International Journal of Current Research and Review
DOI: http://dx.doi.org/10.31782/IJCRR.2020.12163

Scopus

Research Article

IJCRR
Section: Healthcare
Sci. Journal Impact Factor: 6.1 (2018)
ICV: 90.90 (2018)

A Study on the Development of Analysis Model using Artificial Intelligence Algorithms for PTSD (Post-Traumatic Stress Disorder) Data

Jeong Beom Kim

Professor at Namseoul University, Korea

ABSTRACT

A typical mental effect that occurs over a long period of time after experiencing a physical and traumatic event is known as PTSD(post-traumatic stress disorder). Post-traumatic stress disorder is a complex psychological trauma that can endanger self-harm and suicidal impulses, even in serious cases leading to suicide. The root causes of this phenomenon can be found everywhere, but repeated trauma exposure and inadequate medical support related to the professional specificity of firefighters can be cited as the main reasons. As such, firefighters' post-traumatic stress disorder needs to be treated as a mental health issue in the public domain beyond the mental health level of the individual. For the sake of stability, social problems should be covered in a broad sense by advanced research on high-risk analysis tools and preemptive forecasting methods along with a shift in perception of actual risk factors in modern society. Based on AI algorithm and big data, this study aims to develop the analysis model about prevalence of post-traumatic stress disorder caused by mental health damage factors for firefighters exposed to various trauma such as accidents, disasters and stress.

Key Words: Predictive Model, PTSD (post-traumatic stress disorder), Cluster analysis, Machine Learning, AI algorithm

INTRODUCTION

The term post-traumatic stress disorder (PTSD) was first used in the field of psychiatry in the United States in 1980 through the Diagnostic and Statistical Manual of Mental Disorders (DSMIIII). PTSD initially began to draw attention through symptoms of anxiety, sleep disorders and depression experienced by soldiers dispatched to war or conflict areas, and is now commonly used as a generic term for natural disasters, social accidents and trauma suffered by victims of the incident. In particular, in the process of developing an advanced civil society, high-risk PTSD groups are increasing due to the diversification of large-scale disaster and accident that cannot be compared to the past. Although there are obvious differences in the intensity and response patterns of trauma accepted by individuals in certain events, the main purpose of the research is to develop intelligent analysis tools based on big data in accordance with the existing PTSD diagnostic criteria and implement analysis algorithms. The limitations of the methods of diagnosing and investigating these complex factors do not completely eliminate the possibility of contamination of the resulting data by the self-defense instinct of the trauma experience. The focus was on R&D on the big data-based analysis system implementation algorithm that can identify and predict major risk factors of post-traumatic stress disorder (PTSD).

Related study

Due to various social and natural disasters and accidents, physical trauma and mental health damage are increasing. In particular, in addition to direct and physical accident experiences, indirect trauma exposure that is delivered or contacted through family or acquaintances can cause serious stress disorders. In particular, firefighters who perform first aid and disaster prevention activities at various disaster, disaster, and accident sites have a big problem in that they are relatively more prone to trauma and repeatability than the general population. The average number of people in charge of one fire official is 1,300 and the average life expectancy is 58.8 years, ranking among the lowest among government officials. The death rate is very high in relation to mental health, with two

Corresponding Author:
Dr. Jeong Beom Kim, Professor at Namseoul University, Korea.
Email: jbkim@nsu.ac.kr

ISSN: 2231-2196 (Print) **ISSN:** 0975-5241 (Online)
Received: 15.06.2020 **Revised:** 14.07.2020 **Accepted:** 05.08.2020 **Published:** 22.08.2020

of the firefighters who died in 2015 and 12 others dying. The typical post-traumatic stress disorder (PTSD), which is expressed over a long period of time after experiencing events caused by physical and mental shock, is known as the common post-traumatic stress disorder (PTSD), along with depression, sleep disorders and panic disorder. PTSD is a complex psychological trauma that can actually lead to suicide in serious cases such as self-harm and suicidal impulse against the danger group itself. The fundamental causes of this phenomenon can be found in many places, but repeated trauma exposure and lack of adequate medical support can be cited as the main reasons for the occupational specificity of fire officials.

As such, the psychological post-traumatic stress disorder of fire officials needs to be treated as a mental health problem in the public domain beyond the mental health level of the individual, and perceptual changes should be made along with research on advanced analysis tools and preemptive forecasting methods for PTSD high-risk groups among those who perform official duties at the forefront of protecting the property and life of the people, who stabilize the working environment in a broad sense. In this study, we would like to investigate the prevalence and status of PTSD due to mental health damage factors of fire officials exposed to various trauma such as accidents, disasters, and stress. Based on the analyzed data, the classification system for PTSD-occurring factors can also be derived. Risk factor prediction techniques and methods that can be used as key information assets to help treatment and support programs are being studied in many ways.

