International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Investigating the Influence of Feature Normalization on Spoken Language Understanding Performance for the Classification Function

Dr. Sheetal Swapnil Mahadik[1], Mr. Pravin Jangid*[2], Dr. Deven Shah[3]

Submitted: 31/12/2023 **Revised**: 07/02/2024 **Accepted**: 15/02/2024

Abstract: Deep learning models are used for improving the performance of many applications such as image processing, natural language processing, video processing, human-computer interaction (HCI), etc. Commercially available HCI systems such as Apple Siri, Microsoft Cortana, and Alexa incorporate deep learning models to enhance their system performance. One of the tasks of HCI is to classify user utterance to a predefined domain-specific slot such as a request for food, area, etc. This task is accompanied by the spoken language understanding (SLU) unit of HCI. The deep learning model in SLU classifies user utterance to understand the user's intention. The performance of the classification learning model depends upon the quality of the input feature matrix. These feature matrices for SLU are high dimensional and each feature is not on the same scale. Thus, there is no equal contribution from each feature. Therefore, there is a need for applying to feature normalizing techniques to give equal weights to each feature and enhance the classification task in the SLU model. Feature quality in SLU can be improved by pre-processing techniques such as feature normalization, and it will aid to improve the user utterance classification of SLU. The work in this paper investigates the impact of feature normalization techniques on SLU performance for the classification task. The feature normalization techniques investigated for SLU are Z-score, mean-centered, variable stability scaling, min-max normalization, max normalization, decimal scaling normalization, tanh-based normalization, and sigmoidal normalization. The experimentation was done on a publicly available WOZ 2.0 dataset. The feature normalization methods which were more effective in reducing classification error are Z-score and min-max normalization techniques. The less effective techniques in reducing classification errors are decimal scaling, scaling, and log normalization. can result in a page being rejected by search engines. Ensure that your abstract reads well and is grammatically correct.

Keywords: *WOZ 2.0, grammatically, SLU, HCI, experimentation, Z-score, matrix, utterance*

1. Introduction

Normalization is a pre-processing technique applied to many applications. Normalization, in particular, improves the predictive property of the model. Park et al. in paper [1] applied normalization for microarray data. The intensity-based normalization was effective for microarray datasets. In paper [2], a new normalization technique is proposed for financial forecasting data. The proposed normalization which is known as advanced on min-max outperforms the decimal and min-max normalization. Reinhold et al. in the research work [3], experimented with normalization on image data. The results showed that it had improved the image synthesis results. Normalization is also applied to signal processing applications [4]. In another work, normalization is applied to electroencephalograph data [5]. It has improved the performance of the classification task of the brain-computer interface. In paper [6] normalization was used for gene dataset. The research work in the paper investigated normalization for microbiome sequencing datasets [7]. This work displayed good detection capability. Pre-processing normalization aids to improvement in performance in an application such as health record data [8], speaker verification [9], and smoke detection [10]. The work in the paper investigates effective normalization techniques to improve the classification task of SLU.

The learning models are incorporated in many applications, such as computer-based simulation purposes [11]. In another application Recurrent Neural Network (RNN) for disease prediction. Annamalai et al. in their paper [12], incorporated a belief network for face recognition. Thus, machine learning aided the performance of many applications.

This work investigates to are Z-score, mean cantered, Variable Stability Scaling, Min-Max Normalization, Max Normalization, Decimal Scaling Normalization, Tanh Based Normalization, and Sigmoidal Normalization standardize features in SLU. The objective of work in the paper is as follow:

1) To enhance the performance of SLU for the classification task by normalizing features.

2) To investigate six feature normalization techniques from

[1]*Electronic and telecommunication engineering department, Shree L.R Tiwari engineering college, kanakia park, Mira Road (East), Thane, Maharashtra, 401107, India.*
[2]*Computer Engineering department, Shree L.R Tiwari engineering college, kanakia park, Mira Road (East), Thane, Maharashtra, 401107, India*
[3]*Information technology department, Shree L.R Tiwari engineering college, kanakia park, Mira Road (East), Thane, Maharashtra, 401107, India*
* *Corresponding author's Email: sheetal.jagdale2@gmail.com*

machine learning to enhance the performance of SLU.

