Automated COVID-19 dialogue system using a new deep learning network

Khaldoon H. Alhussayni[1], Eman S. Alshamery[2]
[1] Computer Center, University of Babylon, Babylon, Iraq
[2] Software Department, College of Information Technology, University of Babylon, Babylon, Iraq

ABSTRACT

During the coronavirus disease 2019 (COVID-19) pandemic outbreak, it is necessary to apply social distancing measurements and search for an alternative to physical contact due to the spread of viral infection. The interest in task-oriented dialogue systems has grown remarkably in healthcare, using natural language in the dialogue between patients and doctors. However, the doctor's advice is implicit and unclear in most conversations, and the patient may also be nervous when describing symptoms or may have difficulty describing them. Therefore, the patient's description of symptoms is insufficient for a diagnosis by doctors. This study aims to provide suitable medical advice based on the patients' symptoms during the conversation between doctors and patients by proposing a new deep learning method for automated medical dialogue systems. The model is based on an encoder and two stages of learning to make reliable decisions. The encoder extracts important words using text normalization, resulting in two vectors: symptom vectors and doctor utterance vectors. The symptom vectors are represented as a weighted bag-of-words feature. The first stage is used to cluster the patients' utterances by applying Hopfield network while considering the semantic similarity, whereas the second stage extracts an implicit label as a template of advice using clustering. Additionally, the external evaluation model used the applied feedforward neural network classification algorithm using labels obtained in the second stage. The CovidDialog-English dataset is used to evaluate the model. The experimental results indicate the high performance of the feedforward neural network with an F1-score of 0.972 and presents a comparison of three clusters using the k-nearest neighbours and naïve Bayes-based models.

Keywords: Deep Learning, Healthcare, Hopfield network clustering, Natural Language Processing, Task-Oriented Dialog Systems.

Corresponding Author:
Khaldoon H. Alhussayni
Computer Center
University of Babylon
Babylon, Iraq
E-mail: alhussayni@uobabylon.edu.iq, alhussayni83@gmail.com

1. Introduction

Under the conditions of the coronavirus disease 2019 (COVID-19) pandemic, the arrival of patients at hospitals is harmful and has a high risk for transinfection. Transinfection refers to the transmission of the coronavirus from patients who frequent clinics to healthy people. It is necessary for people who suffer from symptoms, such as a fever or difficulty breathing, to consult doctors and seek medical advice. The patients and doctors engage in dialogue in natural language. Given that physical contact has high risk due to the spread of the viral infection, it is necessary to apply social distancing measurements and search for an alternative to this form of interaction.

In recent years, medical dialogue systems have been increasingly required, even more than systems for booking films or airline tickets [1], [2], finding hotels or restaurants [3], [4], or obtaining details from the internet. Medical dialogue systems have proved to have several advantages in the last two decades in diagnosing, monitoring, or supporting treatment [5]. They provide cost-effective, scalable, and personalized medical assistance solutions that can be offered any place or time through web-based or mobile applications [6], [7] as digital interventions.

The Hopfield network [8] offers a valuable context for this research. The proposed self-aggregation network uses the Hebb rule to encode pattern vectors in this work. One advantage of linked memory networks in the Hopfield network is that they can be applied to solve hard problems in combination. The Hopfield network was first studied by [8], [9], who concluded that the recovery of the memory contained in the Hebbian matrices is assured to a *p*-value that is a crucial fraction of the number of network nodes. Applications range from combination optimization to image reconstruction, including various control engineering optimization issues in robotics and content-addressed memory systems.

Clustering is automatically gathered into predefined numbers of clusters based on their (dis-)similarity. In vector space, the dissimilarity or distance is resolved using either a distance metric or a similarity metric, such as the *Euclidean distance* [10], [11] and *cosine similarity* [12]. There are assorted types of clustering approaches, including hierarchical [13], density-based [14], and model-based methods [15]. K-means [16], due to its flexibility and efficiency, is one of the most widely used clustering algorithms. It is also commonly used as a baseline clustering algorithm in text clustering experiments, as it has low complexity and reasonably good efficiency when paired with an effective distance metric.

The proposed model can recognize unknown patterns without training, making it very easy to generalize the model for another dataset in different domains. There are two stages of learning used throughout the proposed model. The first stage involves clustering the patients' utterances to group similar symptoms, determine the advice, and evaluate the clustering learning using silhouette analysis. The second stage extracts a template for each cluster as a label and then measures errors in the percentage using three classification algorithms (feedforward neural network, k-nearest neighbours (KNN), and naïve Bayes) to measure the error using the F1-score.

