Local Triangular Kernel-Based Clustering (LTKC) For Case Indexing on Case-Based Reasoning

Damar Riyadi*[1], Aina Musdholifah[2]
[1]Master of Computer Science, FMIPA UGM, Yogyakarta, Indonesia
[2]Department of Computer Science and Electronic, FMIPA UGM, Yogyakarta, Indonesia
e-mail: ***[1]damar.riyadi@mail.ugm.ac.id**, [2]aina_m@ugm.ac.id

Penelitian ini bertujuan untuk meningkatkan kinerja Case-Based Reasoning dengan memanfaatkan analisis cluster yang digunakan sebagai metode pengindeksan untuk mempercepat pengambilan kasus di CBR. Metode clustering menggunakan Local Triangular Kernel-based Clustering (LTKC). Metode koefisien kosinus digunakan untuk menemukan cluster yang sesuai sedangkan nilai kesamaan dihitung menggunakan jarak Manhattan, jarak Euclidean, dan jarak Minkowski. Hasil dari metode tersebut akan dibandingkan untuk menemukan metode mana yang memberikan hasil terbaik. Penelitian ini menggunakan tiga data uji: penyakit malnutrisi, penyakit jantung, dan penyakit tiroid. Hasil pengujian menunjukkan bahwa CBR dengan indeks LTKC memiliki akurasi dan waktu pemrosesan yang lebih baik daripada CBR tanpa pengindeksan. Keakuratan terbaik pada ambang 0.9 penyakit gizi buruk, diperoleh dengan menggunakan jarak Euclidean yang menghasilkan akurasi 100% dan waktu pengambilan rata-rata 0,0722 detik. Akurasi terbaik pada ambang 0,9 penyakit jantung, diperoleh dengan menggunakan jarak Minkowski yang menghasilkan akurasi 95% dan 0,1785 detik waktu pengambilan rata-rata. Akurasi terbaik pada ambang 0,9 penyakit tiroid, diperoleh dengan menggunakan jarak Minkowski yang menghasilkan akurasi 92,52% dan 0,3045 waktu pengambilan rata-rata. Perbandingan akurasi CBR dengan indeks SOM, pengindeksan DBSCAN, dan pengindeksan LTKC untuk penyakit gizi buruk dan penyakit jantung menghasilkan bahwa mereka memiliki akurasi yang hampir sama.

Keywords— *case-based reasoning, indeks, clustering, LTKC, nearest neighbor retrieval*

Abstract
This study aims to improve the performance of Case-Based Reasoning by utilizing cluster analysis which is used as an indexing method to speed up case retrieval in CBR. The clustering method uses Local Triangular Kernel-based Clustering (LTKC). The cosine coefficient method is used for finding the appropriate cluster while similarity value is calculated using Manhattan distance, Euclidean distance, and Minkowski distance. Results of those methods will be compared to find which method gives the best result. This study uses three test data: malnutrition disease, heart disease, and thyroid disease. Test results showed that CBR with LTKC-indexing has better accuracy and processing time than CBR without indexing. The best accuracy on threshold 0.9 of malnutrition disease, obtained using the Euclidean distance which produces 100% accuracy and 0.0722 seconds average retrieval time. The best accuracy on threshold 0.9 of heart disease, obtained using the Minkowski distance which produces 95% accuracy and 0.1785 seconds average retrieval time. The best accuracy on threshold 0.9 of thyroid disease, obtained using the Minkowski distance which produces 92.52% accuracy and 0.3045 average retrieval time. The accuracy comparison of CBR with SOM-indexing, DBSCAN-indexing, and LTKC-indexing for malnutrition diseases and heart disease resulted that they have almost equal accuracy.

Keywords— *case-based reasoning, indexing, clustering, LTKC, nearest neighbor retrieval*

1. INTRODUCTION

Case-Based Reasoning (CBR) is an approach for problem-solving by utilizing solutions from similar problems that have been experienced before. The method for problem-solving in CBR is based on memory, where the past cases saved in case base are the starting point for solving new problems. In many CBR applications, cases are usually represented as two unstructured sets of attribute-value pairs that represents the problem and solution features [1]. A good CBR system depends on the mechanism for finding the similar case (retrieval process) for new problems. The more cases stored in the case base, the time needed to find similar case will be longer because it should calculate similarity value of the new case with all old cases saved in the case base. Therefore, indexing of cases is needed to speed up the retrieval process of finding similar case. Previous studies focusing on indexing on CBR have been conducted before. Ant Colony Optimization (ACO) approach is used by [2], and [3] used a fuzzy approach for indexing on CBR. Based on the previous studies that have been conducted until 2015, shows that the indexing process on CBR is still a relevant research topic.