3) To analyze the impact of feature normalization techniques on SLU performance.

2. Background

2.1. Spoken Language Understanding

Celikyilmaz et al., in their research work [13], used the conditional random field (CRF) technique for SLU. CRF technique leads to improvement of the F-Score then baseline model. In this paper, learning models are used for sequence tagging. The model presented in paper [14], a technique for Non-Sentential Utterances(NSU) classification is proposed. Semi-Supervised learning techniques and additional features are used for NSU classification. A common issue in predicting those classes is that the parallelism with their antecedent is almost entirely at the semantic level. In this paper, machine learning is used for NSU classification. Sarikaya et al. presented a deep learning network for natural language understanding [15]. Deep Bayesian network (DBN) in this work first learns from unlabelled data and the feature discovered by this model is provided input to the feed-forward neural network. Future work is proposed to extend DBN for slot detection and entity tagging in SLU. In paper [16] learning model is used to extract features from unlabelled data.

Macherey et al. presented a statistical translation method for natural language understanding [17]. In this paper, natural language understanding is described as a translation problem. The results demonstrate that the ME-based approach outperforms the source channel-based approach. In the paper, the learning model is used for SMT. In article [18] a deep learning architecture for semantic decoder component of SDS. The future scope proposed is to extend these techniques to multi-domain. In this paper, machine learning is used for semantic representation. Henderson, in his thesis [19], used discriminative techniques for SLU. The future scope of the thesis is to use an active learning algorithm for labeling data and replace the decoder with the word tracking system. In this thesis, learning is for semantic representation. Celikyilmaz et al. in paper [20] has presented a model that addresses both problems by using semi-supervised learning. The author proposes an entire history of conversation for semantic tagging. This semantic model is used for Semantic tagging and using this method domain adaptation performance was also improved. SLU task involves extracting information from a spoken query. In another research work [21] learning model is used for slot filling

2.2. Techniques for Feature normalization:

Feature normalization is a pre-processing procedure that scales the features to the new range or changes the features. This task is important for equal contribution from each feature. Normalization techniques can be applied to features to enhance model performance. Normalization methods apply scaling and transformation techniques.

Normalization techniques that apply to scale are min-max normalization, mean center, z–score Pareto scaling, and max normalization [22][23][24]. This technique scales the features to a predefined range. In paper [25], three normalization techniques have been experimented with using the decision tree model. The methodology used in the paper is to experiment with decimal, z score, and min-max normalization for different training data. The result displayed the highest accuracy and less growing time for the decision tree with a min-max algorithm. Berg et al. applied the normalization technique to biological data [26]. The proper selection of data pre-processing techniques has an impact on biological data. Normalization techniques that apply transformation are power transform, tanh, and sigmoidal normalization [27][28][29].

3. Proposed Methodology

Feature normalization assigns equal contribution from each feature attribute. This enhances the prediction capability of the model—the work extent these feature normalization techniques to improve the classification task of SLU. The three-input feature vector representing slot value pair, user utterance, and earlier conversation is first input to the normalization unit. Normalization will give an equal attribute to each feature. The normalized feature vector is fed to convolution layers. Feature vector representing user utterance and slot value pair is fed to sematic decoder. It will map user utterance to a semantic form. The output of the semantic decoder is combined with context to generate a summary. The next classifier will classify the user into a slot value pair

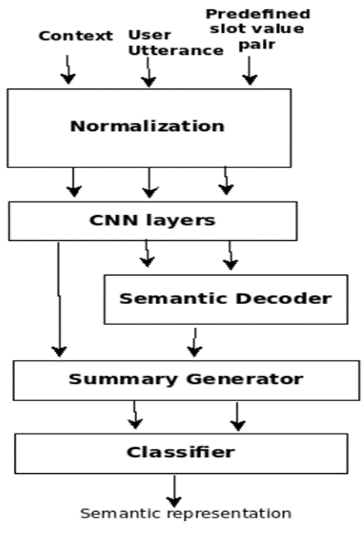

Fig. 1 Proposed methodology for SLU with normalization

The experiment is performed on the WOZ dataset. The normalization techniques that are investigated are Z-score, mean-centered, variable stability scaling, min-max normalization, max normalization, decimal scaling normalization, tanh-based normalization, and sigmoidal normalization. The evaluation parameter for experimentation is F score, precision, recall, detection rate, and accuracy. First, the WOZ dataset, which consists of 00 training dialogues, was used for training and validation. The experiment was conducted for the slot area. Then testing was done for the slot area with 200 dialogues. All evaluation parameters were calculated. The learning curve for the area slot was obtained for a different number of dialogues in the training set. Next same experimentation was repeated for the price range and food slot.