This paper is structured as follows. Section 2 explains the theoretical background, including the Hopfield neural network in Section 2.1, natural language processing (NLP) in Section 2.2, and text preprocessing in Section 2.3. Section 3 describes the proposed methodology, including the dataset described in Section 3.1, the automated COVID-19 dialogue system model architecture containing the encoder model in Section 3.2, and the learning model in Section 3.3. Section 4 discusses the methodology and results. Finally, Section 5 concludes this paper.

2. Theoretical background

2.1. Natural language processing

Natural language processing (NLP) is an approach to how machines can process or comprehend human languages to carry out useful tasks. In addition, NLP integrates computer linguistics, computer science, cognitive science, and artificial intelligence in an interdisciplinary area. Therefore, NLP sustains human language comprehension and development and is concerned with developing innovative functional technologies to promote computer-language interaction. Several NLP applications exist, including speech recognition, natural language understanding, dialogue systems, question answering, sentiment analysis, natural language generation, and natural language summarization [17].

2.2. Hopfield neural networks

The Hopfield neural network [8] consists of one or more fully associated recurrent neurons per layer and generally operates using self-association and optimization. The input and output patterns may be represented as discreet, binary (0,1), or bipolar (+1, –1) in nature, as the main difference lies in the activation function. The key is to determine its weight under stable conditions. The weight of the device is symmetric (i.e., $w_{ji} = w_{ji} = 0$). Weights can be updated during the training of the Hopfield network, and the updates can be made using the following partners for bipolar input patterns:

For a set of binary patterns $s(p)$, $p = 1$ to P, where $s(p) = s_1(p), s_2(p), \ldots, s_i(p), ..., s_n(p)$, and n = the number of network nodes (length of pattern).

The weight matrix is given by the following:

$$w_{ij} = \sum_{p=1}^{p} s_i(p)\ s_j(p) \quad for\ i \neq j \tag{1}$$

A connection matrix is simply a dyadic form given by *W* in Equation (1) to store one pattern. The prescription for storing an idea is obtained from [18], where it is stated that the change in synaptic transmission is proportional to the pre- and post-synaptic neuron signals. For each matrix product, learning the 'Hebbian rule' is the method in which it is properly calculated, as shown below:

$$w_{ij} = \frac{1}{N}\sum_{p=1}^{p} s_i(p)\ s_j(p) \quad for\ i \neq j \tag{2}$$

The pattern itself (or an incomplete or noisy representation) is calculated using the dot product of the pattern vector and the weight matrix to obtain a kept pattern from the Hopfield network. This calculation produces a new vector pattern that is binarized and returned to the system:

$$p_{i+1} = sgn(w\ p_i) \tag{3}$$

$$sgn(p_i) = \begin{cases} +1\ , p_i\ \geq 0 \\ -1\ , p_i\ < 0 \end{cases} \tag{4}$$

The procedure is replicated until the pattern vector is unchanging or after a specified number of iterations have been completed. For easy implementation, it would normally be appropriate to stop after five steps. However, a cost function is associated with the Hopfield network. The issue is that this functionality is not truly a cost function because, unlike typical cost functions, it is not a function of network weights but a function of network states (the values of the neurons).

The error between the real network output given by a sample and the desired output from the training set is generally calculated using a neural network cost function. The idea is to minimize this function using gradient descent. The training set is not labelled in the Hopfield network, as it just needs to examine and memorize certain patterns. Therefore, as nothing indicates the real label for something, there is no notion of 'error'. It is similar to the human brain process: experiences are normally memorized. They are already seen, and that is how they are learned and recalled later.

The energy of a Hopfield network is a decision made by recalling the energy function in Equation (5), which represents the sum of any neuron times its weighted sum. Any neuron flip minimizes the overall energy of the network. If a network has *N* neurons, then the energy function would be a dimensional function of *n*. The energy function is also called the Lyapunov function in the theory of dynamical systems [19]–[21]

$$E\left(p_{ij}\right) = -0.5 \sum w_{ij}\, p_i\, p_j \tag{5}$$

2.3. Text preprocessing

Preprocessing text strives to obtain important features or keywords from text documents to determine the importance of words and documents and the relevance of words within each document. A powerful preprocessor represents the document efficiently to store the document with a strong recall rate for handling requests (precision and recall). This process is the most important and complex, resulting in a collection of index words describing each text. Text preprocessing is referred to as tokenized text or standardized text.