In this study, we referenced two predictive model. One is SVM (Support Vector Machine) and the other is NNM(Neural Network Model) with keras platform.

SVM is a supervised learning model that mainly performs pattern recognition and data analysis. It uses classification and regression. SVM algorithm determines which area the new data will belonged to according to the given data. This classification model is a model that finds the boundary with the largest width in the space where the data is expressed. This model learns how to find the optimal hyperplane and divide it into two groups. The classification model divided by a straight line is called a linear classification model, and when it is divided by a non-straight line, it is a nonlinear model.

Recently NNM(Neural Network Model) is being using as most predictive model. NNM is kind of nonlinear model mainly developed to solve the prediction problem of complex structures, which was developed with a focus on biological neural network structures. This model predicts the optimal output value for input value by connecting several neurons to each other. Neural networks consist of an input layer, a hidden layer, and an output layer. The input layer is composed of nodes corresponding to input variables, so the number of nodes is the same as the number of input variables. In the hidden layer, it is processed as a nonlinear function through linear combination of variable values transmitted from the input layer, and transmitted to the output layer, or in the case of deep learning, to other hidden layers. The output layer is a node corresponding to the output variable, and the number of occurrences is used as the output variable. In the case of the classification model, as many output nodes as the number of classes are generated. A single-layer neural network is a model consisting only of an input layer and an output layer, and a multi-layer neural network is a model consisting of an input layer, a hidden layer, and an output layer. It has very high predictive power, but has difficulties in interpretation, so it is not used well in areas where interpretation is important, such as credit rating, but it is applied to fields such as speech recognition.

The background of research

The background of this study is to use the research data which express early warning of post-traumatic stress disorder by mapping the report data to the 119 situation room. The firefighting field activity report data of the crews are the main source of research data. Based on the question and answer method, there is a variable that the reliability of the result will be deteriorated if the individual firefighters are involved. By analyzing the log of emergency activities, this paper aims to secure the correlation, objectivity and reliability of the results. Fire officials who are responsible for the property and safety of the people are generally occupied with the required high professionalism, speed and accuracy along with physical health. At the same time, a healthy mental power must be supported to prevent secondary accidents that may occur during public service. In 2008, the study by the National Emergency Management Agency (formerly National Fire Service) found that over 80% of firefighters experienced traumatic events, while 35.9% said they needed mental health services.

Among the 2,705 fire department officials from 17 fire departments were selected as the study group. This paper referenced the report and dispatch command data entered into the 119 Situation Room over the past two years. Through collecting survey data and deriving risk group standard data for rescue and paramedics. This paper defined data model of field activity personnel by dispatch order type . The research results are presented in table 1.

Table 1: Fire field activity Personnel by dispatch order type

Classification	PTSD		Depressive Disorder		Sleep Disorder		Intention of Care
	Needed Control	Needed Care	Needed Control	Needed Care	Needed Control	Needed Care	
Number	3,752	2,806	4,886	3,985	13,507	8,804	4,116
Ratio	10.1%	5.6%	13.2%	10.7%	36.4%	21.8%	28.6%

Following is the analysis of research result. First, PTSD risk-inducing factors were used by ICE-R, CESD, Suicide Accident (SBQ-R), and Alcohol Drinking Scale (AUDIT-K). Traumatic risk causing factors were used according to the characteristics and types of field activities of the crew. Second, the actual rate of using the service was only 6.2%, which was pointed out due to lack of substitutes and adjustment of working hours. At present, the lack of human and material resources for fire officials is unresolved, and it is also closely related to the recent issue of the transition to national positions of firefighters, who are highly interested in society. Firefighters currently classified as local officials are determined by the financial status of their local governments, which means that the survival rate of people in the event of disasters, and accidents is proportional to local finances. This is directly related to the treatment of firefighters, that is, the advancement of the working environment and health welfare. The central government support is urgently needed regardless of the phenomenon of the rich and poor. Third, there is a discrepancy in the quality of rescue and emergency services in case of emergency depending on the financial status of the region. Prompt, accurate rescue and first aid services for firefighters can be manifested from the healthy body and spirit, which can be found in the case of developed countries. The US fire department's top priority is "fireman's health and safety," and each state is enacting a law that provides mental health counseling and treatment services to families of firefighters, including firefighter suicide prevention programs. As a quantitative basis for such policy decision making, research on automatic analysis of risk groups and early warning system for big data-based "traumatic stress disorder" is expected to be utilized as an important information asset for solving social problems. The number of PTSD patients in Korea is 7896, which is less than expected (Health Insurance Review and Assessment Service). This is because we do not know or hide that he/she is a patient. There is an atmosphere in Korea where it is difficult to talk about mental health at first, so the prevalence of mental disorders, including PTSD, is lower than in foreign countries. If we are constantly looking for information related to an infectious disease in anxiety, or if we are fleeing past places that create fear, we should be careful with PTSD. In particular, testing is recommended if he/she is in severe stress due to self-restraint among the isolates. Usually treatment works well within 4 weeks of the event. There are 'cognitive behavioral therapy' that reminds of trauma in a safe situation, and 'sensitization and reprocessing of eye movements' that maximizes the information processing function of the brain and blurs the memory.

