Conditional Time Series Imputation for Distinctive Denoising Diffusion Model

THE UNIVERSITY OF SYDNEY

Ximeng Chen(520015656), Ziliang Wang (490035861) and Huicheng Zhang (500244274)

COMP5329 Deep Learning

Research Problem

Time series are widely applied to the environment, finance and medical areas [1-3]. Data imputation is vital to get accurate results from the model [4]. However, traditional imputation methods rarely take the temporal information of data into consideration. Therefore, deep learning imputation methods that can gain temporal relationships in the time series are popularly employed.

Score-based generative models, which gradually add noise until the data is fully polluted, then denoising to generate data samples [5], have some huge improvements in many downstream tasks, such as image generation [6-7]. Some methods like TimeGrad utilised the Denoising Diffusion Probabilistic Model (DDPM) to forecast multivariate time series, however, because it applied RNNs which derived the current state from the previous time step [8], which is not suitable for imputation tasks. As a result, the Conditional Score-based Diffusion model for Imputation (CSDI) is applied to solve this problem [9].

We found adding some non-Gaussian noise can boost the performance in early stages of the training process in image generation and speech generation tasks [10]. Therefore, we proposed a new idea to replace the Gaussian noise in CSDI with Gamma noise and Mixture Gaussian noise to test the performance of imputation quality under Beijing PM2.5 dataset.

CSDI

Conditional score-based diffusion models for probabilistic time series imputation models (CSDI) were employed in order to predict the missing values in the time series. Each time series will be divided into imputation values and conditional observations. In the forward diffusion process, we gradually add noise to target imputations based on the concept of DDPM, while the conditional observations remain the same, as shown in Figure 1.