Feature normalization assigns equal contribution from each feature attribute. This enhances the prediction capability of the model—the work extent these feature normalization techniques to improve the classification task of SLU. The three-input feature vector representing slot value pair, user utterance, and earlier conversation is first input to the normalization unit. Normalization will give an equal attribute to each feature. The normalized feature vector is fed to convolution layers. Feature vector representing user utterance and slot value pair is fed to sematic decoder. It will map user utterance to a semantic form. The output of the semantic decoder is combined with context to generate a summary. The next classifier will classify the user into a slot value pair.

4. Results and Discussion:

A normalization experiment was conducted on WOZ 2.0 dataset on the baseline model proposed in the paper [30]. The evaluation parameters are F score, precision, recall, detection rate, and accuracy. The experimentation was done for the area, price, and food slot.

Precision:

Table 1 shows the precision values for SLU with and without feature normalization. The precision for slot area and price is improved with feature normalization.

Table 1. Precision for SLU

Sr. No	Normalization Method for SLU	Precision for slot area	Precision for slot price
1	SLU without normalization	0.992	**0.988**
2	Z-score normalization	0.995	0.988
3	Tanh based normalization	0.998	0.982
4	Sigmoidal normalization	0.996	0.982
5	Min-max normalization	0.996	0.989
6	Log	0.995	0.99
7	Decimal Scaling normalization	1	0.988
8	Scaling normalization	0.998	0.993
9	Mean Cantered normalization	0.989	0.983
10	Max norm normalization	0.995	0.982

Recall:

Table 2 shows the recall values for SLU with and without feature normalization. The recall for slot area and price is better with feature normalization

Table 2. Recall for SLU

Sr. No	Normalization Method for SLU	Recall For Area slot	Recall For Price Slot
1	SLU without normalization	0.900	0.94
2	Z-score normalization	0.955	0.952
3	Tanh based normalization	0.952	0.961
4	Sigmoidal normalization	0.916	0.947
5	Min-max normalization	0.995	0.953
6	Log	0.956	0.93

7	Decimal Scaling normalization	0.936	0.933
8	Scaling normalization	0.916	0.874
9	Mean Centered normalization	0.942	0.958
10	Max norm normalization	0.954	0.958

Classification error rate:

Table 3 shows the classification error for SLU with and without feature normalization. The classification error for slot area and price is reduced with feature normalization. The set of feature normalization techniques that reduced the classification error significantly are Z-score, max-min, and normalization. The set of feature normalization in which classification error was not reduced are decimal scale, scaling, and log normalization.

5. Conclusion:

HCI systems are a widely used system where users can interact with the computer. It is important to understand the user's motive. SLU assists the HCI system to understand the user's motive. SLU performs this task by classifying user utterance to predefined slots such as area, food, and price, etc. Thus, the performance of SLU depends majorly on the classification model. The classification model is influenced by the input feature matrix dimension and range of values. Therefore, normalization can be applied to scale or transform feature vectors so that they contribute equally. This work in the paper investigated feature normalization methods for SLU. The investigation had two important findings. The first finding is that the SLU classification task is improved by feature normalization. The second finding is that the classification error rate is reduced by feature normalization. The most effective method for reducing classification errors are Z-score normalization and min-max normalization. The less effective methods for reducing classification errors are decimal scaling, scaling, and log normalization.