Preprocessing is a technique that can be mostly separated into various text operations, such as stopping words, stemming, N-gram models, and keyword weighting. The most commonly used words in English are valueless (i.e., pronouns, prepositions, and conjunctions). Such words are called stop words. The first step is to remove the stop words, which is considered extremely important [22]. The root of the word is found using stemming techniques. Stemming transforms the vocabulary into its stems by removing the affixes (i.e., suffixes or prefixes) of a word [23]. The N-gram models are an intra-specified sequence of *n* consecutive elements derived from a broader sequence (e.g., bigrams when *n* = 2 or trigrams when *n* = 3). In N-grams, items typically contain either letters or words. Term weighting is a very critical principle that determines whether learning algorithms perform or fail. Although various words are of different significance levels in the document, the term weight is a significant indicator for each term [24].

3. Proposed methodology

The proposed model for generating medical advice contains two parts: the *encoder model* and *learning model*, as illustrated in Figure 1.

3.1. Dataset

The dataset used in this work is the Covid Dialog-English dataset collected by [25]. It consists of 604 English COVID-19 and other associated pneumonia conversations, with 1,232 statements and 90,664 tokens (English words). As for the conversation structure, each appointment begins with a summary of a patient's medical problems, followed by a dialogue between the patient and doctor.

3.2. Encoder model

To represent the important words, the first step in the model was processing the original data. Therefore, the raw text underwent several text normalization methods, including text tokenization, the removal of non-standard words (e.g., number, acronyms, abbreviations, and stop words). Another common technique of NLP is lemmatization. Its role is to find the root of all the words derived from their base form. This method was used to reduce the cardinality of vocabularies and simplify tasks, such as finding the cluster topic.

The results of the text normalization model were represented in two types of vectors: symptom vectors and doctor utterance vectors. The symptom vectors were represented as weighted using the bag-of-words feature vector, including the patient's descriptions and the previous utterance of patients. Each vector contains weighted information, the presence (1) or absence (0) of words representing the vector using a fixed length for all vectors, where each zero in the vector is converted to (-1) to adapt the vectors as input to the Hopfield neural network. Only the symptoms mentioned in the text were considered, and all other words were excluded because, when giving advice, doctors only rely on symptoms that patients experience.

The doctor utterance vectors were also represented as a weighted bag-of-words feature vector after the text normalization stage. Then, the k-means clustering algorithm was applied to group similar advice, considering the advice to be a label of patient symptoms. The labels were used to evaluate the external performance of the models.