Clustering is an exploration technique which capable of extracting knowledge from a set of data by grouping unlabeled data based on similarity and dissimilarity into clusters so that each cluster contains as similar as possible data [4]. The clustering algorithm used in this study is Local Triangular Kernel-based Clustering (LTKC). LTKC is density-based clustering that determines the density of data points using combination of two nonparametric density estimation procedures. LTKC combines k-nearest neighbour (KNN) and kernel density estimation (KDE). KNN density estimation is extended and combined with triangular kernel function. LTKC uses Bayesian decision rule in order to assign objects to respective clusters. LTKC only requires one input parameter, which is the number of nearest neighbours (k) without having to define the number of clusters because the algorithm finds it automatically [5]. The LTKC algorithm is chosen in this study because based on previous research showed that LTKC produced more accurate clustering results (based on some clustering validation methods) and less processing times when compared to DBSCAN and DENCLUE [5].

The method used for finding similarity value between old cases and new cases in this study is nearest neighbour retrieval. This technique compares each feature of target case (new case) with features of source case (old cases) stored in a case base, then the comparison of each feature is calculated by a similarity function. Previous research conducted by [6], [7], [8], and [9] showed that the method was good enough to be used in the CBR system for diagnosis.

2. METHODS

2.1 Knowledge Acquisition

This study uses three datasets. Malnutrition disease in infants dataset is acquired from previous research conducted by [8] which contains 90 data. From the dataset, 70 data are used as train data and 20 data are used as test data. The malnutrition dataset is originally obtained from RSUP Dr. Sardjito Yogyakarta. Heart disease dataset is acquired from previous research conducted by [7] which contains 135 data. From the dataset, 115 data are used as train data and 20 data are used as test data. The heart disease dataset is originally obtained from RSUP Dr. Sardjito Yogyakarta. Thyroid disease dataset is provided by Garvan Institute and J. Ross Quinlan, New South Wales Institute, Sydney, Australia. Thyroid disease dataset used in this study contains 1428 data. From the dataset, 1000 data are used as train data and 428 data are used as test data. The thyroid disease dataset can be obtained from UCI *Machine Learning Repository*.

2.2 Case Representation

In this study, cases are represented as a frame model. In a frame model, cases are represented in the form of a collection of features that uniquely identify the case and solution of the case. In heart disease and malnutrition disease in infants cases, each feature of the case - age, gender, symptoms or risk factors - has a weight that indicates the level of influence on the diagnosis of disease suffered by the patient. The weight of each feature is given by an expert. The greater weight indicates the greater influence of the feature in determining or diagnosing the patient's disease. The thyroid disease case has its own data characteristics because there is no weight given by an expert in its features. After clustering process has been performed on all old cases saved in case base, the cases are represented again with new additional information, which is the cluster where the case is located and its relation with center of cluster (centroid). Table 1 shows an example of case representation of malnutrition in infants with cluster information.

Table 1 Case representation of malnutrition in infants

No	Case Data	Explanation
A	Symptoms	
1	G011	Wrinkled and baggy skin
2	G016	Whiny
3	G019	Edema
4	G021	Weight loss (thin)
B	Patient Data	
1	Age	23 months
C	Diagnosis	
1	P003	Marasmus-Kwashiorkor
D	*Indexing*	
1	*Cluster*	1

2.3 Indexing

Clustering method used in this study is Local Triangular Kernel-Based Clustering (LTKC). LTKC is used for grouping old cases into clusters by determining similarity and dissimilarity of cases so that each cluster contains as similar as case data. The clustering process groups all cases into their respective clusters and produces the center of cluster (centroid), which is the average value of features for each case in the same cluster. The centroid value will be used as an index for diagnosing new case or problem.

2.3.1 Data Normalization

Data normalization aims to obtain features with smaller values which represent the original data without losing its characteristics. Also, normalization is necessary to avoid an unbalance range on specific features in the case base. Several features will be normalized in this study, especially the feature which has a numeric value. In malnutrition in infants dataset and heart disease dataset, age feature will be normalized. In thyroid disease dataset, there are five features will be normalized, which are age, TSH, T3, TT4, and T4U. The Min-Max Normalization method is used for normalization. The method requires the minimum and maximum value of features. For example, age feature on malnutrition in infants has a minimum value of 0 and a maximum value of 60 months. Equation (1) is the formula of Min-Max Normalization.

$$v' = \frac{v - minA}{maxA - minA} \qquad (1)$$

2.3.2 Cluster-indexing using LTKC

The first step of case base indexing process using LTKC method is initialization by defining the value of k nearest-neighbor and maximum iteration. Then, for each data point (case) in the case base, Euclidean distance is calculated using formula (2). After the distance of each data point has been calculated, create a distance table by sorting the distance value from the smallest to the highest. Initialize the number of clusters which is equal with the number of data point.

In the iteration step, for each data point in a cluster, find k nearest-neighbor based on distance table created before and put them in as members of the cluster. For each data point, find clusters which contains the data point, then calculate triangular kernel value for each cluster by using formula (3), and finally put the data point into the cluster which has the maximal value of triangular kernel. The maximal value of triangular kernel is obtained using formula (4). Then, re-index the cluster label by deleting clusters with has the same members as another cluster.

The iteration step is performed until the cluster structure is unable to change or the iteration has reached the maximum iteration which is defined in the first step. The cluster-indexing process results that each data point (case) with its features grouped into respective clusters, the features of a data point is used to calculate the center of cluster (centroid). The center of cluster (centroid) is obtained with an average value of features of each data point within the same cluster.

$$dist(x_i, x_j) = \sqrt[2]{\sum_{k=1}^{d} (x_{ik} - x_{jk})^2} \tag{2}$$

$$Kt_\omega = \frac{1}{n_\omega} \sum_{j=1}^{n_\omega} \left(1 - dist(x, x_j)\right) \tag{3}$$

$$Clust_\omega = max_{\omega \in \omega_{nci}}\left(Kt_\omega(x_i)\right) \tag{4}$$

2.3.3 Cluster Validation

The cluster validation method used in this study is silhouette coefficient. The method is used to validate the quality of a cluster, how appropriate data are grouped into clusters. The silhouette coefficient method is a combination of cohesion and separation method. The silhouette coefficient value of a data point in clusters is calculated using formula (5) [10].

$$s(i) = \frac{b(i) - a(i)}{max\left(a(i), b(i)\right)} \tag{5}$$

Where $a(i)$ is the average distance value of data point i to another data point within the same cluster, while $b(i)$ is the average distance value of data point i to another data point which is located in different cluster. The average value of all $s(i)$ is the value of *silhouette coefficient* global which is used to validate the clustering result.

2.4 Retrieve and Reuse Process

CBR system with clustering-indexing does not necessarily calculate the similarity value of a new case with all cases saved in the case base, but it only calculates the similarity value with cases located in same cluster or group. The retrieval process is divided into two steps. The first step is finding the appropriate cluster by utilizing the center of clusters (centroids) saved in the database. The second step is calculating similarity value of new case with cases within the same cluster. Figure 1 shows a flowchart of the retrieval process.
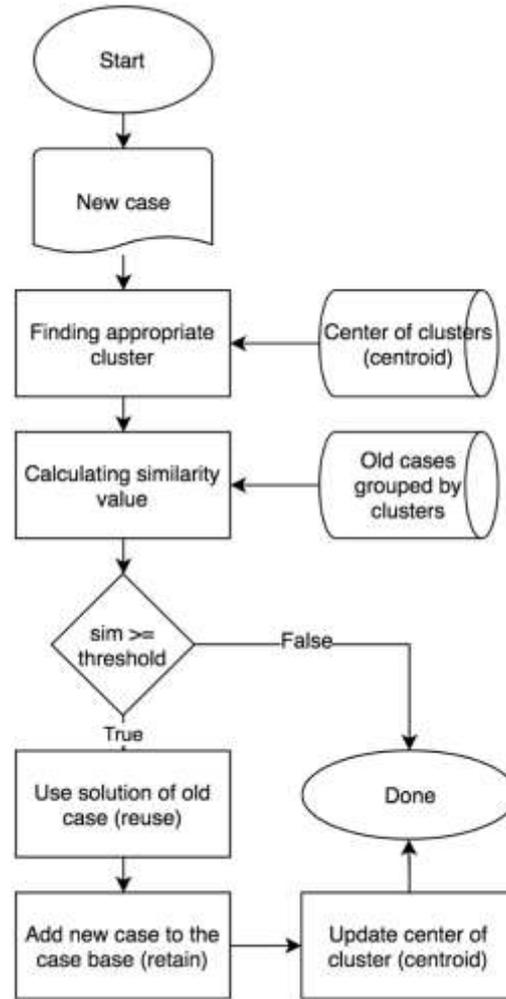
Figure 1 Flowchart of retrieval process

2.4.1 Finding Appropriate Cluster

The diagnosis process of a new case in this CBR system begins with finding the appropriate cluster. The appropriate cluster is obtained by calculating similarity value between new case's features or symptoms value with the value of respective features of centroids (center of clusters) save in the database. Cosine coefficient method is used to perform such task as described in formula (6). The method is used because it is suitable for data with small values, also based on previous research conducted by [11] it produced a good accuracy of diagnosis.

$$Cos(X,Y) = \frac{\langle X,Y \rangle}{\|X\|\|Y\|} \qquad (6)$$

2.4.2 Calculation of Similarity Value

The calculation of similarity value is performed by comparing each feature of a new case with each feature of all old cases saved in the case base; then the comparison result is calculated using a similarity function. There are two types of similarity: local similarity and global similarity. The local similarity is a feature-level similarity, while global similarity is a case-level similarity. The calculation process of local similarity is divided into two types based on the kind of data, formula (7) [12] is used for numeric data, while formula (8) [12] is used for symbolic data. The methods used for calculation global similarity in this study are Manhattan distance similarity as described on formula (9) [13], Euclidean distance similarity [13], and Minkowski distance similarity [13] as described on formula (10) with the value of $r=2$ for Euclidean distance and $r=3$ for Minkowski distance similarity.

$$f(S_i, T_i) = 1 - \frac{|S_i - T_i|}{|f_{max} - f_{min}|} \qquad (7)$$

$$f(S_i, T_i) = \begin{cases} 0, jika\ S_i \neq T_i \\ 1, jika\ S_i = T_i \end{cases} \qquad (8)$$

$$Sim(S, T) = \frac{\sum_{i=1}^{n} f(S_i, T_i) * w_i}{\sum_{i=1}^{n} w_i} \qquad (9)$$

$$Sim(S, T) = \left(\frac{\sum_{i=1}^{n} w_i{}^r * |f(S_i, T_i)|^r}{\sum_{i=1}^{n} w_i{}^r} \right)^{1/r} \qquad (10)$$

2.4.3 Reuse

The reuse process used in this study is performed by obtaining the highest similarity value, and the value is equal or greater than a defined threshold. The result of retrieval process with the highest similarity value is used as a solution of the new case.

2.4.4 Retain

The new case which has been revised by an expert saved in the case base with consideration of the value of centroid (center of cluster) as a new knowledge (retain).

2.5 System Testing

The system testing is performed by applying some new cases; 20 new cases as test data for malnutrition in infants, 20 new cases for heart disease, and also 428 new cases as test data for thyroid disease. The result of diagnosis of CBR system is compared with the result of diagnosis of medical record. This study performs two test scenarios: the first scenario, CBR system compares the accuracy of diagnosis and retrieval time between CBR system without indexing and CBR system with indexing using LTKC. Test data used in this scenario is malnutrition in infants data, heart disease data, and thyroid disease data. The second scenario compares the accuracy of diagnosis between CBR system with LTKC-indexing and CBR system with SOM-indexing and DBSCAN-indexing that has been performed in the previous research [11]. Test data used in this scenario is malnutrition in infants data and heart disease data.

3. RESULTS AND DISCUSSION

3.1 Clustering Process of The Case Base

The LTKC clustering algorithm only requires one parameter which is the value of k nearest-neighbor. But, in order to prevent the clustering process takes too long for specific k value, a maximum value of iteration is applied as an additional parameter. In this study, all possible k values are tested to obtain the optimal k. The optimal k value candidate is obtained by examining silhouette coefficient value of the clustering result. From all tested k value, only the top five will be chosen for further testing by examining its silhouette coefficient value. Then, each candidate will be tested in CBR system to obtain the accuracy of diagnosis on retrieval process of finding similar case. Similarity measure used in the test is Manhattan distance similarity and the lowest threshold value of 0.7 is applied to check whether CBR system can produce the right diagnosis or not. The accuracy of CBR system for each k candidate is compared to find the best k value for CBR system with the highest accuracy of diagnosis. The k value with the highest accuracy and silhouette coefficient value is used as the optimal parameter for LTKC. Table 2 shows the clustering result with its optimal k value.

Table 2 Clustering result with optimal k value

Attributes	Malnutrition data	Heart disease data	Thyroid disease data

Value of k	27	22	558
Number of train data	70	115	1000
Number of cluster	3	6	6
Silhouette coefficient	0.3293	0.2261	0.4040
Accuracy	100%	90%	92,52%

3.2 System Capability Analysis

The capability analysis of diagnostic system aims to determine the ability of the system on producing an accurate diagnosis. There are two scenarios performed in the analysis. The first scenario analyzes the accuracy of diagnosis of CBR system without indexing and the second scenario analyzes CBR system with LTKC-indexing. On both scenarios, the process of finding the appropriate cluster utilizes cosine coefficient method, and the similarity value is obtained by using Manhattan distance similarity, Euclidean distance similarity, and Minkowski distance similarity. The test is performed by applying 20 new cases as test data for malnutrition in infants, 20 new cases for heart disease, and 428 new cases for thyroid disease. Table 3, 4, and 5 shows the comparison of system capability for each scenario.

Table 3 Comparison of CBR system capability for malnutrition in infants data

Similarity Method	Threshold	Accuracy		Average retrieval time	
		Non-indexing	LTKC	Non-indexing	LTKC
Manhattan	0.7	90% (18)	100% (20)	0.1710	0.0703
Euclidean	0.7	90% (18)	100% (20)	0.1735	0.0722
Minkowski	0.7	85% (17)	100% (20)	0.1934	0.0792
Manhattan	0.8	85% (17)	95% (19)	0.1710	0.0703
Euclidean	0.8	90% (17)	100% (20)	0.1735	0.0722
Minkowski	0.8	85% (17)	100% (20)	0.1934	0.0792
Manhattan	0.9	45% (9)	55% (11)	0.1710	0.0703
Euclidean	0.9	90% (18)	100% (20)	0.1735	0.0722
Minkowski	0.9	85% (17)	100% (20)	0.1934	0.0792

Table 4 Comparison of CBR system capability for heart disease data

Similarity Method	Threshold	Accuracy		Average retrieval time	
		Non-indexing	LTKC	Non-indexing	LTKC
Manhattan	0.7	80% (16)	90% (18)	0.8649	0.1661
Euclidean	0.7	95% (19)	95% (19)	0.8682	0.1780
Minkowski	0.7	95% (19)	95% (19)	0.8687	0.1785
Manhattan	0.8	65% (13)	65% (13)	0.8649	0.1661
Euclidean	0.8	95% (19)	95% (19)	0.8682	0.1780
Minkowski	0.8	95% (19)	95% (19)	0.8687	0.1785
Manhattan	0.9	30% (6)	35% (7)	0.8649	0.1661
Euclidean	0.9	65% (13)	65% (13)	0.8682	0.1780
Minkowski	0.9	95% (19)	95% (19)	0.8687	0.1785

Table 5 Comparison of CBR system capability for thyroid disease data

Similarity Method	Threshold	Accuracy		Average retrieval time	
		Non-indexing	LTKC	Non-indexing	LTKC
Manhattan	0.7	91,59% (392)	92,52% (396)	0.8748	0.2933
Euclidean	0.7	91,59% (392)	92,52% (396)	0.9197	0.3027
Minkowski	0.7	91,59% (392)	92,52% (396)	0.9272	0.3045
Manhattan	0.8	91,59% (392)	394 (92,05%)	0.8748	0.2933
Euclidean	0.8	91,59% (392)	396 (92,52%)	0.9197	0.3027
Minkowski	0.8	91,59% (392)	396 (92,52%)	0.9272	0.3045
Manhattan	0.9	89.72% (384)	86,21% (369)	0.8748	0.2933
Euclidean	0.9	91,59% (392)	92,05% (394)	0.9197	0.3027
Minkowski	0.9	91,59% (392)	92,52% (396)	0.9272	0.3045

Based on the test results show that CBR system with cluster- indexing LTKC has better accuracy and faster processing time than CBR without indexing. The best accuracy and processing time on malnutrition cases with a threshold of 0.9, obtained by using Euclidean distance similarity method that produces 100% accuracy with 0.0722 seconds average processing time. The average value has a little difference with the Minkowski distance similarity method. The best accuracy and processing time on heart diseases cases with a threshold of 0.9, obtained by using Minkowski distance similarity method that produces 95% accuracy with 0.1785 seconds average processing time. The best accuracy and processing time on thyroid diseases cases with a threshold of 0.9, obtained by using Minkowski distance similarity method that produces 92,52% accuracy with 0.3045 seconds average processing time.

3.3 Comparison of CBR system accuracy with indexing SOM, DBSCAN, and LTKC

CBR system with SOM and DBSCAN indexing method for malnutrition in infants and heart disease has been developed in previous research [11]. Table 6 sows a comparison of system capabilities in terms of accuracy on CBR systems with SOM, DBSCAN, and LTKC indexing for malnutrition cases in infants. While the comparison of CBR system accuracy for heart disease cases is presented in Table 7.

Table 6 Comparison of CBR system accuracy with cluster-indexing
for malnutrition in infants

Similarity Method	Threshold	LTKC	SOM	DBSCAN
Manhattan	0.7	100% (20)	100% (20)	100% (20)
Euclidean	0.7	100% (20)	100% (20)	100% (20)
Minkowski	0.7	100% (20)	100% (20)	100% (20)
Manhattan	0.8	95% (19)	90% (18)	90% (18)
Euclidean	0.8	100% (20)	100% (20)	100% (20)
Minkowski	0.8	100% (20)	100% (20)	100% (20)
Manhattan	0.9	55% (11)	55% (11)	55% (11)
Euclidean	0.9	100% (20)	100% (20)	100% (20)
Minkowski	0.9	100% (20)	100% (20)	100% (20)

Tabel 7 Comparison of CBR system accuracy with cluster-indexing for heart disease

Similarity Method	*Threshold*	LTKC	SOM	DBSCAN
Manhattan	0.7	90% (18)	85% (17)	85% (17)
Euclidean	0.7	95% (19)	100% (20)	100% (20)
Minkowski	0.7	95% (19)	85% (19)	85% (19)
Manhattan	0.8	65% (13)	65% (13)	65% (13)
Euclidean	0.8	95% (19)	100% (20)	100% (20)
Minkowski	0.8	95% (19)	95% (19)	95% (19)
Manhattan	0.9	35% (7)	30% (6)	30% (6)
Euclidean	0.9	65% (13)	60% (12)	60% (12)
Minkowski	0.9	95% (19)	95% (19)	95% (19)

CBR system with LTKC indexing on some threshold values and similarity methods is better than CBR system with SOM and DBSCAN indexing, but on other threshold values and similarity methods, SOM and DBSCAN is better than LTKC with only 1 case difference. It shows that CBR system with SOM, DBSCAN, and LTKC indexing for malnutrition in infants and heart disease cases resulted that they have almost equal accuracy of diagnosis.

4. CONCLUSION

Based on the datasets used in this study, it can be concluded that CBR system with indexing LTKC for malnutrition in infants, heart disease, and thyroid disease has better accuracy of diagnosis and retrieval time than CBR system without indexing. The CBR system for malnutrition in infants with LTKC-indexing is capable of producing 100% accuracy of diagnosis and 0.0722 seconds average retrieval time on a threshold value of 0.9 using Euclidean distance similarity method. The CBR system for heart disease with LTKC-indexing is capable of producing 95% accuracy of diagnosis and 0.1785 seconds average retrieval time on a threshold value of 0.9 using Minkowski distance similarity method. The CBR system for thyroid disease with LTKC-indexing is capable of producing 92.52% accuracy of diagnosis and 0.3045 seconds average retrieval time on a threshold value of 0.9 using Minkowski distance similarity method. The test of CBR system accuracy with indexing LTKC, SOM, and DBSCAN for the diagnosis of malnutrition in infants and heart disease indicates that all three indexing methods produce almost equal accuracy of diagnosis.