Author contributions

Dr. Sheetal Mahadik: Conceptualization, Methodology, Software, Field study Writing-Original draft preparation, Software, Validation **Mr. Pravin Jangid:** Data curation, Field study **Dr. Deven Shah:** Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Park, T., Yi, S. G., Kang, S. H., Lee, S. Y., Lee, Y. S., & Simon, R. (2003). Evaluation of normalization methods for microarray data. BMC Bioinformatics, 4, 1–13. https://doi.org/10.1186/1471-2105-4-33

[2] Patro, S. G. K., & sahu, K. K. (2015). Normalization: A Preprocessing Stage. Iarjset, 20–22. https://doi.org/10.17148/iarjset.2015.2305

[3] Reinhold, J. C., Dewey, B. E., Carass, A., & Prince, J. L. (2018). Evaluating the Impact of Intensity Normalization on MR Image Synthesis. ArXiv. https://doi.org/10.1117/12.2513089

[4] Cochran, D. (1988). Consequence of Signal Normalization in Spectrum Analysis. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2388–2391. https://doi.org/10.1109/icassp.1988.197121

[5] Pathak, J., Bailey, K. R., Beebe, C. E., Bethard, S., Carrell, D. S., Chen, P. J., Dligach, D., Endle, C. M., Hart, L. A ., Haug, P. J., Huff, S. M., Kaggal, V. C., Li, D., Liu, H., Marchant, K., Masanz, J., Miller, T., Oniki, T. A., Palmer, M., … Chute, C. G. (2013). Normalization and standardization of electronic health records for high-throughput phenotyping: The sharpn consortium. Journal of the American Medical Informatics Association, 20(E2). https://doi.org/10.1136/amiajnl-2013-001939

[6] Giricz, O., Lauer-Fields, J. L., & Fields, G. B. (2008). The normalization of gene expression data in melanoma: Investigating the use of glyceraldehyde 3-phosphate dehydrogenase and 18S ribosomal RNA as internal reference genes for quantitative real-time PCR. Analytical Biochemistry, 380(1), 137–139. https://doi.org/10.1016/j.ab.2008.05.024

[7] Chen, L., Reeve, J., Zhang, L., Huang, S., Wang, X., & Chen, J. (2018). GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. PeerJ, 2018(4), 1–20. https://doi.org/10.7717/peerj.4600

[8] Pathak, J., Bailey, K. R., Beebe, C. E., Bethard, S., Carrell, D. S., Chen, P. J., Dligach, D., Endle, C. M., Hart, L. A., Haug, P. J., Huff, S. M., Kaggal, V. C., Li, D., Liu, H., Marchant, K., Masanz, J., Miller, T., Oniki, T. A., Palmer, M., … Chute, C. G. (2013). Normalization and standardization of electronic health records for high-throughput phenotyping: The sharpn

consortium. Journal of the American Medical Informatics Association, 20(E2). https://doi.org/10.1136/amiajnl-2013-001939

[9] Wei Rao, Man-Wai Mak, Kong-Aik Lee " Normalization of total variability matrix for i-vector/PLDA speaker verification", Published 2015 Mathematics, Computer Science 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[10] Yin, Z., Wan, B., Yuan, F., Xia, X., & Shi, J. (2017). A Deep Normalization and Convolutional Neural Network for Image Smoke Detection. IEEE Access, 5, 18429–18438.

[11] Talasila, V., Madhubabu, K., Mahadasyam, M. C., Atchala, N. J., & Kande, L. S. (2020). The prediction of diseases using rough set theory with recurrent neural network in big data analytics. International Journal of Intelligent Engineering and Systems, 13(5), 10–18.

[12] Sumpavakup, C., Mongkoldee, K., Ratniyomchai, T., & Kulworawanichpong, T. (2020). Touch and step voltage evaluation based on computer simulation for a mass rapid transit system in Thailand. International Journal of Intelligent Engineering and Systems, 13(5), 159–169.

[13] Kusuma, E. J., Shidik, G. F., & Pramunendar, R. A. (2020). Optimization of Neural Network using Nelder Mead in Breast Cancer Classification. International Journal of Intelligent Engineering and Systems, 13(6), 330–337.

[14] Fibriani, I., Widjonarko, W., Prasetyo, A., Raharjo, A., & Irawan, D. (2020). Multi Deep Learning to Diagnose COVID-19 in Lung X-Ray Images with Majority Vote Technique. International Journal of Intelligent Engineering and Systems, 13(6), 560–568.

