2015년 추계학술발표대회 논문집 제22권 제2호(2015. 10)

Electropulsegraph 및 파형분류 프레임워크

박진수*, 최동학**, 민세동***, 박두순****
순천향대학교
e-mail :{ *vtjinsoo, **cdh0191, ***sedongmin, ****parkds}@sch.ac.kr

Electropulsegraph and Wave Classification Framework

JinSoo Park*, Dong Hag Choi**,Se Dong Min*** and Doo-Soon Park****
*,**Industrial Academy Cooperation Foundation, SoonChunHyang University
***Dept. of Medical IT Engineering, SoonChunHyang University
****Dept. of Computer Software Engineering, SoonChunHyang University

Abstract

Electropulsegraphy is a medical device that was invented by an orient medical physician and a few engineers to help the physicians to diagnose patients in more systematic way by analyzing waveforms generated from the device. Data generated form the device has been collected for over several decades, and undergoes functional upgrades today. The device generates 33 waveforms that reflect the states of patients. As one of those upgrading efforts, we strive to develop an intelligent algorithm that makes the diagnostic process automatically, which was previously done manually for a long period of time. The logistic regression algorithm is used for our classification problems, which is one of those well-known algorithms for various classification problems such as character recognition systems. Out of the 33 waveforms, we only use 5 waveform data (Type1 toType5) as training data sets to estimate the parameters of the logistic regression. And the parameters are used to classify waveform inputs chosen at random.

1. Introduction

Doctors of oriental medicine have used a semi-diagnostic device that generates 33 waveforms from different organs reflecting the physical states of patients. The data have been collected for over several decades and used today for its functional upgrades to make the diagnostic process automatic. In this article, we provide the results of this preliminary study on the automatic diagnostic system that classifies the waveforms from the organs to help doctors to diagnose patients more efficiently. It employs one of the well-known machine learning algorithms-logistic regression.

We strive to apply the logistic regression algorithm for the classification of waveforms generated from our dedicated device, thus the physicians can use the classification results for their diagnostic decisions on the patients later. In [1], the author suggests a method to model electroencephalography waveforms with semi-supervised deep belief nets for fast classification and anomaly measurement. The corresponding experimental results showed that the fast classification and anomaly measurement of EEG waveforms are possible with deep belief network algorithms. However, the algorithm assumes only five different classes of EEG waveforms as target waveform classes, which is not appropriate to more than five-type of waveform classification problems.

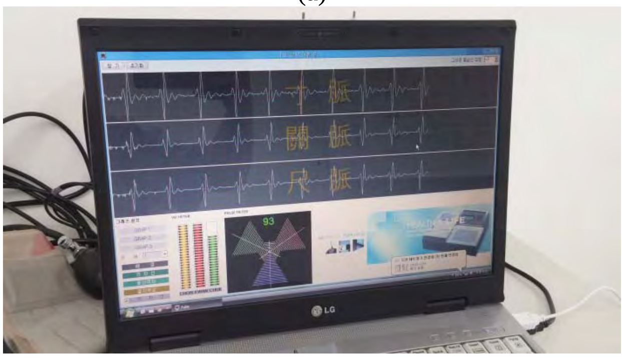
The logistic regression algorithm is one of the well-known machine learning algorithms, and it has been used for multiclass classification purposes today. One of those application areas can be character recognition systems [2]. The algorithm is used to classify hand-written characters based on derived parameters by the logistic regression algorithm. We put our goal in developing classification systems for more than thirty different waveforms simultaneously. As a preliminary study, we describe a method for the multiclass classification used in this study and classification results performed for five different types of waveforms.

2. Classification Results

The dedicated device generates 33 waveforms reflecting the states of internal organs respectively, and amplitude for each wave has a periodic wave patterns generally as shown in Fig. 1. We use the data collected for over decades, and only some of the data recently collected are used for our experiment.