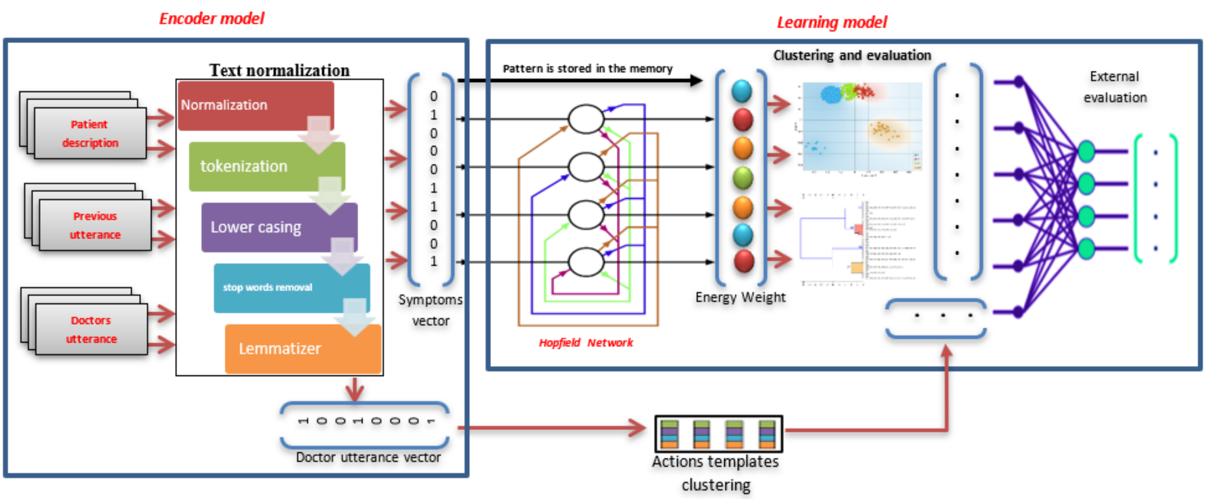


Figure 1. Automated COVID-19 dialogue system model architecture

4.1. Learning model

The second part of the proposed method is the learning model. The Hopfield network algorithm was applied to learn all patterns that represent the symptom vectors. Hopfield network algorithm computed the energy value of each pattern. The learning of patterns was only applied whenever the pattern was unknown to the Hopfield network (i.e., not stored in the network memory or not recognized yet). Otherwise, the energy value was calculated directly without training.

The proposed model achieves network outcomes with one piece of medical advice for each cluster by clustering a similar pattern. This method increases the network's ability to learn and recognize unknown patterns and decreases training time. In addition, the stored memory capacity of the patterns was reduced by 40% to provide more capacity for the Hopfield network memory. The network was designed by fixing the N x N recurrent connections given in Equation (2) and assigning zeros as initializing values. All patterns, denoted by P in Equation (2), were given as the network input.

Through learning, the dynamics of the networks were observed until a balance was reached. The energy value (also called the weight) was computed for each pattern. The learned patterns were stored in the weight matrix, and each pattern was compared with the original one at each bit. The number of incorrect bits was measured. The probability that the entire vector was recovered exactly was considered. The matrix weight W determines the weights between the network nodes and the features of N patterns. To normalize the weight matrix, the number of patterns was split, setting the diagonal elements to 0.

The energy values represent the cost function values of the Hopfield network. Therefore, each pattern has value. Clustering patterns consider the energy values to determine the medical advice of the cluster. Hierarchical clusters and k-means algorithms were applied to the proposed model to evaluate the result performance, and the k-means++ algorithm was used to initialize the values of the k-means algorithm. Both the Euclidean distance metric and the normalized values were used for the hierarchical cluster algorithm.

Four clusters were selected to be advice templates. The N-gram technique considered two words, representing the most-frequent top two words in the text preprocessing with more informative terms in the symptoms and advice for COVID-19. The semantic synonyms (e.g., *stay home*, *self-quarantine or self-isolate*) were also considered in this process.

Table 1: Samples of each cluster based on the patient's descriptions

Index	Patient description	Cluster id
439	Good day. I have a body temperature of 38,39 degrees Celsius accompanied by headaches and a cough. I've been tested for covid, but it's negative. I have trouble urinating and have a pain in my penis as well as my anus. Thank you?	C1
584	Hi I am 39 years old and have flu like symptoms as of yesterday. Runny nose, sore throat. No fever yet. I returned from Germany 19 days ago. Should I get tested for cov19?	C1
533	If you feel not quite well 3 or 4 evenings in a row, no fever, minor dry cough, little stuffy, maybe scratchy throat very minor, just crummy feeling in general and evenings only (feel fine during the day) could that be a mild case of Coronavirus? Thanks.	C2
3	I have chills, breathing problems and cough with white phlegm. Could it be Coronavirus infection?	C2
306	Nasal congestion and feeling a bit fluey?	C4
545	I travelled to Mauritius and do not have symptoms. Should I get tested for covid19?	C4
573	I have mild irritation in my chest but I am not coughing, I just feel a tingling sensation on my thought. I have no fever, no aches should I be worried about COVID 19?	C3

Finally, the proposed model was evaluated using the silhouette score and cross-validation data in the network. In addition, for a more accurate evaluation, three classification algorithms were used based on the energy clustering values of the Hopfield network: the feedforward neural network, KNN, and naïve Bayes.

