An Improved Items Recommendation for Memory-Based Collaborative Filtering Technique

Kittisak Onuean[1],
Sunantha Sodsee[2],
and Phayung Meesad[3]
Faculty of Information Technology,
King Mongkut's University of Technology North Bangkok, Thailand
[1]on.kittisak@gmail.com
[2]sunanthas@kmutnb.ac.th
[3]phayung.m@it.kmutnb.ac.th

Abstract **- This research aimed to improve Improved Items Recommendation for Memory-based Collaborative Filtering Technique. In this research: 1) to concern rating item about users and 2) to know new prediction rating based on Memory-based Collaborative Filtering Technique. The experimental results from tested propose equation appear γ is [0, 1) and from the equation, γ is equal to 0.2. Get the minimization then recommend item was improved. Mean Absolute Error (MAE) compared to exiting approach was decrease 99.9 %.**

Keywords **- Memory-Based Collaborative Filtering Technique, Mean Absolute Error, Data Sparsity**

I. INTRODUCTION

Nowadays, recommendation system is a system that collects information from the customer's preference in the preferred items and services to predict relevant products or services that might also be preferred. Currently, there are many techniques used to create items recommendation system such as movies, music, programs, websites, tourism, and education. The recommendation system will use preference level or behaviour of the user in calculating relationship from the information to recommend new registered users [1] Currently, both government and private sectors used the recommendation system in creating advantage in the competition by support and create satisfaction for the customer by recommending suitable things for them. There are many types of recommendation system used [2] such as digital government recommendation system and digital business recommendation system. There are many techniques used in the recommendation system including: 1) Content-based (CB) Recommendation Techniques, 2) Collaborative Filtering (CF)-based Recommendation Techniques, 3) Knowledge-based (KB) Recommendation Techniques, 4) Hybrid Recommendation Techniques, 5) Computational Intelligence-based Recommendation Techniques, 6) Social Network-based Recommendation Technique, 7) Context Awareness-based Recommendation Techniques, and 8) Group Recommendation Techniques, which are recommendation systems for the preference of specific group and a recommendation system was created with Collaborative Filtering Technique in this research as well. Main procedures to create recommendation systems with collaborative filtering techniques include similarity computation, neighbors filtering, predicting and recommendations. The performance evaluation of the recommendation system with Collaborative Filtering Technique [3] could be done by analysing on the error value resulted from Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). There is a research on merging user and Item-based Collaborative Filtering to alleviate data sparsity [4], which created memory-based collaborative filtering (CF) that calculate similarity on the basis of users and items. Therefore, in this research, the equation to improve items recommend value was integrated.

II. LITERATURE REVIEWS

A. Memory-Based Collaborative Filtering Technique

Memory-based Collaborative Filtering Technique [5] is a technique that applies preference ranking history of the user towards the items in the form of user preference ranking table, then search for similarity between active users or new products in the system then select K value for the prediction and recommend items for active users, as seen in the below images.

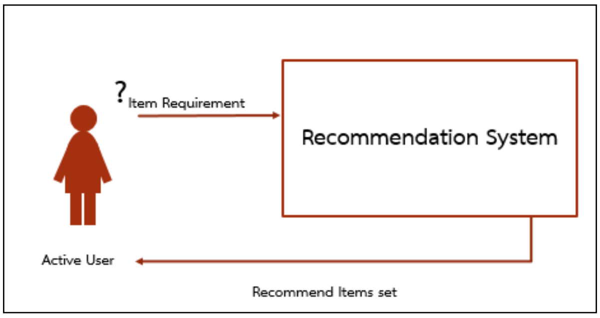

Fig. 1 Item Recommendation Using Memory-Based Collaborative Filtering Technique.

From fig. 1, it is an image showing the process of items recommendation using Memory-based Collaborative Filtering Technique, which started when an active user wants a recommended items from the system. The recommendation system will filter item similar to information of the user to predict according to their preference and recommend them. The process of this system could be seen in fig. 2, work process of Memory-based Collaborative Filtering Technique.

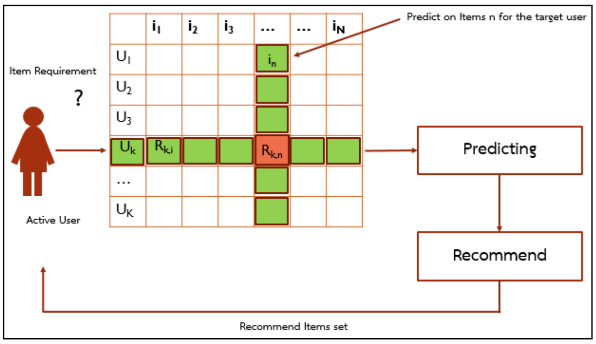

Fig. 2 Work Process of Memory-Based Collaborative Filtering Technique.

From fig. 2, work process of Memory-based Collaborative Filtering Technique, the process will start by using old data in the system and arrange in the form of user and item information table. Then, when a user log in or want to see the recommended list, the system will calculate the similarity of the user with the old data to neighbor filtering and create the recommendation. This process of recommending will use preference ranking of the user and recommend the list according to that. From the said work process of Memory-based Collaborative Filtering Technique, it could be summarized as followed.

B. Similarity Computation

Similarity computation value of active users is a recommendation system process using users' history to find the similarity. The popular algorithms used to find this similarity [1, 6] are: 1) Pearson's Correlation (COR), 2) Cosine (COS), 3) Jaccard (JACC) [7], and 4) Euclidean Distance(DIS) [8].

C. Neighbor Filtering

Neighbor filtering is a selection of data used in the prediction that could be divided into two types [9] including: 1) Nearest Neighbor Algorithms: The search for current user similarity using old data in searching for the similarity and find nearest items to recommend active users as they usually prefer in what recorded in their search history. The nearest neighbor will be acquired in both user's view and product's view and 2) Top-k Recommendation: It is a method that uses data for prediction by determining top-k from the user preference information, which will be calculated from dimensional table of user-product, in which the relationship between users and items will be calculated to find the similarity.

D. Prediction and Recommendation

The prediction and recommendation of item in the recommendation system [6] usually uses two methods in predicting the preference ranking of active users. The Weighted Sum of Others' Ratings (WS) was frequently used for predicting from the user while Simple Weighted Average (SWA) was frequently used for predicting from the product, then test the performance and accuracy of the recommendation system by predicting the preference and recommend the items. The researcher used weight value technique in improving the prediction result by creating the recommendation system. In the research on merging user and Item-based Collaborative Filtering to alleviate data sparsity [4], Similarity-based collaborative filtering was used to create the recommendation system, which is a creation of recommendation system using similarity value based on users and items. The research used this system with each type of data sparsity in the form of a hybrid approach and represent between user-based CF and item-based CF. The presentation of BiUCF testing with the sparseness of data that has density less than 1%. This kind of data usually bear the problem of using preference ranking to predict the items for users. The improvement of prediction and recommendation using BiUCF could be done with the following equation.

$$Pred_{u,i} = w_u \times Pred_{u,i}^U + w_i \times Pred_{u,i}^I \quad (1)$$

$Pred_{u,i}^U$ and $Pred_{u,i}^I$ are output values from preference prediction of User-based CF and Item-based CF while w_u and w_i are weight values output from error value of the real preference ranking. w_u is output from User-based CF from equation (1).

$$w_u = w_u + \gamma(E_{u,i} - \alpha w_u) \quad (2)$$

And w_i is output from Item-based CF from equation (1).

$$w_i = w_i + \gamma(E_{u,i} - \alpha w_i) \quad (3)$$

From equation (2) and (3), it could be seen that w_u and w_i are output values from $E_{u,i} - \alpha w_u$ and $E_{u,i} - \alpha w_i$ in weighing where $E_{u,i}$ could be obtained as below:

$$E_{u,i} = R_{u,i} - Pred_{u,i} \quad (4)$$

From equation (4), $R_{u,i}$ value is the real preference ranking value, then test the performance with MAE and RMSE. In this research, ML 1M data was used (Movie Lens 1M: 6,040 users and 3,952 movies). The first test was conducted on 1,004 users and 1,091 movies.