5. SUGGESTION

Further research needs to use a particular method to determine the optimal k value as the LTKC parameter without having to try all possible k values available to speed up the process of finding the best clustering result of the base case. It is also necessary to use additional validation methods of clustering results so that the value of the cluster validation more represent the diagnostic accuracy of the CBR system. Related to the dataset used, it is also necessary to use some kind of certainty factor of diagnosis that considers the number of features exists in a particular case so that the similarity value of diagnosis produced by CBR system also considers the level of confidence.

REFERENCES

[1]   R. Weber, *Case-Based Reasoning: A Text Book*. Berlin: Springer-Verlag, 2013.

[2]   J. Shu, "The Application of Ant Colony Optimization in CBR," *Adv. Intell. Syst. Comput.*, vol. 212, 2013 [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-37502-6_143. [Accessed: 10-Nov-2017].

[3]   A. Sarkheyli and D. Söffker, "Case Indexing in Case-Based Reasoning by Applying Situation Operator Model as Knowledge Representation Model," *IFAC-PapersOnLine*, vol. 28, no. 1, pp. 81–86, 2015 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S240589631500049X. [Accessed: 10-Nov-2017].

[4]   I. H. Witten, *Data Mining Practical Machine Learning Tools and Techniques*, 4th ed., vol. 18 Suppl, no. 1. Cambridge, Massachusetts: Morgan Kaufmann, 2016.

[5]   A. Musdholifah, S. Zaiton, and M. Hashim, "Cluster Analysis on High-Dimensional Data : A Comparison of Density-based Clustering Algorithms," vol. 7, no. 2, pp. 380–389, 2013 [Online]. Available: http://ajbasweb.com/old/ajbas/2013/February/380-389.pdf. [Accessed: 10-Nov-2017].

[6]   E. Faizal and S. Hartati, "Case-Based Reasoning untuk Mendiagnosa Penyakit Cardiovascular dengan Metode Weighted Minkowski," S2 Ilmu Komputer, Universitas Gadjah Mada, 2013.

[7]   E. Wahyudi and S. Hartati, "Case-Based Reasoning untuk Diagnosis Penyakit Jantung," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 1, p. 1, 2017 [Online]. Available: https://journal.ugm.ac.id/ijccs/article/view/15523. [Accessed: 10-Nov-2017].

[8]   Nurfalinda, S. Hartati, and A. Musdholifah, "Case Based Reasoning dan Rule Based Reasoning untuk Diagnosis Penyakit Gizi Buruk pada Balita," S2 Ilmu Komputer, Universitas Gadjah Mada, 2016.

[9]   S. H. Tedy Rismawan, "Case-Based Reasoning Untuk Diagnosa Penyakit THT (Telinga Hidung dan Tenggorokan)," *IJCCS-Indonesian J. Comput. ...*, vol. 6, no. Sistem Pakar, pp. 67–78, 2013 [Online]. Available: http://journal.ugm.ac.id/index.php/ijccs/article/view/2154. [Accessed: 10-Nov-2017].

[10] B. Desgraupes, "Clustering Indices," *CRAN Packag.*, no. April, pp. 1–10, 2013 [Online]. Available: cran.r-project.org/web/packages/clusterCrit. [Accessed: 10-Nov-2017].

[11] H. Santoso, "Metode Indexing pada Case Based Reasoning (CBR) menggunakan Density Based Spatial Clustering Application with Noise (DBSCAN)," S2 Ilmu Komputer, Universitas Gadjah Mada, 2017.

[12] A. M. Salem, M. Roushdy, and R. A. Hodhod, "A Case Based Expert System for Supporting Diagnosis of Heart Diseases," *Heart Fail.*, no. 5, pp. 33–39, 2005 [Online]. Available: http://www.cs.ru.ac.za/courses/honours/ai/casebaseddiagnosis.pdf. [Accessed: 10-Nov-2017].

[13] H. Núñez *et al.*, "A Comparative Study on The Use of Similarity Measures in Case-Based Reasoning to Improve The Classification of Environmental System Situations," *Environ. Model. Softw.*, vol. 19, no. 9, pp. 809–819, 2004 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364815203002020. [Accessed: 10-Nov-2017].