[15] Annamalai, P. (2020). Automatic face recognition using enhanced firefly optimization algorithm and deep belief network. International Journal of Intelligent Engineering and Systems, 13(5), 19–28.

[16] Celikyilmaz, A., 2010, "Convolutional Neural Network Based Semantic Tagging with Entity Embeddings," Microsoft research , pp. 1–7.lligent Engineering and Systems, 13(6), 560–568.

[17] Dragone, P., 2015, "Non-Sentential Utterances in Dialogue: Experiments in Classification and Interpretation." In Proc. of the 19th Workshop on the Semantics and Pragmatics of Dialogue.

[18] Sarikaya, R., Hinton, G. E., and Deoras, A., 2014, "Application of Deep Belief Networks for Natural Language Understanding," IEEE Trans. Audio, Speech Lang. Process., 22(4), pp. 778–784.

[19] Sarikaya, R., Celikyilmaz, A., Deoras, A., Jeong, M., and Corporation, M., 2011, "Shrinkage Based Features for Slot Tagging with Conditional Random Fields," Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2011, 1309-1312.

[20] Macherey, K., Bender, O., Ney, H., and Member, S., 2009, "Applications of Statistical Machine Translation Approaches to Spoken Language Understanding," IEEE Transactions on Audio, Speech, and Language Processing (Volume: 17, Issue: 4, May 2009) 17(4), pp. 803–818.

[21] Barahona, L. M. R., Gasic, M., Mrkšić, N., Su, P.-H., Ultes, S., Wen, T.-H., and Young, S., 2016, "Exploiting Sentence and Context Representations in Deep Neural Models for Spoken Language Understanding," Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, The COLING 2016 Organizing Committee, Osaka, Japan, pp. 258--267.

[22] Henderson, M. S., 2015, "Discriminative Methods for Statistical Spoken Dialogue Systems," Phd thesis University of Cambridge.

[23] Celikyilmaz, A., 2013, "Enriching Word Embeddings Using Knowledge Graph for Semantic Tagging in Conversational Dialog Systems." AAAI Spring Symposium.

[24] Sarikaya, R., Celikyilmaz, A., Deoras, A., Jeong, M., and Corporation, M., 2011, "Shrinkage Based Features for Slot Tagging with Conditional Random Fields," Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2011, 1309-1312.

[25] Luai Al Shalabi, Zyad Shaaban, Normalization as a preprocessing enginefor data mining and the approach of preference matrix, in: Interna-tional Conference on Dependability of Computer Systems, IEEE, 2006, pp.207–214.

[26] 2Robert A van den Berg, Huub CJ HoefslootJohan A Westerhuis, Age K Smilde, Mariët J van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMCGenomics 7 (1) (2006) 142.

[27] Shih-Wei Lin, Kuo-Ching Ying, Shih-Chieh Chen, Zne-Jung Lee, Particle optimization for parameter determination and feature selection ofsupport vector machines, Expert Syst. Appl. 35 (4) (2008) 1817–1824.

[28] Luai Al Shalabi, Zyad Shaaban, Normalization as a preprocessing enginefor data mining and the approach of preference matrix, in: Interna-tional Conference on

Dependability of Computer Systems, IEEE, 2006, pp.207–214.

[29] Robert A van den Berg, Huub CJ HoefslootJohan A Westerhuis, Age K Smilde, Mariët J van der Werf, Centering, scaling, and transformations:improving the biological information content of metabolomics data, BMCGenomics 7 (1) (2006) 142.

[30] Antonio Reverter, Wes Barris, Sean Mcwilliam, Keren A Byrne, Yong HWang, Siok H Tan, Nick Hudson, Brian P Dalrymple, Validation of al-ternative methods of data normalization in gene co-expression studies,Bioinformatics 21 (7) (2004) 1112–1120.

[31] Olav M. Kvalheim, Frode Brakstad, Yizeng Liang, Preprocessing of analyticalprofiles in the presence of homoscedastic or heteroscedastic noise, Anal.Chem. 66 (1) (1994) 43–51.

[32] Kevin L. Priddy, Paul E. Keller, Artificial neural networks: an introduction,vol. 68, SPIE press, 2005.