Development of Analysis Model

SVM(Support Vector Machine) Model

The distance between the support vector and the decision boundary is called the margin. The goal of SVM is to find the decision boundary that maximizes this margin. The decision boundary is a boundary for determining different classification values. The support vector is the foremost data point that influences the creation of the decision boundary. The margin corresponds to the distance between the decision boundary and the support vector. A characteristic of linear SVM is to classify data without using the kernel. In this model, we can control the size of the margin by controlling the cost. Using kernel tricks, if linear separation is not possible in a given dimension, the decision boundary is found through the effect of moving the data to a higher dimension. The Gaussian RBF kernel adjusts the curvature of the decision boundary by adjusting the standard deviation of the Gaussian function applied to the data points. The standard deviation adjustment variable is called gamma. The **figure 1** below shows the difference in crystal boundary curvature according to the size of gamma.

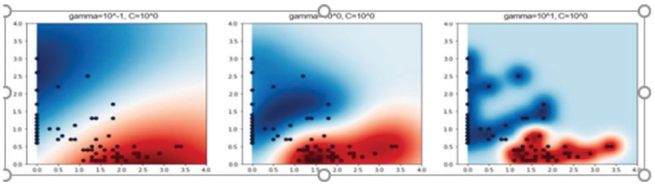

Figure 1: The difference in crystal boundary curvature according to the size of gamma.

We selected cost and gamma as adjustable parameters to increase accuracy. Cost is the margin width adjustment variable. The larger the margin, the narrower the width, and smaller the wider the margin. Gamma is the standard deviation control variable of the kernel. The smaller the data point, the greater the influence of the data point, the smoother the boundary, and the larger the data point, the

less the data point affects the decision boundary. The advantages of choosing the SVM model are as follows. First, by using kernel tricks, it is strong in classifying data with various characteristics. Second, it is possible to cope with over fitting by adjusting the parameters (Cost, gamma). Third, classification with high accuracy can be expected with deep learning even with little learning data depicted in Figure 2, Figure 3 and Table 2.

Following is a analysis model using learning data from questionnaire data. Sklearn module is used to get the optimal result.

```
# import main library
    import SVC GridSearchCVclassification_reportaccuracy_score numpy
    from sklearn.model_selection
    #define parameters
    def svc_param_selection(X, Y):, svm_parameters = [{'kernel': ['rbf'],
    'gamma': [0.00001,0.0001, 0.001, 0.01, 0.1, 1],
    'C': [0.01, 0.1, 1, 10, 100, 1000]
    }], clf = GridSearchCV(SVC(), svm_parameters, cv=10) clf.fit(X_train, y_train.values.ravel()) print(clf.best_params_)
    return clf
    X_train = train[['ICE-R', 'CESD', 'SBQ-R', 'AUDIT-K']
    Y_train = train[['Classification']
    clf = svc_param_selection(X_train, y_train.values.ravel(), 10)
    #visualize result
    plt.figure(figsize=(20, 20))
    xx, yy = np.meshgrid(np.linspace(0, 5, 200), np.linspace(0, 5, 200))
    for (k, (C, gamma, clf)) in enumerate(classifiers):
    Z = clf.decision_function(np.c_[xx.ravel(), yy.ravel()])
    plt.subplot(len(C_canditates), len(gamma_candidates), k + 1) plt.title("gamma=10^%d, C=10^%d" % (np.log10(gamma), np.log10(C)),
    size='medium')
    #test model
    X_test = train[['ICE-R', 'CESD', 'SBQ-R', 'AUDIT-K'], Y_test = train[['Classification']
    y_true, y_pred = y_test, clf.predict(X_test)
    print("accuracy : "+ str(accuracy_score(y_true, y_pred)) )
```