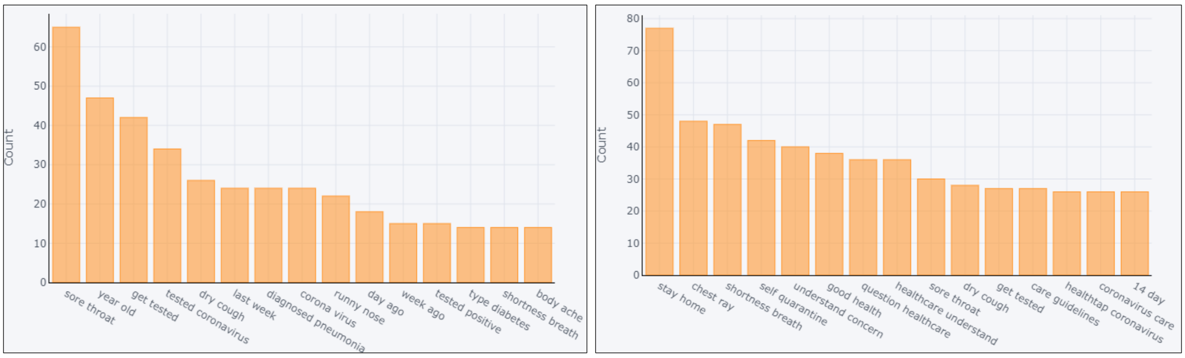
5. Discussion and results

As part of the procedure, patients were asked to describe the symptoms they experienced. Based on the domain terms used in their description, about 40 symptoms were extracted from the texts. Most advice and symptoms in the database text have a length of two words, as presented in Figure 2(a,b): *stay home, get tested, sore throat*, and *dry cough*. Therefore, the N-gram technique was considered to determine the subject of the cluster, where *n* represents two words because the one-gram does not make any sense for both advice and symptoms.

get tested, sore throat, and *dry cough*. Therefore, the N-gram technique was considered to determine the subject of the cluster, where *n* represents two words because the one-gram does not make any sense for both advice and symptoms.

The clustering results were visualized using the hierarchical (dendrogram) technique, as illustrated in Figure 3(a). The results present two clusters by the indices of each pattern. One cluster with (0, 1, 2) indices contains the *viral infection* topic cluster, and the second cluster with (3, 4) indices contains the *stay home* or *self-quarantine* topic cluster.

A random pattern with index (306) in the dataset and index (5) in Figure 3(b) was added to the Hopfield network after checking the clusters regarding whether they were correctly added (i.e., whether the new pattern was appended to the cluster with appropriate medical advice compared to



(a) The patient's description (b) The doctor's response

Figure 2 Top 15 two-grams in dialogue text after normalize text

the doctor's response in the dialogue database). The results reveal that the new pattern collected with the cluster of the *self-quarantine* topic has the same medical advice in the desired output (action template). The same process is repeated, but this time the random selection chose a pattern with index (3) in the dataset and index (6) in Figure 3(c), and it outperformed the results. As expected from the proposed model, the results were achieved by clustering most of the patient's descriptions, with each cluster containing similar medical advice, depending on the description given by the patient, as listed in Table 1.

The first cluster introduced numerous pieces of medical advice focused mostly on patients not infected by COVID-19, which means that the patient did not need to self-quarantine yet. According to the patient symptoms, the patients may have been infected by a viral disease, as shown in the sample with indices 439 and 584.

Most instances of *self-quarantine* were grouped in the second cluster. The most recurring medical advice given by doctors through the dialogue included '*stay home to be socially isolating*', '*monitor your temperature*', and '*rest and drink fluids*'.

The third cluster varied in identifying patterns throughout the network with general and non-specific questions about the coronavirus pandemic disease. Most uninfected people suffer from anxiety or fear for various reasons, such as travelling to a high-risk area, communicating with a person later found to be infected, or asking about symptoms that distinguish COVID-19 from other diseases (e.g., pneumonia, influenza, and colds) due to the similarity between the symptoms and the difficulty of diagnosing COVID-19.

Questions from patients with pneumonia and tuberculosis were added to the fourth cluster. Patients asked about pneumonia and tuberculosis symptoms and their effects on those with COVID-19 disease. The appropriate treatment for each case was suggested.

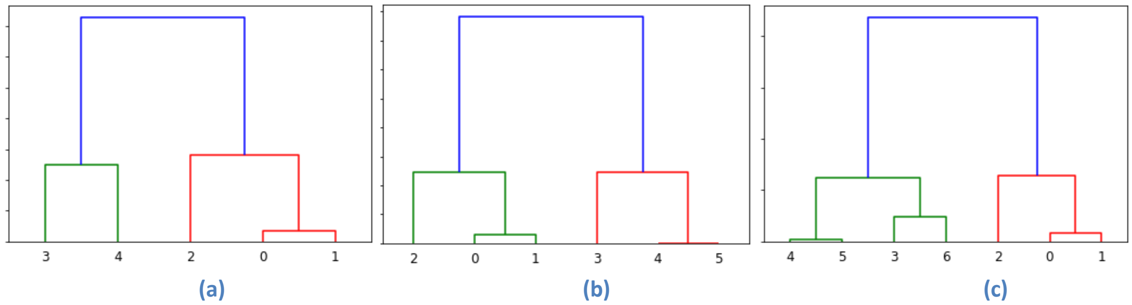


Figure 3 plot of dendrogram clusters as following: (a) first five (0-4) patterns clustering, (b) appending new pattern with number 5 to the network, (c) appending new pattern with number 6 to the network