E. Evaluation

Mean Absolute Error (MAE) [1, 6, 10] is a measure of difference between an accurate value and an estimated value from a sample agent. If MAE is less, it defines that a sample agent could estimate an accurate value similar to the result of the test under an equation.

$$MAE = \frac{1}{S}\sum_{i,j}\left|R_{i,j} - \hat{R}_{i,j}\right| \quad (5)$$

$R_{i,j}$ is defined as a value obtained from a prediction of a sample agent. $\hat{R}_{i,j}$ is thereby an actual value while S is a number of the data used in a sample agent.

Root Mean Square Error (RMSE) is a measure of a discrepancy similar to square root of a standard deviation. It is found to have less value when representing a sample agent and a low value to an actual value as in the below equation.

$$RMSE = \sqrt{\frac{1}{S}\sum_{i,j}\left(R_{i,j} - \hat{R}_{i,j}\right)^2} \quad (6)$$

To define $R_{i,j}$, it is a predicted value from a sample agent. $\hat{R}_{i,j}$ is an accurate value, and S is a number of the data in a sample agent.

Both MAE and RMSE are the measure of discrepancy from both of the averages based on a calculation. If it equals 0, it represents the best value because no discrepancy occurred from the calculation. RMSE results in a higher value than MAE due to the fact that the discrepancy is squared. If considering large discrepancy or huge mistake, RMSE has higher possibility to provide false results.

III. RESEARCH METHODOLOGICAL AND METHODS

For the research methodology and process, the researcher had separated this research into three parts including method of Memory-based Collaborative Filtering Technique, method of recommendation system based on active user preference, and test and evaluation of the research, which could be explained as followed.

A. Memory-Based Collaborative Filtering Technique

The researcher proceeded the development of a collaborative filtering technique that support sparsity data starting with the Memory-based Collaborative Filtering Technique.

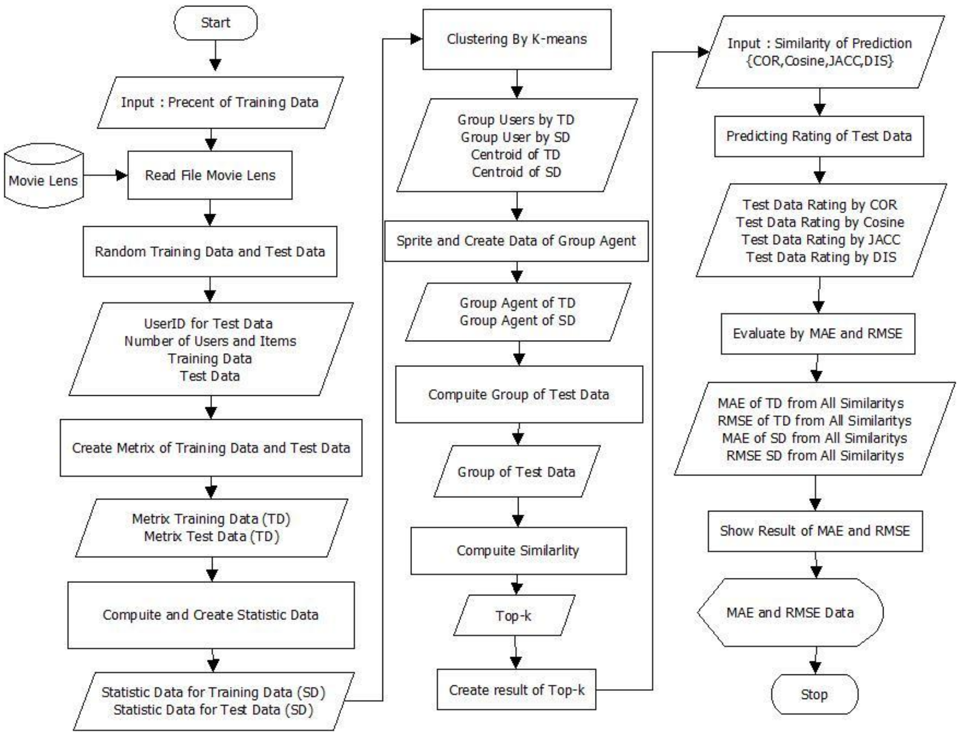

Fig. 3 Method of Memory-Based Collaborative Filtering Technique.

