, LWPR is keeping each model for further predictions in memory. It uses multiple locally weighted linear models which are combined for approximating non-linear functions. Adding new data points requires only an update to the existing models or the creation of a new model if there is no trustworthy mode available. This makes it unnecessary to save large training data in the memory, because the models are updated incrementally. Furthermore, LWPR is using an online version of the dimensionality reduction method Partial Least Squares (PLS) to handle redundant and irrelevant input data. The goal of PLS is to reduce the dimensionality locally to find optimal local projections and eliminate subspaces of the input space that minimally correlate with the output. This ensures that redundant and irrelevant input dimensions are ignored

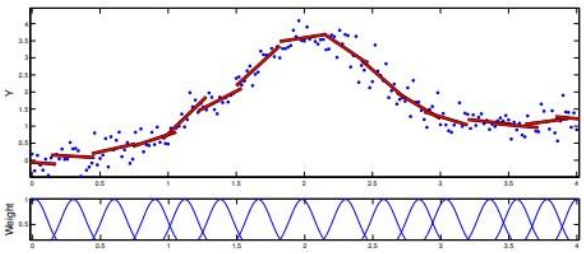

Fig. 2 Example for Locally Weighted Projection Regression (LWPR)

It is containing in the upper graphic the data points (x,y) (blue dots) and the local linear models (red lines). The bottom graphic shows the Receptive Fields which show the activation weight of the corresponding model in above shown figure.

III. EXPERIMENT

A.Experimental data

The data were used to test patient discharge data according to the incidence route in California. The total number of records of data is 4873, and each characteristic variable is as follows.

Table. 2 Experiment Data Attributes

attribute	type	value
Year	nominal	2009 - 2014
Oshpd_ID	numeric	ID set to 10000 or higher
Facility_Name	String	Facility Name
Type_Of_Control	String	Control type (distance, city, etc.)
County_Name	String	State
Admission_Source_Route	String	Admission route
Count	numeric	Occurrences

B. Experimental Process

In the experiment, we want to analyze which routes are used in California and which facilities are mainly used. From the results of the LWL algorithm, it is the main goal to grasp the performance of the LWL algorithm with heavy weight on the hospital routes. We used WEKA v3.6.1 developed by Waikato University as an experimental tool. We will use Weka's LWL algorithm, and the LWL algorithm is a kind of classifier.

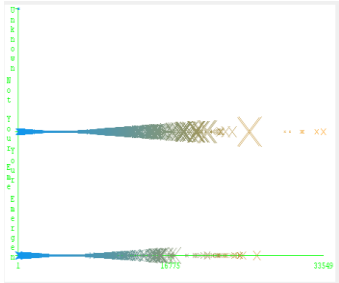

Fig. 3 Visualization of emergency rooms

It is shown an analysis of the use of the emergency room. The results are larger than those used without the emergency room.

The color of each point means the number of times, from blue to orange, the number of times increases. As a result, people in California can see that they do not use the emergency room by almost a third of the difference.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

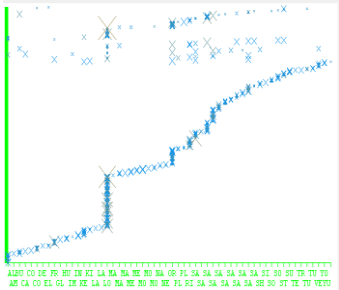

Fig.4Visualization of hospitalization types

It is shown an analysis of which facilities in each region are used as hospitalization types. The total number of facilities is 571, and the average of Count is 12. As a result of these conjectures, various facilities will be used, and the cluster will not be concentrated in specific areas of the facility.

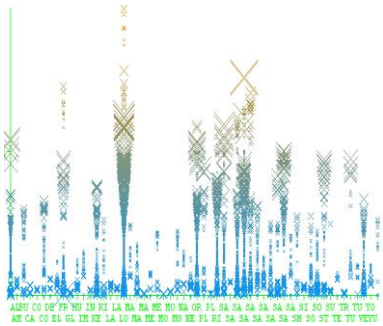

Fig.5 Visualization of hospitalization types

Above figure shows the counts for each region in California. In each region, the highest number of counts is LOS ANGELES (1215), the second is ORANGE (365), and the third is SAN DIEGO (275). LOS ANGELES was more than three times higher than other counts.