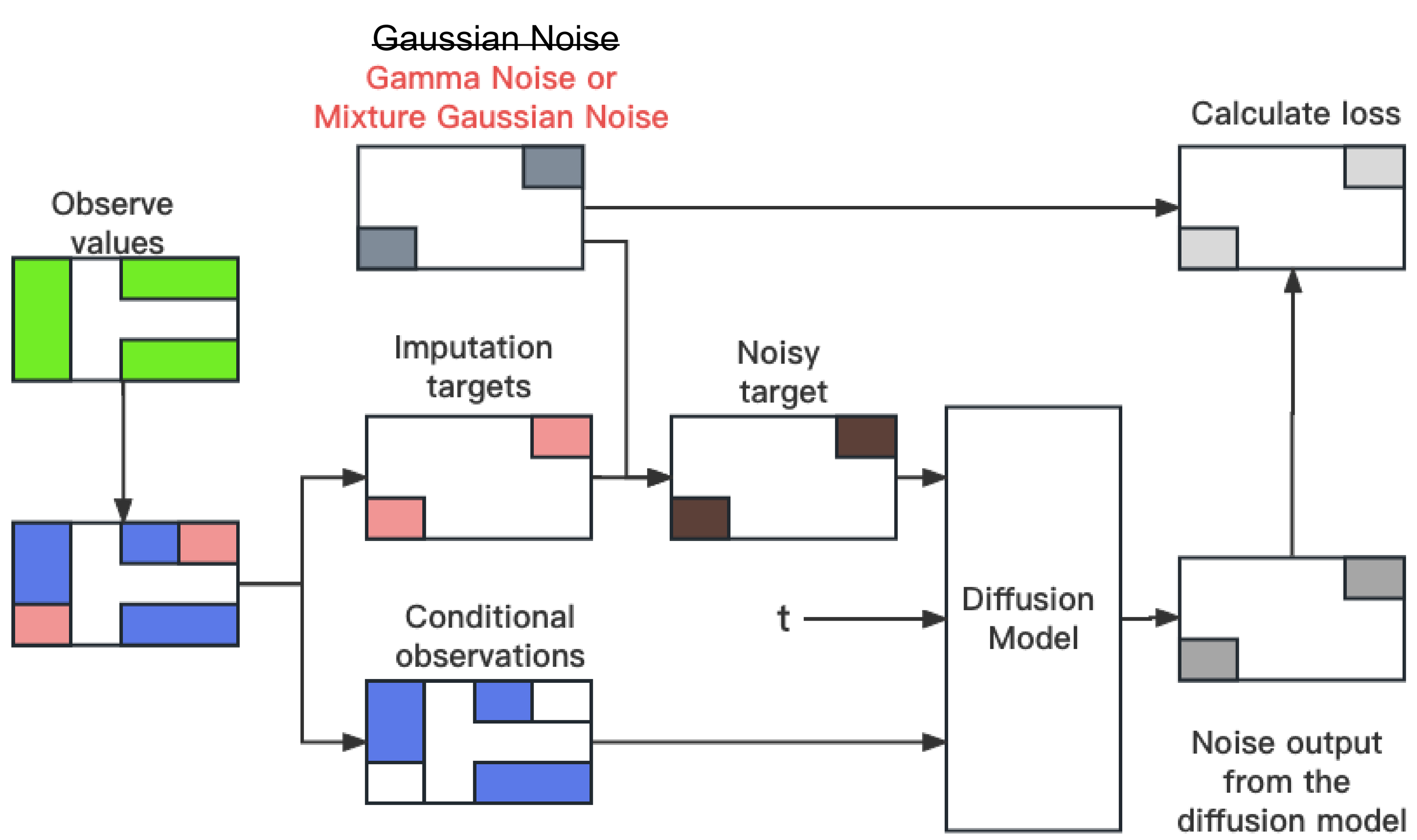

Figure 1 The training process of CSDI

The combination of the noisy targets, conditional observations and timestamps serve as the inputs for the reverse process of the CSDI model, which means that the reverse process progressively transforms random noise into plausible time series on the condition of the previous values [11].

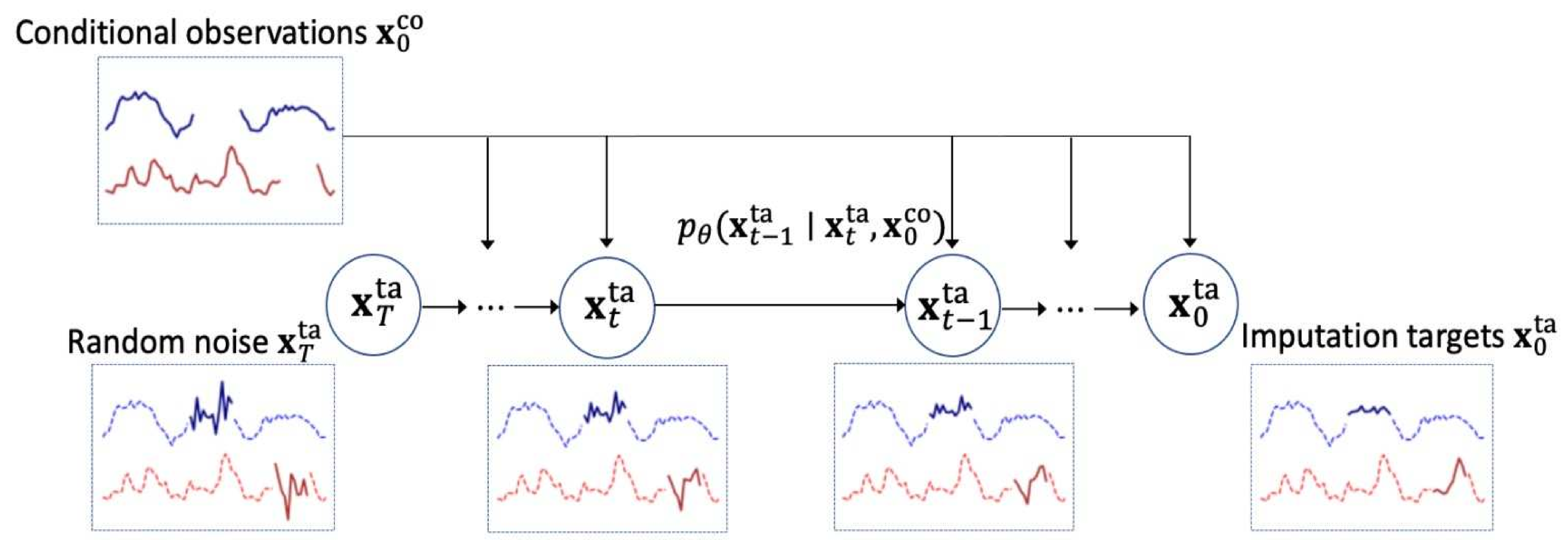

Figure 2 The inverse process of time series imputation with CSDI [9]

Noise

We proposed an approach to replace Gaussian noise in CSDI with Gamma and Mixture Gaussian noise for time series imputation (Figure 1).