Among the thirty-three different waveforms obtained from the device, we use only five of them. Thus training data sets can be obtained for those waveforms respectively. In this article, we only report the training and classification results for the five different probing waveforms – Type 1, Type 2, Type 3, Type 4 and Type 5.

(a)



(b)

Fig. 1. Electropulsegraphy and waveforms

Fig.2. illustrates samples of waves from the four different types where each type of wave has relatively similar amplitudes and frequencies respectively. Each waveform is composed of 783 integer values as amplitude values. The number of training data is different for each probing types because when patients with rare symptoms come to see doctors, it sometimes led to inadequate number of data for that particular type of symptoms.

We emphasize that our research goal is to decide the classes of waveforms for unknown waveforms. During the training process, we first estimate the parameters of the training data set, and the estimated parameters are used to classify the unknown waves using the estimated parameters.

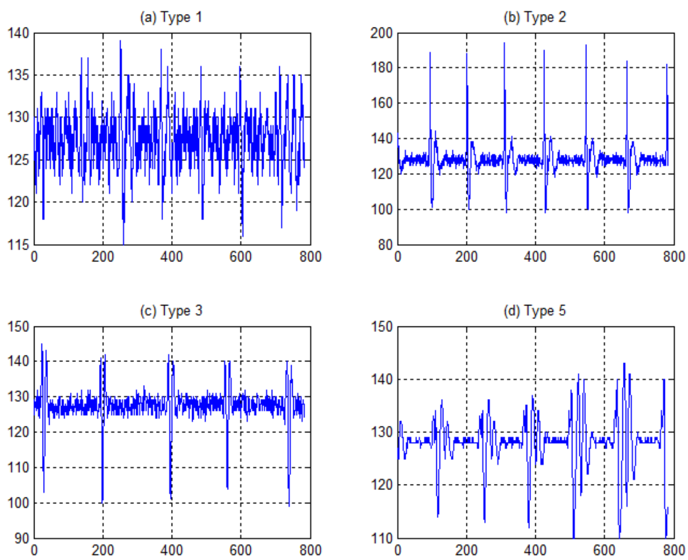


Fig. 2. Sample waveforms from four different types(Type4 is intentionally omitted for space reasons)

In the estimation process, we need to determine the learning rate (α), a parameter that relates to the entire learning speed. As the learning rate is changing from 0.04 to 0.4, we examine the change in training accuracy. The accuracy is fairly stable, so we choose 0.1 for α.

Table 1 shows the classification accuracy of training data set measured at four different probe points- Type 1, Type 2, Type 3, Type 4 and Type 5. The number of data for each probe points is 166, 12, 19, 14 and 53 respectively. And as shown in table 1, the estimation accuracy is fairly high for each probe points except Type 5 respectively.

Table 1. Estimation accuracy for each training data set

Probe points	Number of data	Estimation accuracy (%)
Type 1	166	100.00
Type 2	12	91.67
Type 3	19	89.47
Type 4	14	100.00
Type 5	53	60.38

We expect that as we increase the number of training data set and execute a few preprocessing procedure, we can achieve better classification results later.

3. Conclusions

This research addresses a classification framework for the classification of electropulsegraph waves. The data collected for several decades are used to train the data set. We employ the logistic regression algorithm for this classification work that has been used for various multiclass classification problems. Through the experiment, we demonstrated that the algorithm works well for the medical waveform classification.

Acknowledgement

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, and run under the C-ITRC(Convergence Information Technology Research Center) (IITP-2015-IITP-2015-H8601-15-1009) program supervised by the IITP(Institute for Information & communications Technology Promotion).

References

[1] D F Wulsin, J R Gupta, R Mani, J A Blanco, B Litt, Modeling electroencephalgraphy waveforms with semi0supervised deep belief nets: fast classification and anomaly measurement, (2011) 1-13

[2] Kinjal Basu, Radhika Nangia, Umapada Pal, Recognition of Similar Handwritten Characters using Logistic Regression, 2012 10th IAPR International Workshop on Document Analysis Systems (2012) 200-204