Table 2: Precision rate and accuracy of Cost and Gamma

	precision	recall	f1-score	Support
C	0.87	0.87	0.87	8
SG	0.91	0.91	0.91	12
average / total	0.89	0.89	0.89	20
accuracy : 0.89				

PLS regression model is described as blow.

```
    # import main library
    from sklearn.cross_decomposition
    #components
    principalComponents, _ = PLSRegression(n_components=2).fit_transform(train[X_train.columns], train['RUL'])
    principalDf = pp.DataFrame(principalComponents, columns = ['PLS component 1', 'PLS component 2'])
    finalDf = pp.concat([principalDf, train['RUL']], axis=1)
    import matplotlib seaborn
    cutter = []
    for i in range(4):
    cutter.append(10 ** i)
    i = i + 1
    finalDf['rul_range'] = pd.cut(finalDf['RUL'], cutter)
    ax = sns.scatterplot(x = 'PLS component 1',
    y = 'PLS component 2',
    #decomposition
    import PLSRegression from sklearn.cross_decomposition
    pca_coeff = pd.DataFrame(pca.components_[0])
    pls2.fit(train[X_train.columns], train['RUL'])
    pls_coeff.reset_index()
    coeff_merged = pd.concat([corr, pca_coeff, pls_coeff], axis=1, sort=False)
    for col in coeff_merged.columns:
    coeff_merged[col] = coeff_merged[col] / np.std(coeff_merged[col])
    coeff_merged[col] = coeff_merged[col] - np.min(coeff_merged[col])
```

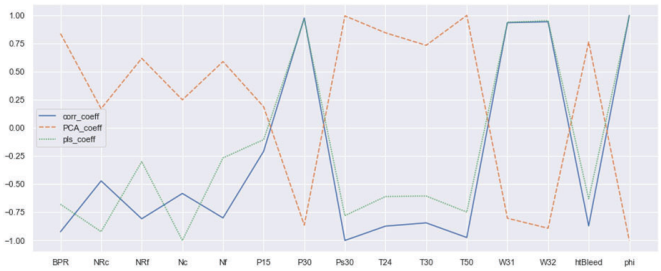

Figure 2: PLS regression.

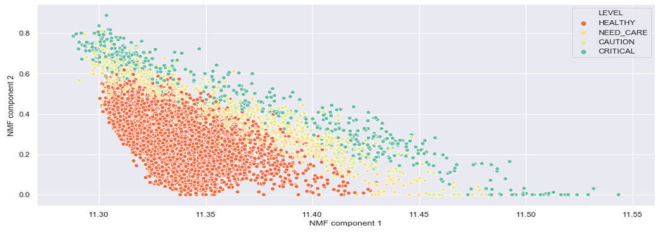

Figure 3: PCA distribution

RNN (Recurrent Neural Network) and Keras Model

If we use RNN(Recurrent Neural Network) analysis, research data can be analyzed in statement type. The program code and figure 4 show the example of using Keras and RNN analysis .

The first phase is to creation of data set. In this phase, following activities are included.

Import original data or generate data through simulation. Create training set, and verify data set. Check format conversion is performed to train and evaluate the deep learning model.

The second phase is to configure model. In this phase, following activities are included.

After creating the sequence model, add and configure the necessary layers. When you need a more complex model, use the Keras function API. The third phase is to set up model training process. In this phase, following activities are included. Set up for learning before learning. Define the loss function and optimization method. Keras uses the compile () function. The fourth phase is to train model. In this phase, following activities are included. Train with a model constructed using a training set. In Keras, we use the fit () function. The fifth phase is to check training process. In this phase, following activities are included. When training the model, we measure the loss and accuracy of the training set and verification set. Determine the learning situation by looking at the loss and accuracy trend according to the number of repetitions. The sixth phase is to verify model. In this phase, following activities are included. Evaluate the trained model with the prepared test set. Keras uses the evaluate () function. The seventh phase is to use model finally. In this phase, the output of the model with random inputs is described. Keras platform uses the predict () function.

```
import np_utils from keras.utils, import PTSD data.csv from keras.datasets

import Sequential from keras.models, import Dense, Activation from keras.layers

(x_train, y_train), (x_test, y_test) = mnist.load_data()

x_train = x_train.reshape(600, 784).astype('float32') / 255.0

x_test = x_test.reshape(100, 784).astype('float32') / 255.0

y_train = np_utils.to_categorical(y_train)

y_test = np_utils.to_categorical(y_test)

model.add(Dense(units=64, input_dim=28*28, activation='relu'))

model.add(Dense(units=10, activation='softmax'))

model.compile(loss='categorical_crossentropy', optimizer='sgd',

 metrics=['accuracy'])

hist = model.fit(x_train, y_train, epochs=20, batch_size=10)

print('## training loss and acc ##')

print(hist.history['loss']), print(hist.history['acc'])

loss_and_metrics = model.evaluate(x_test, y_test, batch_size=12)

print('## evaluation loss and_metrics ##')
```