IV. EXPERIMENTAL RESULTS

It is resulted as the output value when the LWL algorithm is executed as follows

```
Correlation coefficient                  0.5844
Mean absolute error                   3114.2508
Root mean squared error               3907.4643
Relative absolute error                 82.9247 %
Root relative squared error             81.6577 %
Total Number of Instances             4872
```

Fig.6 Experimental Result of LWL

After experimental execution, Correlation coefficient ranked as 0.5244 and average of absolute error 3223.0814. It shows as normal. In other words, one out of three to four people use the emergency room, so it can be assumed that a relatively large number of patients use the emergency room. When we analyzed the use of facilities in each state, there was no phenomenon of cluster concentration in one place. More than half of the data was output at LOS ANGELES, and the path not using the emergency room was about one third more than the path used by the emergency room. In other words, one out of three to four people use the emergency room, so it can be assumed that a relatively large number of patients use the emergency room. When it is analyzed as the use of facilities in each state, there was no phenomenon of cluster concentration in one place.

V. CONCLUSION

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.[1] Data mining is an interdisciplinary subfield of computer science with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. In the paper the information of hospitalized pathways is researched that stored electronically as a hospital information system. The database stores all the data related with medical actions, including accounting information, laboratory examinations, and patient records described by medical staff. Incident or accident reports are not exception: they are also stored in HIS as clinical databases. All the clinical inputs are shared through the network service in which medical staff can retrieve their information from their terminals. Since all the clinical data are distributed stored and connected as a large-scale network, it is expected that similar techniques in data mining, web mining or network analysis can be applied to the data. Dealing with cyberspace in a hospital will give a new challenging problem in hospital management in which spatiotemporal data mining, social network analysis and other new data mining methods may play central roles. In this study, we use the LWL algorithm among classification algorithms. Identifying these various admission pathways and finding out the weight of various hospitalization pathways can develop toward the response and quality of medical care. Therefore, it is classified into LWL algorithm to predict various hospitalization routes and specific gravity.

ACKNOWLEDGMENT

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. 2016M3A9B694241).

REFERENCES

1. Frank, Eibe, Chang Chui, and Ian H. Witten. "Text categorization using compression models." (2000)
2. Witten, Ian H., and Eibe Frank. "WEKA-Waikato Environment for Knowledge Analysis." Internet: http://www. cs. waikato. ac. nz/ml/weka,[Mar. 2, 2008] (2000).
3. Witten, Ian H., et al. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.
4. Appel, Charlotte Weiling, et al. "Risk of use of antidepressants among children and young adults exposed to the death of a parent." Epidemiology 27.4 (2016): 578-585.
5. Naeher, Luke P., et al. "Use of isotope ratios to identify sources contributing to pediatric lead poisoning in Peru." Archives of Environmental & Occupational Health 58.9 (2003): 579.
6. Crespo, M., et al. "Emotional state in employed caregivers of frail elderly relatives." European Psychiatry 33 (2016): S342-S343.
7. Schmidt, Tobias, et al. "Social cognition and social problem solving abilities in individuals with alcohol use disorder." Journal of clinical and experimental neuropsychology 38.9 (2016): 974-990.
8. Ali, A., & Haseeb, M. (2019). Radio frequency identification (RFID) technology as a strategic tool towards higher performance of supply chain operations in textile and apparel industry of Malaysia. *Uncertain Supply Chain Management, 7*(2), 215-226.
9. Misganaw, Chalachew, Metasebia Mesfin, Megdelawit Tesfaye, and Andualem Derese. "Retrospective study on outcome of in-patient treatment of severe acute malnutrition in Jimma University specialized hospital from September 2011-September 2012." Journal of Diagnostics 1. 2 (2014): 18-28.
10. Ghorbaniparsa, Fatemeh, and Hamideh Ofoghi. "Comparing Patatin Class I and Camv 35s Promoters in Expression of Human Calcitonin Gene in Potato (Solanum Tuberosum Cvs. Kardal And Marfona)." The International Journal of Biotechnology 5.4 (2016): 52-61.
11. Sandygulova, Anara. Adapting child-robot interaction to reflect age and gender. Diss. University College Dublin (Ireland), 2016.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication