Computer Engineering and Applications Vol. 11, No. 2, June 2022

Point of Interest (POI) Recommendation System using Implicit Feedback Based on K-Means+ Clustering and User-Based Collaborative Filtering

Sulis Setiowati[1], Teguh Bharata Adji[2], Igi Ardiyanto[3]

[1]*Department of Electrical Engineering, Jakarta State of Polytechnic*
[2,3]*Department of Electrical Engineering, Faculty of Technique Universitas Gadjah Mada (UGM)*
**sulis.setiowati@elektro.pnj.ac.id*

ABSTRACT

Recommendation system always involves huge volumes of data, therefore it causes the scalability issues that do not only increase the processing time but also reduce the accuracy. In addition, the type of data used also greatly affects the result of the recommendations. In the recommendation system, there are two common types of data namely implicit (binary) rating and explicit (scalar) rating. Binary rating produces lower accuracy when it is not handled with the properly. Thus, optimized K-Means+ clustering and user-based collaborative filtering are proposed in this research. The K-Means clustering is optimized by selecting the K value using the Davies-Bouldin Index (DBI) method. The experimental result shows that the optimization of the K values produces better clustering than Elbow Method. The K-Means+ and User-Based Collaborative Filtering (UBCF) produce precision of 8.6% and f-measure of 7.2%, respectively. The proposed method was compared to DBSCAN algorithm with UBCF, and had better accuracy of 1% increase in precision value. This result proves that K-Means+ with UBCF can handle implicit feedback datasets and improve precision.

Keywords: Point of Interest, Recommender Systems, Collaborative Filtering, Implicit Feedback, K-Means+.

1. INTRODUCTION

The tendency of users in utilizing technology is to maximum impact of increasing information every time. Many people are confused in determining the right choice of overloaded information. A recommendation system (RS) can be developed to overcome the problem. The recommendation system is a method for recommending items to users from a pile of relevant information [1, 2]. At present, it has been widely applied in various fields to provide services for its users such as www.amazon.com, www.google.com, www.netflix.com. Further, recommendation system uses the information to create a recommendation for user over the increasing number of choices. Recommendation system will recommend items according to different user preferences. The focus of this research is the Point of Interest (POI) recommendation system that is used to facilitate users finding the location in accordance with their preference. Before the recommendation system evolves, people use conventional methods for selecting items or locations. They search information one by one of each item, then decide the best option for them. This is certainly inefficient and ineffective if the information grows over time. That is the reason why POI recommendation system is one of solution that can provide

recommendations precisely and suit to user preferences with fast processing time [3]. In this study, POI recommendation system uses implicit feedback or binary rating data.

The data is categorized into two, namely implicit feedback and explicit feedback. Over the years, researches related explicit feedback (scale of 1-5) have been widely developed in the movie, e-commerce dataset and others. The recommendation system using this dataset produces an accuracy up to 80% [4, 5]. However, research on the POI recommendation system using implicit feedback still needs to be enhanced because it needs the appropriate method in processing data to produce recommendations. To date, the precision that can be achieved from previous research related to implicit feedback is an average of 4-7% [6]-[8].

Accordingly, developing an appropriate method to overcome the problems and improve a recommendation quality is urgently needed. The collaborative filtering becomes the most popular technique due to high accuracy compared to other techniques [3, 5, 9-11]. In addition, collaborative filtering system usually is used for large database [12]. The current research uses User-Based Collaborative Filtering (UBCF) whereas employing user similarities for recommending POI. In the previous research, this method has never been used for POI recommendation system. The UBCF is a method that can improve recommendation performance because of its easily implemented, have sufficient accuracy, and fit in offline evaluation [13].

Referring to the problem in the handling of binary rating data and low precision value compared with other recommendation systems, it needs an approach to carry out the problem. In some research, clustering process on dataset proved to enhance the dataset quality. K-means is a clustering algorithm used in this research. Up until now, there was many developments about K-means [14]. This algorithm is compatible to be implemented in the clustering of the POI recommendation system that involves a geospatial dataset so that the clustering result will be maximum. To generate optimal clustering, the K parameter in the algorithm is automatically defined using Davies-Bouldin Index (DBI) method based on the approximately estimation of the distances between clusters and their dispersion to obtain a final value that represents the quality of the partition [15]. The DBI measurement approach is conducted to maximize the inter-cluster and minimize the intra-cluster distance. By using this method, the best value of clustering number is obtained to make optimal cluster area. In this research, researcher attempt to implement K-means algorithm with K value optimization using DBI method (K-Means+) and User-Based Collaborative Filtering (UBCF) approach to deal with implicit feedback data type in order to improve precision.

2. RELATED WORKS

To date, research on the POI recommendation system has evolved to improve accuracy and overcome previous problems. Collaborative filtering has become one of the most successfully and widely used recommendation technique, aiming at helping people reduce the amount of time they spend to find out the items they are interested [16]. Recommendation systems have been developed with different methods and techniques, and future research trends and challenges, such as sparse data, scalability, synonymous, shilling attacks, and privacy protection. Since collaborative filtering's shortcomings are obviously, many works on collaborative

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

filtering have been carried out. In his research, Mao Qinjiao [17] proposed traditional collaborative filtering algorithms with binary similarity to develop recommendation systems. They focus on the implicit feedback on which filtering approach is constructed to provide users with Top-N recommendation. The experiments showed the traditional recommendation performance can be effectively improved by proposed methods. Researches on implicit feedback are more common in information retrieval. The literature [18] studied the implicit feedback and a Bayesian feedback approach was proposed to build user profiles. Implicit feedback was used to build a special aspect model for recommendations and optimize the algorithm on this model [19]. George and Thorsten [20, 21] also analyzed various feedback of users and provide a practicing method. However, these methods are not generic, and usually need to design the systems corresponding to catch the specific implicit feedback. Research proposed by [16] summarize the traditional CF-based approaches and techniques used in RS and study some recent hybrid CF-based recommendation approaches and techniques, including the latest hybrid memory-based and model-based CF recommendation algorithms. The results showed k-means can solve high time complexity but the challenge is centroid selection

3. METHODOLOGY

3.1. POI RECOMMENDATION SYSTEM

POI recommendation system is a subclass of filtering information system that seeks to predict a “rating” or “preference” that a user will possibly assign a rating to an item or location. There are three approaches in the recommendation system, which are collaborative filtering, content-based filtering, and hybrid method. Collaborative filtering is a common approach to overcome problem related to recommendation system [22]-[24]. In addition, collaborative filtering is a technique that broadly used by researchers to develop recommendation system in e-commerce, documents, and others [5, 9, 13]. The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on an issue, A is more likely to have B's 13 opinion on a different issue than that of a randomly chosen person.

Generally, collaborative filtering can be defined as a filtering process using the opinion of other people based on users' past behaviour [25]-[27]. As mentioned before, there are several rating prediction systems used to calculate similarities such as scalar rating or explicit feedback and binary rating or implicit feedback. Explicit rating is a numerical value of 1-5 that represents the rating of an item given by the user actively. While the implicit rating implies that the system automatically obtains passive user preferences by monitoring user action or binary rating by choosing “ever” or “never” [28]. Implicit rating data is obtained from the history of users who visiting locations in certain area. Table 1 gives an example of implicit rating data from several users. The table consists of users and locations.

TABLE 1.
Rating matrix of user-items using implicit feedback

Item/User	M	K	P	B	PB	US	MM
Cole	Ever	Ever	Never	Ever	Never	Never	Never
Jon	Never	Ever	Never	Ever	Never	Ever	Never
Don	Ever	Ever	Never	Ever	Never	Ever	Never
Lisa	Ever	Never	Ever	Ever	Ever	Never	Ever
Kyle	Never	Never	Ever	Never	Never	Ever	Ever

where M, K, P, B, PB, US, MM are Malioboro, Keraton Yogyakarta, Prambanan, Borobudur, Parangtritis Beach, Ulen Sentalu, and Merapi Mountain, respectively. Table 1 can be represented as user-item rating matrix with m user $u_1, u_2, u_3, \ldots\ldots\ldots, u_m$, and n item $i_1, i_2, i_3, \ldots\ldots, i_n$, where each user gives evaluation as rating input in collaborative filtering. Representation is done by changing "Ever" to "1" and "Never" to "0". As shown in Table 2, rating "1" means the user has visited the location and rating "0" means the user has never visited the location.

TABLE 2.
Rating matrix of user-items using implicit feedback

Item/User	i_1	i_2	i_3	i_4	i_5	i_6	i_7
u_1	1	1	0	1	0	0	0
u_2	0	1	0	1	0	1	0
u_3	1	1	0	1	0	1	0
u_4	1	0	1	1	1	0	1
u_5	0	0	1	0	0	1	1

3.2. K-MEANS$^+$

K-Means$^+$ is a combination between Davies-Bouldin Index (DBI) methods to determine the optimal K value (the number of clusters) and K-means algorithm for spatial data clustering. The combination of two methods aim to overcome the weaknesses in the K-means algorithm that the result of clustering is very depending on the value of K defined. Using the DBI method, the most optimal K value of the dataset will be generated so that the clustering results are more optimal. The K-means algorithm is one of an unsupervised algorithm that grouping data based on the cluster's central point (centroid) closest to the data. It is used to have not labelled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. Each centroid defines one of the clusters and each data point is assigned to its nearest centroid iteratively, based on the Euclidean distance squared. Data points are clustered based on feature similarity. The Euclidean distance is defined by following (1).

$$D(x,y) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\|x_i - y_j\|)^2 \quad (1)$$

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

where $\|x_i - y_j\|$ is the Euclidean distance between x_i and y_j, c_i is the number of data points in i^{th} cluster and c is the number of cluster centers (centroid). The next centroid can be calculated by (2).

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i \quad (2)$$

where c_i is the number of data points in i^{th} cluster. Iteration is conducted in one clustering process to the specific threshold that was previously determined. The K-means clustering algorithm is more commonly known for its ability to group large amounts of data quickly and efficiently. This method has been used to solve recommendation system's problem in the movie, e-commerce, music, and others [29, 30]. Meanwhile, Davies-Bouldin Index (DBI) is internal evaluation method that measures a cluster based on cohesion value and separation value. In the grouping process, cohesion is defined as the sum of data closeness/proximity with centroid from following cluster (inter-cluster), while the separation is based on the distance between the centroid and its cluster (intra-cluster) [31]. The distance of intra-cluster can be determined with (3).

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} \{\|x - z_i\|\} \quad (3)$$

From the above formula, C_i is the sum of data from i cluster, z_i is centroid cluster i, and x is a data. The inter-cluster distance is defined by (4).

$$d_{i,j} = \|z_i - z_j\| \quad (4)$$

With z_i is centroid cluster i and z_j is centroid cluster j. Thereafter, Davies Bouldin Index (DBI) can be calculated by the (5).

$$DBI = \frac{1}{K} \sum_{i=1}^{k} R_i \, , qt \quad (5)$$

where K is the number of clusters, and

$$R_i, qt = \max j, j \neq 1 \left\{ \frac{S_i, q + S_j, q}{d_{i,j}, t} \right\} \quad (6)$$

the lowest DBI value, (non-negative ≥ 0), s the cluster obtained from the given value of K [34].

3.3. USER-BASED COLLABORATIVE FILTERING

User-based collaborative filtering (UBCF) is a method of memory-based that uses similarity between users to recommend items. The concept of UBCF is based on inter-user relationships that are analyzed from historical information. This research uses user-based collaborative filtering due to the following advantages [13]:

1. Suitable for offline evaluation using training and testing datasets.
2. The user-based algorithm is easy to implement, simple but produces precise, fairly decent accuracy.

3. Model-based has several disadvantages such as many models are very complex, too sensitive to changes in data and not all model-based theory can be applied with real data.
4. The result of item-based, accuracy is lower than user-based and the number of locations is very large and unbalanced by the number of users, so similarity calculation between items will results in low precision in this dataset.

In a user-based collaborative filtering approach, the process begins by calculating the similarity value between users. The similarity among users can be calculated using Pearson Correlation Coefficient, Cosine similarity, and others. Furthermore, Pearson Correlation Coefficient (PCC) is often used in similarity calculations, but can only be used in the interval or distributed data while cosine similarity more appropriate to calculate binary data [32, 33]. Therefore, we calculate similarity by using cosine similarity by (7).

$$sim\ (a,b) = \frac{n(A \cap B)}{\sqrt{n(A)n(B)}} \quad (7)$$

With *n(A)* is the number of items selected by user *A*, *n(B)* is the number of items selected by user *B*, and *n(A∩B)* is the number of items selected by both A and B users. The Equation (7) can be used to calculate similarity of user 1 based on data from Table 2.

$$sim(Cole, Jon) = \frac{2}{\sqrt{(3)(4)}} = 0.57$$

$$sim(Cole, Don) = \frac{3}{\sqrt{(3)(4)}} = 0.86$$

$$sim(Cole, Lisa) = \frac{2}{\sqrt{(3)(5)}} = 0.52$$

$$sim(Cole, Kyle) = \frac{0}{\sqrt{(3)(3)}} = 0$$

Following the calculation of similarity, a user similarity matrix will be established to quantify the user's nearest neighbors for the predictive generation process. Thus, conclusion can be made from the example, that users who have the highest similarity with Cole are Don, Jon, Lisa and Kyle, respectively.

The model building is derived from the calculation similarity of the entire user that is used as a model for testing. When the testing process, machine learning will seek similarity of the active user from the model that has been created. Once the similarity of active users is found, proceed with generation prediction to build a recommendation system. In user-based collaborative filtering, the nearest neighbor of an active user is selected based on similarity to the user. There are following steps in building recommendations for active users:

1. Find the N nearest neighbor with the greatest similarity value.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

2. Calculate the predicted value of items selected by the nearest neighbor user but has never been selected by the active user, with the (8).

$$Pred(a,p) = \sum_{r \in u} sim(a,u) \times r_u \quad (8)$$

With a is the active user, p is the item that is calculated predictions and r_u is the value "1", indicating that the nearest neighbor user with user u, has already selected item p. From the similarity calculation, it can be calculated the prediction value of each location that has not been visited by the active user (Cole) based on N nearest neighbor.

$$Pred(Cole,P) = \sum_{r \in u} \big(1 \times sim(1,4)\big) + \big(1 \times sim(1,5)\big) = \sum_{r \in u} (1 \times 0.52) + (1 \times 0) = 0.52$$

$$Pred(Cole,PB) = \sum_{r \in u} \big(1 \times sim(1,2)\big) + \big(1 \times sim(1,4)\big) = \sum_{r \in u} (1 \times 0.57) + (1 \times 0.52) = 1.09$$

$$Pred(Cole,US) = \sum_{r \in u} \big(1 \times sim(1,3)\big) + \big(1 \times sim(1,2)\big) + (1 \times sim(1,5))$$
$$= \sum_{r \in u} (1 \times 0.86) + (1 \times 0.57) + (1 \times 0) = 1.43$$

$$Pred(Cole,MM) = \sum_{r \in u} \big(1 \times sim(1,4)\big) + \big(1 \times sim(1,5)\big) = \sum_{r \in u} (1 \times 0.52) + (1 \times 0) = 0.52$$

Prediction is calculated for rating "0" which means the active user has never visited the location. From the calculation results obtained that the prediction for P (Prambanan) is 0.52, PB (Parangtritis Beach) is 1.09, US (Ulen Sentalu) is 1.43 and MM (Merapi Mountain) is 0.52. Further, the value is sorted to generate a location recommendation.

Based on the location candidates, it was selected only the N items with the highest predicted value, which is assumed to have the greatest opportunity value to be chosen by the active user. N items are displayed as a list of recommended items for active users. The calculation result that the recommended location for target user "Cole" are Ulen Sentalu, Parangtritis Beach, Prambanan and Merapi Mountain according to the highest weight.

4. EXPERIMENT

This research uses Gowalla dataset and Foursquare dataset. The Gowalla dataset extracted from the Location-based Social Network (LSBN) that was launched in 2007. The dataset contains 6.264.203 check-in records made by 196.591 users involving 1.280.956 locations over a 627 day time period from February 04, 2009 to October 23, 3 2010. The Gowalla dataset also contains 950.327 friendship link. The Gowalla dataset consists of five attributes namely; user, check-in time, latitude, longitude, and location-id. The second dataset is the Foursquare dataset which has users check-in of New York and Tokyo cities from April 12, 2012 to February 16, 2013. New York City contains 227.428 check-ins and the city of Tokyo contains 573.703 check-ins. This dataset contains eight attributes including User id, Venue

id, Venue category id, Venue category name, Latitude, Longitude, Time zone offset in minutes, UTC time.

The datasets describe the user who check-in at a certain time in a particular location. These datasets are implicit feedback because it contains no rating. In pre-processing, the rating will be built based on existing data to produce a binary rating of "1" and "0". The value of "1" represents that the user has visited a specific location while the value of "0" represents that the user has never visited the location. From the Gowalla dataset, five datasets are taken in different sizes from the center of Austin, Texas, USA within a certain radius. The preprocessing dataset started by scraping data for Austin city, Texas as many as five datasets of varying amounts. From the check-in amount of 6.442.890, this experiment filtered to only 2.561.126 data. The central point is Austin with a radius of between 350 until 3.200 km. The attributes used in this dataset are user-id, location-id, latitude, and longitude. The number of Foursquare datasets used from New York City is 227.428 check-ins with attributes; user id, venue id, latitude, and longitude. Table 3 describes the pre-processing dataset used in this experiment.

TABLE 3.
Rating matrix of user-items using implicit feedback

Size (MB)	Dataset	User	Location	*Check-in*	Radius (km)
Gowalla					
16	Gowalla 1	1537	59972	244462	350
25	Gowalla 2	2295	106452	378118	1200
36	Gowalla 3	2555	166269	512831	2200
45	Gowalla 4	3126	212883	643002	2700
53	Gowalla 5	3418	246369	782713	3200
Foursquare					
15	Foursquare	1083	38333	227428	-

Sparsity and density of datasets can be calculated using (9) and (10).

$$Sparsity = 1 - \frac{number\ of\ checkin}{(number\ of\ users \times number\ of\ location)} \quad (9)$$

$$Density = 1 - Sparsity \quad (10)$$

The datasets have an average sparsity of 99% which means large of data is not filled or considered sparse. In an implicit dataset, the value "0" is not an empty value, but represents that user never visited a particular location. The average density is 0.2% which means the value of "1" in the dataset is 0.2%. A large number of locations and the low intensity of a location visited by the user, resulting in greater sparsity. This problem greatly affects the results of the recommendations and evaluation. Fig. 1 explains a flowchart of the POI recommendation system.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

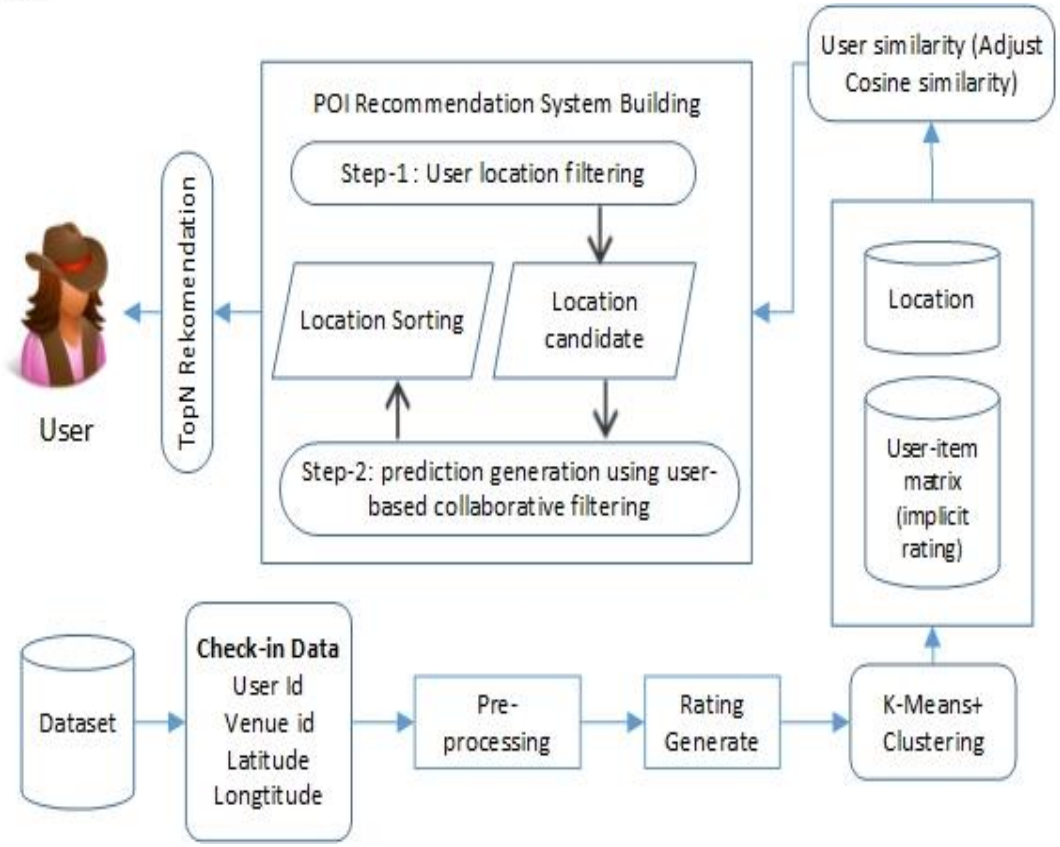


FIGURE 1. POI recommendation system research flow diagram

The datasets are clustered by spatial data (latitude and longitude) using K-Means^{+} after the preprocessing stage. In the POI recommendation stage, a similarity calculation is performed between users using cosine similarity for generation prediction. Precision, recall, and F-Measure are used to evaluate the result of recommendations. This measurement aims to determine the accuracy of resulting recommendations and conformity with user demand. Precision, recall, and F-Measure can be determined by (11), (12) and (13).

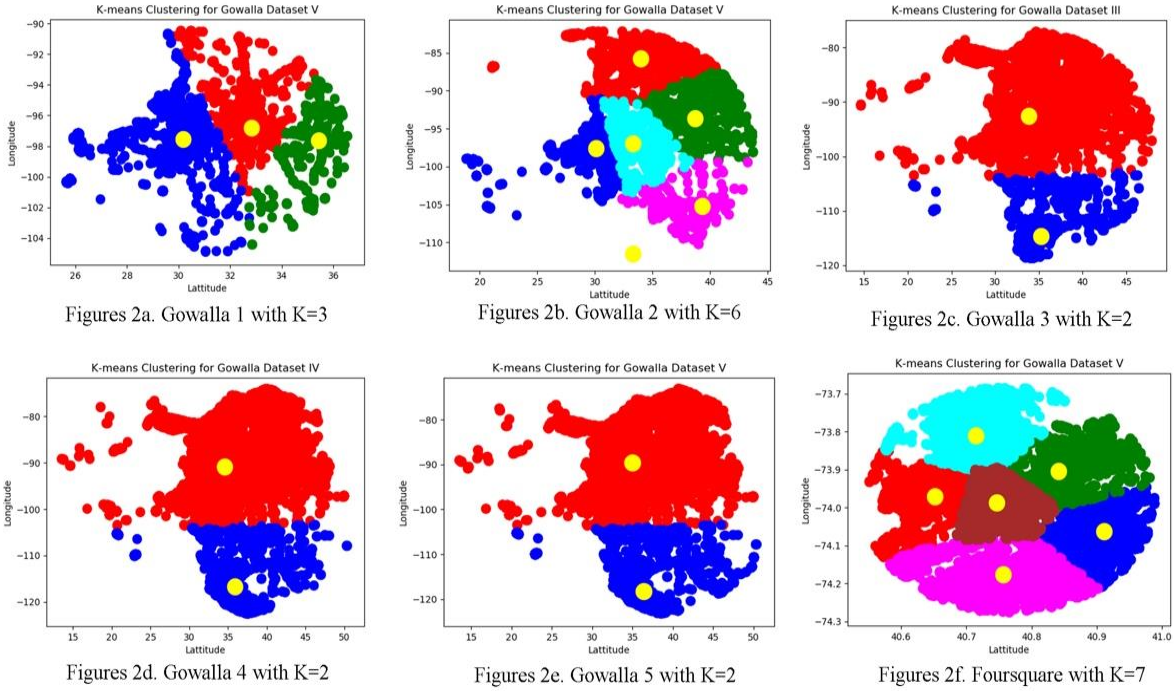
$$precision = \frac{TP}{TP+FP} \tag{11}$$

$$recall = \frac{TP}{TP+FN} \tag{12}$$

$$f - measure = 2 \times \frac{precision \times recall}{precision + recal} \tag{13}$$

5. RESULT AND DISCUSSION

The best cluster value was determined from Equation 5 from DBI method which resulted in the smallest and non-negative value ≥ 0. In this paper, intra-cluster and inter-cluster values were calculated on Gowalla and Foursquare datasets with longitude and latitude attributes. From Table 4 it was observed that there was a variation between the DBI values of datasets with 9 clusters defined because the given dataset was different.



Figures 2a. Gowalla 1 with K=3
Figures 2b. Gowalla 2 with K=6
Figures 2c. Gowalla 3 with K=2
Figures 2d. Gowalla 4 with K=2
Figures 2e. Gowalla 5 with K=2
Figures 2f. Foursquare with K=7

FIGURE 2. The result of dataset clustering using K-Means+

The smallest DBI value of each dataset represents the maximum number of clusters for the dataset. A good cluster is determined from smallest cohesion and the largest separation. The smallest DBI value for Gowalla 1 dataset was 0.413 with cluster number 3, Gowalla 2 was 0.445 with cluster number 6, Gowalla 3 was 0.522 with cluster number 2, Gowalla 4 was 0.507 with cluster number 2, Gowalla 5 was 0.484 with the cluster number of 2 and Foursquare was 0.742 with the number of clusters was 7, respectively.

TABLE 4.
The DBI values of the Gowalla and Fousquare datasets use the K-Means algorithm

	The total number of Cluster								
Dataset	2	3	4	5	6	7	8	9	10
Gowalla 1	0.498	0.413	0.420	0.437	0.473	0.652	0.686	0.555	0.653
Gowalla 2	0.649	0.578	0.548	0.522	0.445	0.560	0.541	0.520	0.495
Gowalla 3	0.522	0.576	0.598	0.573	0.597	0.600	0.616	0.573	0.525
Gowalla 4	0.507	0.537	0.664	0.600	0.657	0.655	0.573	0.540	0.544
Gowalla 5	0.484	0.504	0.662	0.578	0.631	0.531	0.557	0.532	0.537
Foursquare	0.926	0.899	0.747	0.788	0.749	0.742	0.752	0.827	0.838

The K value derived from the optimization will be used for datasets clustering using K-means algorithm. Clustering is used to cluster locations based on spatial parameters thus resulting in better recommendations quality. Clustering aimed to group data according to the user's area when doing check-in and provide location recommendations based on the closest distance to the user. In addition, the clustering is conducted offline to the efficiency of the recommendation process. The

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

result of clustering with each dataset is showed in Fig. 2. Before building a POI recommendation system, the dataset is divided by 8:2 percentage for training and testing data [6, 35, 36].

The User-Based Collaborative Filtering approach is able to generate POI recommendations for users. The system can provide a location recommendation that has not been visited by the user. With a user-based approach, the system can find similarity between users, so the system has many references in recommending locations. This research uses Given-N method [37] to evaluate the algorithm on the sparse data condition by determining the N value of 5, 10, 15, 20, 40 to find the variation of its evaluation. Table 5 is the result of recall evaluation with the value of N were 5, 10, 15, 20, 40, respectively. From Table 5 also appeared that the recall value tends to decrease as the data grow larger.

TABLE 5.
Results of the evaluation of recall on K-Means$^+$ and UBCF method

	Recall				
Dataset	5	10	15	20	40
Gowalla 1	0.042	0.059	0.061	0.064	0.068
Gowalla 2	0.032	0.020	0.021	0.021	0.021
Gowalla 3	0.035	0.041	0.042	0.043	0.044
Gowalla 4	0.030	0.038	0.041	0.041	0.042
Gowalla 5	0.029	0.035	0.036	0.037	0.037
Foursquare	0.021	0.025	0.031	0.033	0.032

Significant decrease occurs in the Gowalla 2 dataset and stabilizes in the 3 to 5 datasets while the recall value of the Foursquare dataset tends to be lower. The highest recall value is the Gowalla 1 dataset which is 0.064 when N = 20. This value can be interpreted that the system succeeded in rediscovering an information is 12 6.4%. In respect with Foursquare, the largest recall value obtained when N = 20 is 0.033. Precision and recall values tended to have inversely proportional value. This is similar to this research that the average of recall value was low while the precision value showed the opposite. The precision value will decrease as the dataset gets more and sparsity gets bigger. Afterward, Table 6 showed that the highest precision value was the Gowalla 1 where the dataset is 0.086 when N = 20. This value can be interpreted that the level of accuracy between the information requested by the user and the answer given by the system is 8.6%. This value is relatively small compared with the research of movie and e-commerce recommendation systems whose precision value reach up to 80% for explicit feedback. However, for implicit feedback, the accuracy was better and increased by 50% compared to previous studies [7, 38, 39].

Quality recommendations can be searched using F-Measure to determine harmonic weights of precision and recall. The F-Measure value ranges of 0 to 1. The quality of recommendation will be better if the F-Measure value approaches the value 1. Table 7 is the F-Measure value of the K-Means$^+$ and UBCF method. As shown in Table 7, the F-Measure value increases with the higher N value on each dataset. It can be explained that the quality of recommendations will be better as

higher the N value. It was also found from table that the highest F-Measure value of the Gowalla 1 dataset is 0.072 when N = 20 with a precision value is 0.086 and recall is 0.064. In the Foursquare dataset, the F-Measure value is 0.036 with a precision value is 0.041 and recall is 0.033.

TABLE 6.
Precision values of Gowalla and Foursquare datasets with N = 5-40

	Precision				
Dataset	5	10	15	20	40
Gowalla 1	0.055	0.067	0.071	0.086	0.076
Gowalla 2	0.032	0.037	0.039	0.039	0.040
Gowalla 3	0.050	0.059	0.061	0.061	0.063
Gowalla 4	0.045	0.057	0.061	0.061	0.062
Gowalla 5	0.047	0.057	0.059	0.060	0.060
Foursquare	0.026	0.031	0.039	0.041	0.040

TABLE 7.
The value of f-measure using Gowalla and Foursquare dataset with N=5-40

	F-*Measure*				
Dataset	5	10	15	20	40
Gowalla 1	0.047	0.062	0.065	0.072	0.072
Gowalla 2	0.024	0.022	0.023	0.023	0.023
Gowalla 3	0.041	0.049	0.050	0.050	0.051
Gowalla 4	0.036	0.046	0.049	0.049	0.050
Gowalla 5	0.036	0.043	0.045	0.046	0.046
Foursquare	0.023	0.028	0.035	0.036	0.035

In this research, the proposed method will be compared with DBSCAN and UBCF [35] using Gowalla 1 and Foursquare datasets with N = 20. Table 7 is a comparison of two methods namely; DBSCAN with UBCF (Method I) and K-Means^{+} with UBCF (Method II). Comparison result showed that the precision value of method I is 0.077 for Gowalla 1 and 0.017 for Foursquare. On the other hand, precision of method II is 0.086 4 for Gowalla 1 and 0.041 for Foursquare dataset. The value indicates that method II improves the precision of recommendation system by 1% from method I. Foursquare dataset, method II produces higher precision value compared with method I because of the functions of K value optimization. Table 8 is a result of evaluation of DBSCAN with UBCF and K-Means^{+} with DBSCAN methods.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

TABLE 8.
The results of evaluation of two methods with the same dataset

Dataset	Method I			Method II		
	DBSCAN + UBCF			KMEANS + UBCF		
	Precison	Recall	F-measure	Precision	Recall	F-measure
Gowalla 1	0.0773	0.0969	0.08605	0.0868	0.0642	0.0728
Foursquare	0.0177	0.0388	0.02434	0.0412	0.0334	0.0369

K-Means$^+$ with UBCF yields a precision value of 0.0868 which means the accuracy level between an information request and an answer to that request information is 8.6%. In this research, POI recommendation systems provide POI recommendations for user's preferences. While the recall value of K-Means$^+$ with UBCF is 0.072 which means the system can find information relevant to the user as much as 7.2% of the number of existing items. The F-Measure value of method I is higher than method II. This is because the sparsity value generated after clustering using DBSCAN tend to be lower, resulting in better recommendations. However, DBSCAN generates noises or outliers which is 66% from Gowalla 1 dataset. Although the results are better, less represent the entire users because of the process of removing outliers. The quality of the POI recommendation system is highly depend on the data processed. If the data has a full rating or low sparsity then the recommendation 1 results will be in accordance with user preferences.

6. CONCLUSION

The proposed method succeeded in giving a recommendation to users according to their preference based on clustering area. Looking at the evaluation results of the proposed method with the previous research, the Davies-Bouldin Index (DBI) method can provide the best number of clusters in the K-means algorithm so that the evaluation results of the POI recommendation system increase. Implicit feedback (binary rating) can be handled using a user-based collaborative filtering approach with K-Means$^+$ clustering by using Davies-Bouldin Index (DBI) method for optimization. K-Means$^+$ algorithm is used in clustering process because K-means is a suitable algorithm for spatial dataset. The Davies-Bouldin Index method provides an optimal K value for K-means to produce a good cluster. It is worth it to be noted that the POI recommendation system using K-Means$^+$ with User-based Collaborative Filtering can generate precise POI recommendations for users. This precision increased by 1% compared with DBSCAN and UBCF. The quality of POI recommendations increased by 7.2%. with F-Measure. The proposed method improves the precision and quality of POI recommendation. However, the percentage of precision value is relatively low compare with explicit recommendation system and needs to be improved using other methods. The problem of sparsity data greatly affects the precision of recommendations so that it needs to be solved by the appropriate method.

ACKNOWLEDGEMENTS

We gratefully appreciate the Faculty of Engineering, Universitas Gadjah Mada, and Indonesia Endowment Fund for Education (LPDP) for providing Research Grant to support this research.

REFERENCES

[1] Jiawei H, Micheline K, Jian P. Data Mining: Concepts and Techniques. San Francisco, CA, USA: Morgan Kaufmann, 23 2012.
[2] Aysun B, Birgul K. Current State and Future Trends in Location Recommender Systems. I.J. Information Tech- 25 nology and Computer Science, 2017; 6: 1-8. doi: 10.5815/ijitcs.2017.06.01
[3] Sulis S, Teguh BA, Igi A. Context-based awareness in location recommendation system to enhance recommendation quality: a review. In: Intenational Conference on Information and Communications Technology; Yogyakarta, Indonesia; 2018. pp. 90-95.
[4] Fernando O, Antonio H, Jesus B, Jeon HK. Recommending items to group of users using matrix factorization based collaborative ltering. Information System. 2016; 345: 313-324. doi: 10.1016/j.ins.2016.01.083
[5] Lakshmi TP, Sreenivasa DP, Siva NN, Srikanth Y. Movie Recommender System Using Item Based Collaborative Filtering Technique. In: International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS); Pudukkottai, India; 2016. pp. 1-5
[6] Huayu L, Yong G, Defu L, Hao L. Learning user ' s intrinsic and extrinsic interests for point-of-interest recommendation: a unified approach. In: The Twenty-Sixth International Conference on Arti cial Intelligence (IJCAI-17); Melbourne, Australia; 2017. pp. 2117{2123.
[7] Yiding L, Tuan-Anh NP, Gao C, Quan Y. An experimental evaluation of point-of-interest recommendation in location-based social networks. In: Proceedings of the VLDB Endowment; 2017. pp. 1021-1021.
[8] Shanshan F, Xutao L, Yifeng Z, Gao C, Yeow MC et al. Personalized ranking metric embedding for next new poi recommendation. In: The Twenty-Fourth International Conference on Artificial Intelligence (IJCAI 2015); Buenos Aires, Argentina; 2015. pp. 2069-2075.
[9] Qilong B, Xiaoyong L, Zhongying B. Clustering collaborative ltering recommendation system based on svd algorithm. In: 4th IEEE International Conference on Software Engineering and Service Science (ICSESS); China; 2013. pp. 963-967.
[10] Fidan k, Gurel Y, Adnan K. A mobile and web application-based recommendation system using color quantization and collaborative ltering. Turkish Journal of Electrical Engineering & Computer Sciences, 2015; 23: 900-912. doi: 10.3906/elk-1212-145
[11] HaiHong E, JianFeng W, MeiNa S, Qiang B, YingYi L. Incremental weighted bipartite algorithm for large-scale recommendation systems. Turkish Journal of Electrical Engineering & Computer Sciences, 2016; 24: 448-463. doi: 10.2906/elk-1307-91

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[12] Gabor T, Istvan P, Bottyan N, Domonkos T. Scalable collaborative filtering approaches for large recommender systems. Journal of Machine Learning Research, 2009; 10: 623-656. doi: 10.1016/j.eswa.2016.09.040
[13] Fidel C, Victor C, Diego F, Vreixo F. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable , high-performance recommender systems. ACM Transactions on the Web, 2011; 5: pp. 1-19 doi: 10.1145/1921591.1921593
[14] Shadi IA, Mohammad M. K-means algorithm with a novel distance measure. Turkish Journal of Electrical Engineering & Computer Sciences, 2013; 21: 1665-1684. doi: 10.3906/elk-1010-869
[15] Maria H, Yannis B, MichaLIS V. Clustering validity checking methods: part II. ACM SIGMOD Record, 2002; 19 31(3); 19-27. doi: 10.1145/601858.6011862
[16] Rui C,Qingyi H, Yan-Shuo C, Bo W, Lei Z, Xiangjie K. A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks, IEEE Access, 2018; 6: 64301- 22 64320. doi:10.1109/ACCESS.2018.2877208
[17] Mao Q, Feng B, Pan S. A study of Top-N recommendation on user behavior data. in: IEEE International Conference on Computer Science and Automation Engineering; Zhangjiajie, China; 2012. pp. 582-586. doi: 10.1109/CSAE.2012.6272839
[18] Philip Z, Yi Z. Bayesian adaptive user pro ling with explicit & implicit feedback. In: ACM CIKM International Conference on Information and Knowledge Management, Arlington, Virginia, USA. 2006. pp. 397-404
[19] Yifan H, Yehuda K, Chris V. Collaborative filtering for implicit feedback datasets. In: Eighth IEEE International Conference on Data Mining; Pisa, Italy. 2008. pp. 263-272
[20] George K. Evaluation of item-based top-n recommendation algorithms. In: the tenth international conference on Information and knowledge management; Atlanta, Georgia, USA. 2001. pp. 247-254
[21] Thorsten J, Laura G, Bing P, Helene H, Geri G. Accurately interpreting clickthrough data as implicit feedback. In: SIGIR '05, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 154-161.
[22] Maria H, Yannis B, MichaLIS V. Clustering validity checking methods: part II. ACM SIGMOD Record, 2002; 31(3); 19-27. doi: 10.1145/601858.6011862
[23] Minakshi P, Anjana NG. Personalized recommender system using collaborative filtering technique and pyramid maintenance algorithm, International Journal of Computer Applications, 2016; 136(8): 25-31. doi: 10.5220/0006513202750282
[24] Tuan HD, Seung RJ, Hyunchul A. A novel recommendation model of location-based advertising: context-aware collaborative filtering using GA approach. Expert System with Applications, 2012; 39(3): 3731-3739. doi: 10.1016/j.eswa.2011.09.070
[25] Selcuk C, Abdulhamit S. Tourism demand modelling and forecasting using data mining techniques in multivariate time series: a case study in Turkey. Turkish Journal of Electrical Engineering & Computer Sciences, 2016; 24: 3388-3404. doi: 10.3906/elk-1311-134

[26] Francesco R, Lior R, Bracha S. Recommender Systems Handbook. New York, USA. Springer, 2011.
[27] Chiu-Chang T, Chi-Fu H, Zong-Han W. Collaborative location recommendations with dynamic time periods. Pervasive and Mobile Computing, 2017; 35(1): 1-14. doi: 10.1016/j.pmcj.2016.07.008
[28] Jian W, Jianhua H, Kai C, Yi Zhou, Zuoyin T. Collaborative filtering and deep learning based recommendation system for cold start items. Expert System with Applications, 2017; 69: 29-39. doi: 10.1016/j.eswa.2016.09.040
[29] Dwi KSU. Item collaborative filtering untuk rekomendasi paket wisata pada franchise tour and travel (in Indonesian). Bachelor, Maulana Malik Ibrahim State Islam University, Yogyakarta, Indonesia, 2014.
[30] Gilda MD, Mehregan M. A new collaborative filtering algorithm using k-means clustering and neighbors ' voting. In: 11th International Conference on Hybrid Intelligent Systems (HIS); Melacca, Malaysia; 2011. pp 179-184.
[31] Georgios P, Xiangliang Z, Wei W. Clustering recommenders in collaborative. IFIP Internationl Federation for Information Processing, 2011; 358: 82-97. doi: 10.1007/978-0-387-85691-96
[32] Xiaonan G, Sen W. Hierarchical Clustering Algorithm for Binary Data Based on Cosine Similarity. In: 8th International Conference on Logistics, Informatics and Service Sciences (LISS); Toronto, Canada; 2018. pp. 1-6. Doi: 10.1109/LISS.2018.8593222
[33] Manzhao B,Shijian L,Ji H. A Fast Collaborative Filtering Algorithm for Implicit Binary Data. In: IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design; Wenzhou, China; 2009. pp. 973-976. Doi: 10.1109/CAIDCD.2009.5374935
[34] Widiarina. Klastering data menggunakan algoritma dynamic k-means (in Indonesian). Jurnal Teknik Komputer AMIK BSI, 2015; 1(2): 260-265. doi: 10.31294/jtk.v1i2.259
[35] Citradevi M, Geetharamani G. An analysis on the performance of kmeans clustering algorithm for cardiotogram data cluster. International Journal on Computational Sciences & Aplications, 2012; 5: 11-20. doi: 10.5121/ijcsa.2012.2502
[36] Zhengwu Y, Haiguang L. Location recommendation algorithm based on temporal and geographical similarity in location-based social networks. In: 12th World Congress on Intelligent Control and Automation (WCICA), Guilin, China, 2016. pp. 1697-1702
[37] Huayu L, Yong G, Richang H, Hengshu Z. Point of interest recommendations: learning potential check-ins from friends. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Fransisco, CA, USA, 2016. pp:975-984
[38] John SB, David H, Cael K. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedding of the 14th Annual Conference on Uncertainty in Artificial Intelligence, San Fransisco, USA, 1998. pp. 43-52
[39] Jia-Dong Z, Chi-Yin C. Point-of-interest recommendations in location-based social networks. SIGSPATIAL Special, 2013; 3: 26-33. doi: 10.1109/CC.2015.7385525
[40] Lei G, Haoran J, Xinhua W, Fangai L. Learning to recommend point-of-interest with the weighted bayesian personalized ranking method in LBSNs. Information, 2017; 8(20) : 1-19. doi: 10.3390/inf8010020

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)