Gamma noise: The Gamma distribution is one of the most widely used distribution in real world applications as it can simulate many practical scenarios [12]. Gamma noise is suitable for right-skewed distribution, which always existed in meteorological data [13].

Mixture Gaussian noise: According to [14], mixture Gaussian distributions are frequently used in data analysis and pattern recognition. Because a mixture Gaussian distribution combines several single Gaussian distributions, which can better fit the shape of the data following multiple distributions.

Experiments and Results

Data: We used the Beijing PM2.5 dataset with 13% missing values from 2014/05/01 to 2015/04/30 as the ground truth file. It is clear that the data has a unimodal distribution with right-skewed, which is suitable for Gamma distribution.

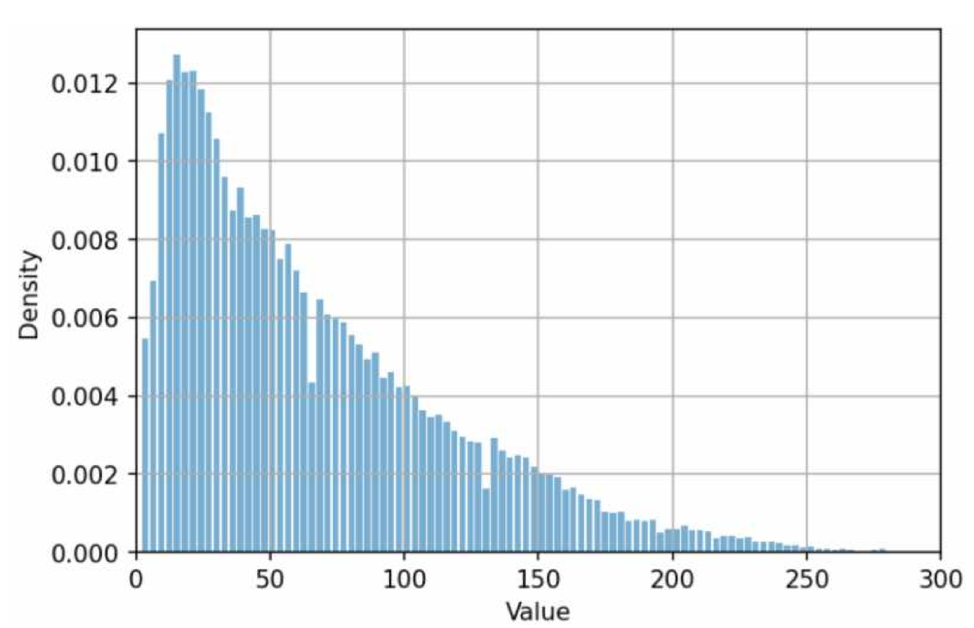

Figure 3 The Distribution of Data

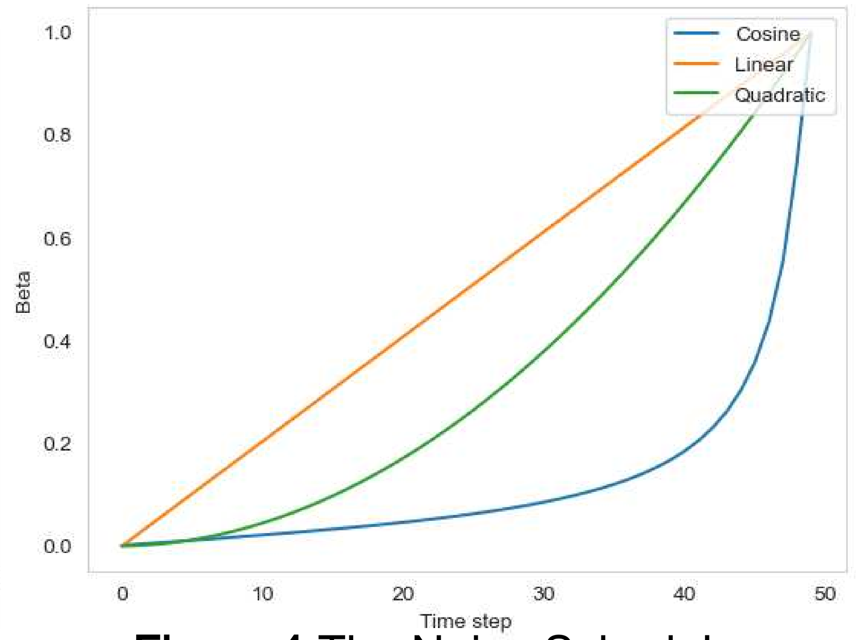

Figure 4 The Noise Schedule

Noise schedule refers to the noise-adding types, to simulate different scenarios in real life, we utilised three types of noise schedules to generate the noise as Figure 4 shows.

