Recommendation an Emergency Patient Destinations by LightGBM

Ryota Kawaguchi, Masao Kubo, Hiroshi Sato
Information Engineering, National Defence Academy of Japan, Hashirimizu 1-10-20
Yokosuka, Kanagawa 239-0811, Japan
E-mail: em60041@nda.ac.jp, masaok@nda.ac.jp, hsato@nda.ac.jp
www.mod.go.jp/nda/english/

Daizoh Saitoh
National Defence Medical College of Japan, Namiki 3-2
Tokorozawa, Saitama 359-8513, Japan
www.mod.go.jp/ndmc/

Takao Oya
Saitamaseibu Fire Bureau, Keyakidai 1-13-11
Tokorozawa, Saitama 359-1118, Japan
www.saisei119.jp/

Abstract

We propose to adopt the LightGBM method for recommending a hospital through we conduct some experiments by using patient transport data, and we found that location information, age, degree of injury are important elements in the situation of selecting a hospital. In addition, we found that accuracy can be obtained without detailed personal information, and the task of selecting a destination hospital can be decentralized.

Keywords: emergency patient, LightGBM, Machine Learning.

1. Introduction

In this study, we propose a method to recommend an emergency patient destination to reduce the burden of selecting a destination hospital for rescue teams which run in parallel with first aid and selecting a destination hospital. Since we use only patients' basic status that may be allowed to be used as input, it can be implemented even in local governments with insufficient understanding of IT utilization in medical care. Concretely speaking, we use patient summary information, location information, and conversations with call centers as input data for machine learning. The data used for the verification here are the data of the 5 most frequently used hospitals in the case of emergency transportation in western Saitama prefecture. The destination hospital was predicted based on the above data, and the accuracy was obtained by comparing it with the actual destination. We have investigated several machine learning methods. As a result, the LightGBM, a decision tree-based machine learning method, achieved the best accuracy compared to some other methods, with a score of about 70%. In addition, as a result of analyzing the emergency transport data, it was found that age, degree of injury, and location information are important for improving the accuracy.

2. Previous Researches

According to the data released by the Fire and Disaster Management Agency of Japan, the number of emergency patients is increasing year by year due to the aging of the population[1]. Therefore, there is an urgent need to reduce the burden of transport operations for rescue teams. In addition, the transportation time from the report to the hospital is increasing. We think that it needs to streamline transportation operations to save critically ill patients who require hospital treatment at an early stage. Against this background, there are some previous studies based on emergency transport data. In the study [2], machine learning methods are used to determine whether hospitalization is necessary or not. In the study [3], it was investigated the conditions of patients who are denied hospitalization. In the study [4], machine learning predicts whether a hospital can admit a patient, but the algorithm is not disclosed because it contains personal information. Therefore, in this research, detailed personal information is not used as input data for machine learning. Not handling the detailed personal information will make it difficult to predict, but will facilitate system introduction.

3. Data Set

The dataset contains 155,369 records taken to 287 hospitals in western Saitama Prefecture. We use only the records of the 5 most used hospitals in this dataset. Since the name and address of the hospital are not recorded, they are called A to E. The number of extracted records is 63,829, which includes 10 elements of information such as notification hour, notification month, day property, sex, age, degree of injury, consciousness level, location information, conversation with a call center, and destination hospital number. The breakdown of this dataset is shown in Table 1. There are 26,186 records to Hospital A, accounting for about 41% of the total. Even the smallest E has about 11%. In the following, Summary Vector(SV) includes information of notification hour, notification month, day property, sex, age, degree of injury, consciousness level. Location Vector(LV) includes information about the location. Conversation Vector(CV) includes information of conversation with a call center. All elements of these 3 vectors were normalized to the range 0 to 1.

Table 1. Data Set Breakdown.

Hospital	Number	Percentage
A	26,186	41.0
B	12,661	19.8
C	9,644	15.1
D	7,848	12.3
E	7,490	11.7
Total	63,829	100.0

3.1. *Summary Vector(SV)*

A SV includes 7 elements. Notification hour and notification month are the information when the call center is notified. The day property is a holiday or weekday. The degree of injury has one of mild, moderate, severe, or death. Consciousness level is an integer from 0 to 300, where 0 is normal and 300 is unconscious. Convert each of these elements from 0 to 1. For example, 0 for men and 1 for females.

3.2. *Location Vector(LV)*

A LV is anonymized location location information of a 2-dimensional. The dataset contains information up to the village section of a patient's location to protect personal information. Convert this information into a vector with two-dimensional elements. First, convert the address to latitude and longitude using the Yahoo API[5]. Then normalize to the range 0 to 1.

3.3. *Conversation Vector(CV)*

A CV includes 300 elements. Each conversation in this data describes in short sentences for example "I'm bleeding in contact with a car". The sentences are converted into a 300-dimensional vector using the Doc2Vec model learned on Japanese Wikipedia[6].

4. Experiments

We conducted the following three experiments.

1) Examination of prediction algorithm: First, we examined the optimal prediction algorithm. We compared the five methods of Support Vector Machine (SVM), k-nearest neighbor method, logistic regression, neural network, and LightGBM, and clarified which method is suitable for this problem. For the input data, 309-dimensional vectors of SV (7-dimensional), LP (2-dimensional), and CV (300-dimensional) were used. we calculate the score with the following formula (1).

$$Score = \frac{Number\ of\ Correct\ Answers}{Number\ of\ All\ Data}. \qquad (1)$$

2) Sensitivity analysis of input vectors: Next, we clarify what vector is important. We experiment using the best prediction method in the above experiment. We experiment using the best prediction method in the above experiment with combination of various vectors.
3) Sensitivity analysis of Summary Vector: Third, we clarify what element is important in SV. We experiment with combinations of all elements of SV.