Figure 4 represents the hierarchical clustering for all patterns, which generates a general idea of the results. The silhouette analysis is useful to analyse the distance of separation between the clusters to select an optimum value for the number of clusters. Figure 4(b) illustrates how close each point in a single cluster is to points in the surrounding clusters and offers a way to evaluate the parameters that are visually similar to other clusters. The range of this measure is [-1, 1]. Silhouette coefficients near +1 mean that the sample is far from the surrounding clusters. A value of zero implies that the sample is at or close to the judgement border between two clusters and that -1 samples may have been allocated to an incorrect cluster.

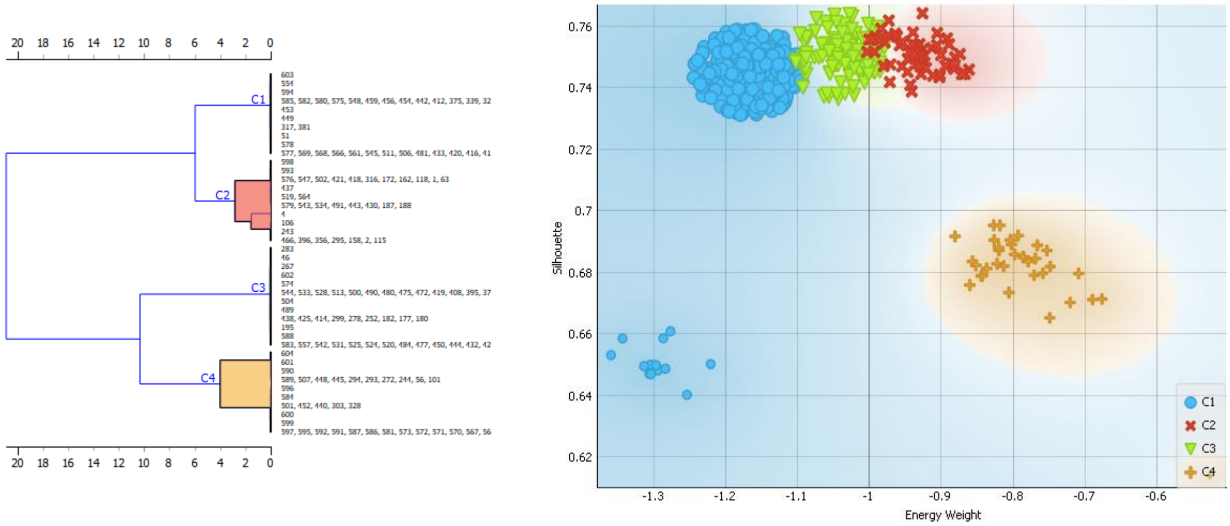
Table 2. Average silhouette score per cluster

No of clusters	Silhouette score
2	0.796
3	0.896
4	0.947

Table 2 presents the silhouette score per number of clusters, being 0.796, 0.896, and 0.947 for Clusters 2, 3, and 4, respectively. The cluster values of Silhouettes 2 and 3 are a poor selection for the given data due to the inclusion of clusters of lower-average figures and the large variations in the number. In the case of decisions between 3 and 4, the silhouette analysis is more comparable, but it is best to use four clusters. The cohesion rates are relatively higher due to the similarity between the symptoms of influenza and COVID-19, which explains the use of external evaluations through classifications.

Another approach used to evaluate the proposed model is labelling data using doctor utterances for each piece of dialogue in the dataset. The k-means clustering algorithm was used to label the data while maintaining the dialogue index because the evolution model considers the index. According to the labelling results, three classifier algorithms (feedforward neural network, KNN, and naïve Bayes) were used to evaluate the performance of the designed model. The F1-score, precision, and recall were used to measure the differences between the samples predicted by an estimator.

The results of each algorithm were compared with their ground truth to measure the accuracy of the estimates. It introduces the association between precision (*P*) and recall (*R*) and achieves the maximum value at one and the worst value at zero [26].