Epochs	20 (Early Stage of training)	80 (Late stage of training)
CSDI (Gaussian noise) – RMSE/MAE/CRPS	22.451/11.219/0.124	19.010/9.700/0.110
CSDI (Gamma noise) – RMSE/MAE/CRPS	19.492/10.748/0.123	18.175/9.508/0.108

Table 1 The results of CSDI with Gamma noise and Gaussian noise

This table presents the difference between our proposed imputation methods (Gamma noise) and the original CSDI. Clearly, adding Gamma noise hugely improved the convergence speed of training in the first 20 epochs especially for RMSE and MAE. It also has a slightly better performance after 80 epochs of training.

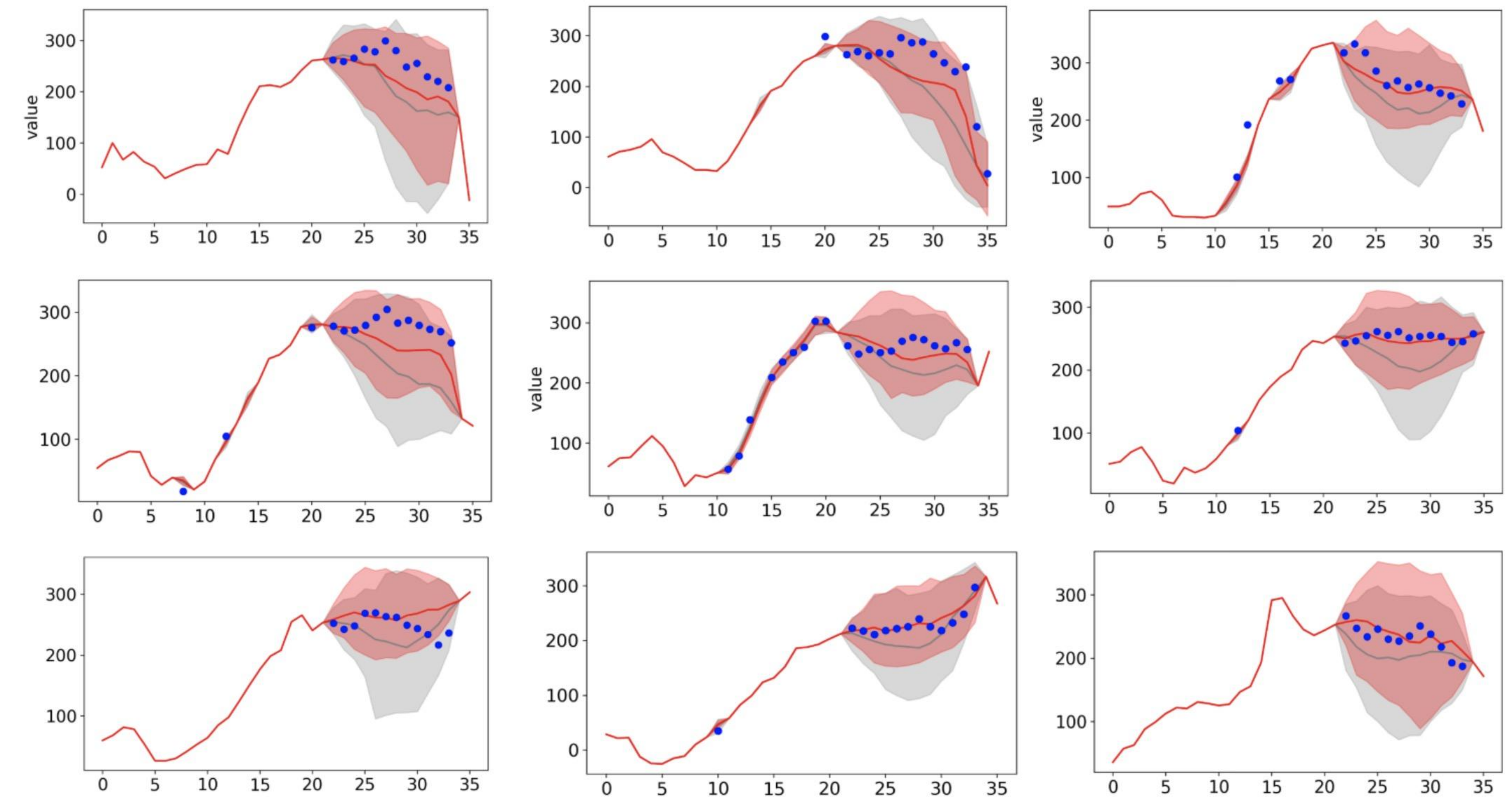

Figure 5 Comparison of Gamma Imputation and original Imputation in CSDI. The blue points show the ground-truth imputation targets. Grey and red lines represent the results of the original CSDI and Gamma CSDI. The shaded areas are confidence intervals.

Figure 5 shows the imputation results of traditional CSDI and our method after 80 epochs of training. In most cases, our model has a smaller confidence interval compared to Gaussian noise. Our method also simulated better quality imputation results which are closer to the ground truth data. These two observations are sufficient to show our model enhances the performance of traditional CSDI, and it can fit the patterns of data distribution better.

References

[1] E. Afrifa-Yamoah, U. A. Mueller, S. M. Taylor, and A. J. Fisher, “Missing data imputation of high-resolution temporal climate time series data,” Meteorological Applications, vol. 27, no. 1, Jan. 2020, doi: https://doi.org/10.1002/met.1873.
[2] W. Pongsena, P. Ditsayabut, N. Kerdprasop, and K. Kerdprasop, “Deep Learning for Financial Time-Series Data Analytics: An Image Processing Based Approach,” International Journal of Machine Learning and Computing, vol. 10, no. 1, pp. 51–56, Jan. 2020, doi: https://doi.org/10.18178/ijmlc.2020.10.1.897.