4.1. *Comparison by Prediction Algorithms*

The ratio of training data to test data was 9: 1, and each experiment was randomly divided. Table 2 shows averages of the accuracy of each method when 10 experiments were performed. From this result, it was found that LightGBM gives the best accuracy. We decided to use LightGBM for the subsequent experiments.

Table 2. Comparison by Prediction Algorithm.

Method	Score[%]
SVM	67.8
k-NN(k=10)	59.8
Logistic Regression	66.3
Neural Network	68.8
LightGBM	70.7

4.2. *Comparison by Input Vectors*

Next, we analyzed the important vectors among the three input vectors. Compare the accuracy by combinations of the 3 vectors. The prediction method used was LightGBM, which was the most accurate in the first experiment. The results are shown in Table 3. From this result, it was found that all vectors contribute to the improvement of accuracy, but LV is an important factor for improving accuracy.

Table 3. Comparison by Input Vectors.

Input vector	Number of input element	Score[%]
SV	7	44.4
LV	2	65.4
CV	300	41.7
SV + LV	9	69.7
SV + CV	307	45.1
LV + CV	302	67.4
SV + LV + CV	309	70.7

*SV:Summary Vector
LV:Location Vector
CV:Conversation Vector

4.3. *Comparison by Summary Vector Elements*

Third, we experimented with combination of various elements of SV by LightGBM. PV and CV are always used in this experiment. There are 2 ^ 7 = 128 patterns for all combinations. The results are shown in Table 4. Looking at Score Ranking 1 to 5 in Table 4, all 7 elements were needed because almost all 7 elements were used. In particular, the top 31 cases always included age and degree of injury, indicating that these two factors are important.

Table 4. Comparison by Summary Vector Elements.

Score Ranking	Notification Hour	Notification Month	Day property	Sex	Age	Degree of injury	Consciousness level	Score[%]
1	○	×	○	×	○	○	○	70.7
2	×	×	○	○	○	○	○	70.6
3	○	×	○	○	○	○	○	70.6
4	○	○	○	○	○	○	○	70.6
5	○	○	×	○	○	○	○	70.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
124	○	×	×	×	×	×	×	67.4
125	×	×	○	×	×	×	×	67.3
126	×	○	×	×	×	×	×	67.3
127	○	○	×	×	×	×	×	67.3
128	×	○	○	○	×	×	×	67.0

* ○:use element, ×:not use element

5. Discussion

In the first experiment, the five prediction methods were evaluated with all available information. As a result, it makes LightGBM the most effective for this problem. In this study, about 41% (63,829 / 155,369) of the total transport data was extracted and used as the input data. It shows that software can replace the job in about 70% of cases. In the second experiment, the sensitivities of three vectors(SV, LV, CV) were investigated. As a result, it founds all three vectors is indispensable. It was also found that LV was the most important. In the third experiment, the sensitivities of the seven elements of SV were investigated. As a result, it became clear that it is effective to use all seven elements. It was also found that gender and degree of injury are particularly important.
From the above three results, it was found that location information, age, and degree of injury should always be collected when constructing a system for recommending a destination hospital. Furthermore, the task of selecting the destination hospital can be decentralized by starting hospital selection from the time of notification.

6. Conclusion

In this study, to reduce the burden of selecting a destination hospital, we investigated a method of recommending a destination hospital from patient information without detailed personal information. As a result, LightGBM was excellent, and we were able to achieve a score of about 70% for the accuracy of selecting a hospital. We also tried various input data patterns and found that age, degree of injury and location information contributed to improving learning accuracy in particular.
Since information on the above three important elements can be transmitted from the any person to the call center, it is possible to promptly recommend a hospital and reduce the burden of selecting a hospital for an on-site rescue team. We also found that accuracy can be obtained without detailed personal information, and what kind of information is referred by the rescue team when selecting a hospital.

References

1. Fire and Disaster Management Agency of Japan, 2020, https://www.fdma.go.jp/pressrelease/houdou/items/c941509de3f85432709ea0d63bf23744756cd4a5.pdf
2. Satoshi Iwai, et al. "Admission predicting system using pre-arrival information in emergency departments", AIM, SIG-AIMED-001-04, 2015.
3. Mika Kumagai, et al. Japanese Society of Public Health, 63(3) 116-124, 2018.
4. "NTT Data News Release", https://www.nttdata.com/jp/ja/news/release/2018/112600
5. Yahoo JAPAN!, https://developer.yahoo.co.jp/webapi/map/
6. Yuki Okuda, https://yag-ays.github.io/project/pretrained_doc2vec_wikipedia/

Authors Introduction

Mr. Ryota Kawaguchi

He received his Bachelor's degree from the Department of Information Engineering, Kagawa University, Japan in 2018. He is currently a Master course student in National Defense Academy in Japan.

Dr. Masao Kubo

He is an Associate Professor of the Department of Computer Science at National Defense Academy in Japan. He graduated from the precision engineering department, Hokkaido University. He received his Ph.D. degree in Computer Science from the Hokkaido University.

Dr. Hiroshi Sato

He is an Associate Professor of the Department of Computer Science at the National Defense Academy in Japan. He obtained a degree in Physics from Keio University in Japan and degrees of Master and Doctor of Engineering from Tokyo Institute of Technology in Japan.