(a) Hierarchical clustering for all patterns (b) K-Means clustering for all patterns

Figure 4 clustering for all patterns

The following are the defined pairs of items in each cluster to compute the F1-score from the precision and recall pairs:

- True positive (*TP*): the number of elements that belong to the same cluster and the same class,
- False positive (*FP*): the number of elements that belong to the same cluster and a different class,
- True negative (*TN*): the number of elements that belong to a different cluster and a different class,
- False negative (*FN*): the number of elements that belong to a different cluster and the same class.

Then, *P*, *R*, and the F1-score are calculated as follows:

$$P = \frac{TP}{TP + FP} \quad , \quad R = \frac{TP}{TP + FN} \quad , \quad F1 - score = \frac{2PR}{P + R}$$

The results indicate that the neural network outperforms both the KNN and naïve Bayes for two, three, four, five, and six clusters, as illustrated in Figure 5.

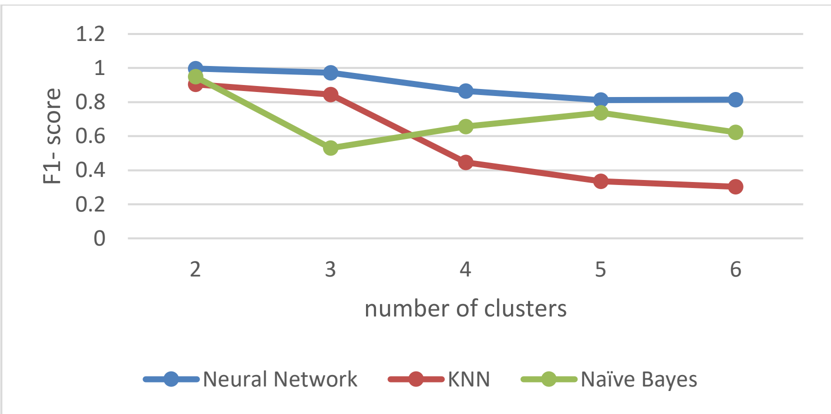


Figure 5. F1-scores per number of clusters for the feedforward neural network, naïve Bayes, and KNN.

Unlike the silhouette score evaluation, the best values of the F1-score in the external evaluation using three clusters are presented in Table 3. With the feedforward neural network estimator being 0.972, the obtained precision and recall have the same value (0.974), whereas the KNN and naïve Bayes estimators performed less efficiently with two, three, and four clusters (except for naïve Bayes, which obtained a high F1-score with two clusters).

Table 3. Evolution for feedforward neural network, naïve Bayes, and KNN algorithms using two, three, and four clusters

	Neural Network			**KNN**			**Naïve Bayes**		
No of clusters	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall
2	0.879	0.883	0.881	0.508	0.667	0.629	0.949	0.952	0.949
3	**0.972**	0.974	0.974	0.845	0.838	0.886	0.530	0.921	0.442
4	0.866	0.877	0.867	0.447	0.696	0.560	0.656	0.890	0.601

6. Conclusion

This research aimed to deliver appropriate medical advice for patients through a dialogue system. Despite the unclear utterance of the patient when describing symptoms, the Hopfield neural network exhibited high performance in clustering similar cases based on energy values, where the results demonstrated cohesion and great convergence in the elements of the same cluster, despite the lack of clarity of utterance, even for humans. This issue leads to the difficulty of choosing how many clusters should determine the number of pieces of advice in the database. Therefore, various numbers of clusters were experimented with, finding that three clusters provide the best evaluation value compared to the other numbers of clusters. Labels of the dataset were provided based on semantic similarity because most of the doctor's utterances were implicit. Therefore, the feedforward neural network outperformed the other base models in the external evaluation and provided a high percentage of correct decisions according to the experimental results.

7. Reference

[1] K. H. Alhussayni, A. Zamyatin, and S. E. Alshamery, "Bidirectional RNN-based attention model for jointly intent detection and slot filling," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 10, no. 11, pp. 404–410, 2018.

[2] C.-W. Goo *et al.*, "Slot-Gated Modeling for Joint Slot Filling and Intent Prediction," *Naacl2018*, vol. 2, pp. 753–757, 2018, [Online]. Available: http://www.aclweb.org/anthology/N18-2118.

[3] K. H. Alhussayni, A. Zamyatin, and S. E. Alshamery, "Dialogue state tracking accuracy improvement by distinguishing slot-value pairs and dialogue behaviour," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 2, pp. 1057–1064, Feb. 2020, doi: 10.11591/ijeecs.v21.i2.pp1057-1064.