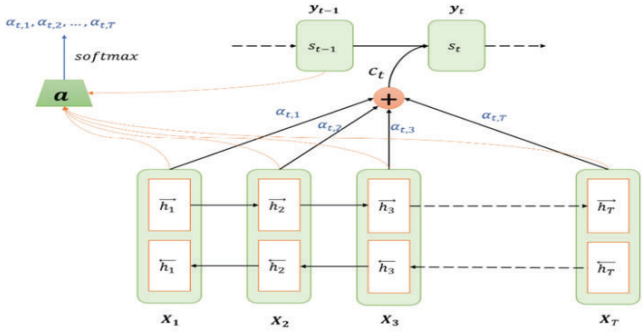

Figure 4: Keras/RNN analysis flow.

CONCLUSION

In this study, the final results were presented as follows: First, SVM(Support Vector Machine)analysis model was used. Second, NNM(Neural Network Model)analysis model was used. PTSD risk factors were used by category of event impact scale (IES-R), depression scale (CESD), suicide accident (SBQ-R), and drinking scale (AUDIT-K). Trauma risk factors according to characteristics and types of field activities were studied. Next paper will aim to develop more practical predictive models using various AI algorithms.

ACKNOWLEDGEMENTS

Funding for this paper was provided by Namseoul University.

Conflict of Interest: Nil

Funding Sources: Namseoul University.

REFERENCES

1. Fire and Disaster Prevention Agency. National firefighters psychological evaluation training analysis results Report. 2014.
2. Brock A, Donahue J, Simonyan K. Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096. 2018 Sep 28.
3. Singh S, Krishnan S. Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020: 11237-11246.
4. De S, Smith SL. Batch normalization biases deep residual networks towards shallow paths. arXiv preprint arXiv:2002.10444. 2020 Feb 24.
5. Big Data Computing Technology. Hanbit Academy. 2016;14-31.
6. Real E, Liang C, So DR, Le QV. Automl-zero: Evolving machine learning algorithms from scratch. arXiv preprint arXiv:2003.03384. 2020 Mar 6.
7. Bello I, Zoph B, Vasudevan V, Le QV. Neural optimizer search with reinforcement learning. arXiv preprint arXiv:1709.07417. 2017 Sep 21.
8. Real E, Aggarwal A, Huang Y, Le QV. Regularized evolution for image classifier architecture search. InProceedings of the aaai conference on artificial intelligence 2019 Jul 17 (Vol. 33, pp. 4780-4789).
9. Liu H, Simonyan K, Vinyals O, Fernando C, Kavukcuoglu K. Hierarchical representations for efficient architecture search. arXiv preprint arXiv:1711.00436. 2017 Nov 1.
10. Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196. 2017 Oct 27.
11. Seo J, Kim J, Moon J, Kang BJ, Im EG. Clustering-based feature selection for internet attack defense. International Journal of Future Generation Communication and Networking. 2008 Dec 1;1(1):91-8.
12. Ogiela MR, Ogiela U. Linguistic cryptographic threshold schemes. International Journal of Future Generation Communication and Networking. 2009 Mar;2(1):33-40.
13. Wooil Kim. Speech Enhancement Based on Feature Reconstruction for Automatic Speech Recognition System with Unknown Structure. International Journal of Image and Signal Systems Engineering. 2017 Feb 1;1(1):7-14.
14. A Study on the Analysis of Post-traumatic Stress in Fire Officials. Ajou University Industrial-Academic Cooperation Group, National Emergency Management Agency. 2008.
15. Ryu Ji-ah, Ha Eun-hee, Jung Choi Kyung-hee, Kim Ji-eun, Park Shin-won and Kim Hyun-joo. Fire officials and post-traumatic stress disorder. 2017.
16. R&D Trends and Implications in Mental Health. Korea Health Industry Promotion Agency R&D Promotion Headquarters. 2013.
17. Hwang, Jae-Joon & Jung, Yun-Hoa & Cho, Bong-Hae & Heo, Min-Suk. An overview of deep learning in the field of dentistry. Imaging Science in Dentistry. 2019:49.