[3] N. Wagner, Z. Michalewicz, M. Khouja, and R. R. McGregor, “Time Series Forecasting for Dynamic Environments: The DyFor Genetic Program Model,” IEEE Transactions on Evolutionary Computation, vol. 11, no. 4, pp. 433–452, Aug. 2007, doi: https://doi.org/10.1109/tevc.2006.882430.
[4] P. J. García-Laencina, P. H. Abreu, M. H. Abreu, and N. Afonoso, “Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values,” Comput. Biol. Med., vol. 59, pp. 125–133, 2015.
[5] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851, 2020.
[6] Y. Song and S. Ermon, “Generative modeling by estimating gradients of the data distribution,” Advances in neural information processing systems, vol. 32, 2019.
[7] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” arXiv [cs.LG], 2020.
[8] K. Rasul, C. Seward, I. Schuster, and R. Vollgraf, “Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting,” Proceedings of the 38th International Conference on Machine Learning, Jan. 2021.
[9] Y. Tashiro, J. Song, Y. Song, and Stefano Ermon, “CSDI: Conditional Score-based Diffusion Models for Probabilistic Time Series Imputation,” 35th Conference on Neural Information Processing Systems, Dec. 2021, doi: https://doi.org/10.48550/arxiv.2107.03502.
[10] Eliya Nachmani, R. San-Roman, and L. Wolf, “Non Gaussian Denoising Diffusion Models,” Jun. 2021.
[11] Y. Song, C. Durkan, I. Murray, and Stefano Ermon, “Maximum Likelihood Training of Score-Based Diffusion Models,” 35th Conference on Neural Information Processing Systems, Jan. 2021
[12] Eric, Oti, and Francis, “ A study of properties and applications of gamma distribution, ” African Journal of Mathematics and Statistics Studies, vol. 4, no. 2, pp. 52 – 65, 2021.
[13] L. Yan, “Confidence interval estimation of the common mean of several gamma populations,” PLoS One, vol. 17, no. 6, p. e0269971, 2022.
[14] F. Najar, Sami Bourouis, Nizar Bouguila, and Safiya Belghith, “A Comparison Between Different Gaussian-Based Mixture Models,” Oct. 2017, doi: https://doi.org/10.1109/aiccsa.2017.108