[4] V. Zhong, C. Xiong, and R. Socher, "Global-Locally Self-Attentive Dialogue State Tracker," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 1458–1467, 2018, [Online]. Available: http://arxiv.org/abs/1805.09655.

[5] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare Journal*, vol. 6, no. 2, pp. 94–98, Jun. 2019, doi: 10.7861/futurehosp.6-2-94.

[6] J. Pereira and Ó. Díaz, "Using Health Chatbots for Behavior Change: A Mapping Study," *Journal of Medical Systems*, vol. 43, no. 5, pp. 1–13, 2019, doi: 10.1007/s10916-019-1237-1.

[7] J. Martínez-Miranda *et al.*, "Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour," *Journal of Medical Systems*, vol. 43, no. 8, pp. 1–18, Aug. 2019, doi: 10.1007/s10916-019-1387-1.

[8] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 8, pp. 2554–2558, 1982, doi: 10.1073/pnas.79.8.2554.

[9] J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, “‘Unlearning’ has a stabilizing effect in collective memories,” *Nature*, vol. 304, no. 5922, pp. 14–20, 1983, doi: 10.1038/304158a0.

[10] A. Conti, M. Guerra, D. Dardari, N. Decarli, and M. Z. Win, “Network experimentation for cooperative localization,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 2, pp. 467–475, Feb. 2012, doi: 10.1109/JSAC.2012.120227.

[11] R. Nandakumar, K. K. Chintalapudi, and V. N. Padmanabhan, “Centaur: Locating devices in an office environment,” *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, pp. 281–292, 2012, doi: 10.1145/2348543.2348579.

[12] S. Tata and J. M. Patel, “Estimating the selectivity of tf-idf based cosine similarity predicates,” *SIGMOD Record*, vol. 36, no. 2, pp. 7–12, Jun. 2007, doi: 10.1145/1328854.1328855.

[13] H. Kim and S. Lee, “An Intelligent Information System for Organizing Online Text Documents,” *Knowledge and Information Systems*, vol. 6, no. 2, pp. 125–149, Mar. 2004, doi: 10.1007/s10115-003-0103-z.

[14] A. Hinneburg and D. A. Keim, “A General Approach to Clustering in Large Databases with Noise,” *Knowledge and Information Systems*, vol. 5, no. 4, pp. 387–415, Nov. 2003, doi: 10.1007/s10115-003-0086-9.

[15] S. Zhong and J. Ghosh, “Generative model-based document clustering: A comparative study,” *Knowledge and Information Systems*, vol. 8, no. 3, pp. 374–384, Sep. 2005, doi: 10.1007/s10115-004-0194-1.

[16] R. Cordeiro De Amorim and B. Mirkin, “Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering,” *Pattern Recognition*, vol. 45, no. 3, pp. 1061–1075, Mar. 2012, doi: 10.1016/j.patcog.2011.08.012.

[17] D. Li and L. Yang, *Deep learning in natural language processing*. 2018.

[18] D. O. Hebb, *Organization of behavior. New York: Wiley*. 1950.

[19] M. G. Anderson D, *Artificial Neural Networks Technology*. 1992.

[20] B. Krose and P. van der Smagt, *introduction to Neural*. 1996.

[21] C. Leave, J. A. Freeman, and D. M. Skapura, *Neural Networks: Algorithms, Applications and Programming Techniques*. 1992.

[22] X. B. Xue and Z. H. Zhou, “Distributional features for text categorization,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 3, pp. 428–442, Mar. 2009, doi: 10.1109/TKDE.2008.166.

[23] D. Sharma, “Stemming Algorithms: A Comparative Study and their Analysis,” *International Journal of Applied Information Systems*, vol. 4, no. 3, pp. 7–12, Sep. 2012, doi: 10.5120/ijais12-450655.

[24] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.

[25] W. Yang *et al.*, "On the generation of medical dialogues for COVID-19," *medRxiv*. medRxiv, pp. 1–6, May 15, 2020, doi: 10.1101/2020.05.08.20095810.

[26] T. Mairittha, N. Mairittha, and S. Inoue, "Automatic labeled dialogue generation for nursing record systems," *Journal of Personalized Medicine*, vol. 10, no. 3, pp. 1–24, 2020, doi: 10.3390/jpm10030062.