Fig. 3, shows the method of Memory-based Collaborative Filtering Technique starting by determining the percentage of required data for the model creation and testing, next the system will call data from Movie Lens then randomly generate the value according to the initial value and create training data and test data set. After acquiring the data, it will be used in creating the matrix of that data, then use training data in making a statistic dataset (SD), which will yield two data matrix types, traditional data (TD) and statistic data (SD), then cluster. In this research, K-means algorithm [12-13] was used for clustering as it is a popular and general model used for clustering dataset according to preference and K-means clustering process.

K-means clustering would yield representatives from each cluster for data filtering and recommendation. After analysing test data for suitable representation group, similarity would then be calculated and selected from the group. After that, the similarity value of COR, COS, JACC, and DIS would be used to predict preference and recommend the items. The result of that would then be used for evaluation of performance in predicting the said algorithm.

B. Items Recommendation Based on Active User Preference

From the method of Memory-based Collaborative Filtering Technique, it would result in an effective similarity value for predicting the preference of the active users and will be used for recommend items set according to the preference of the active users.

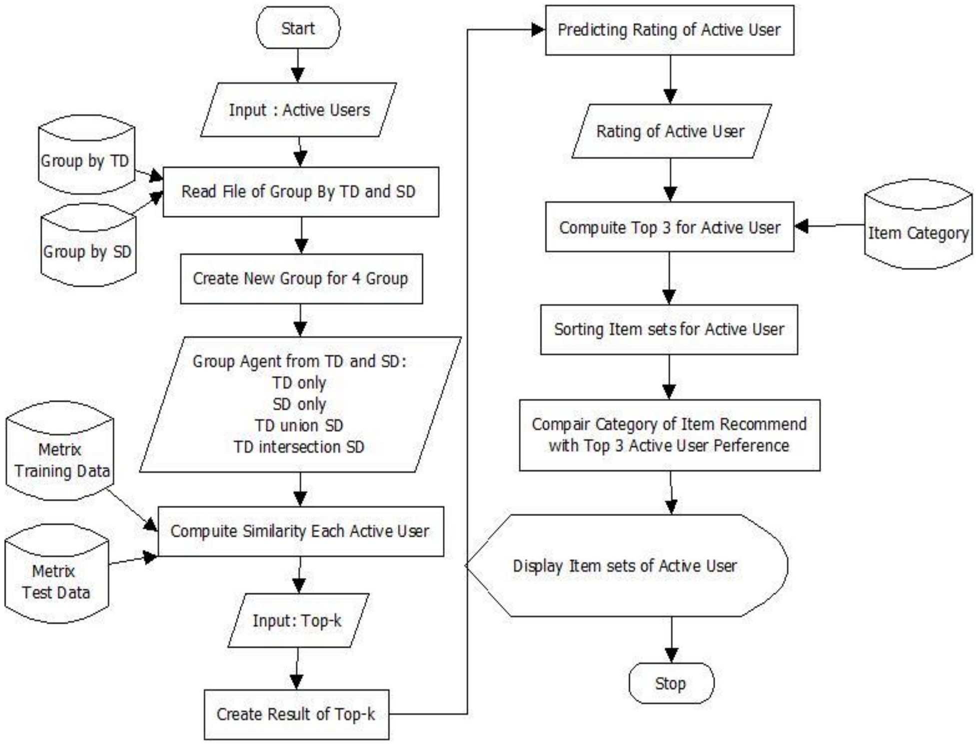

Fig. 4 Method of Recommendation System Based on Active User Preference.

From fig. 4, showing method of recommendation system based on active user preference by randomly select three active users then divide the with TD and SD datasets to create a new cluster with four elements including TD, SD, TD∪SD, and TD∩SD. After that, find the similarity of the users with the representative and neighbor filtering, then predict and recommend the items for the users while finding top three preferences of each user to compare with each user's preference.

C. Improvement Item Rating

The improvement of item rating for item recommendation is a method that uses the process of recommending items for active users. From equation (1), it is an equation used in creating Similarity-based Collaborative Filtering (CF) recommendation system, which weight the user and item similarity value calculation. The research on the collaborative filtering technique that support the modification of data sparsity uses User-based CF technique.

$$Pred_{u,i} = w_u \times Pred^U_{u,i} \tag{7}$$

Therefore, equation (7) was used in the improvement on item rating. The consideration of item set based on users in this research has difference process and data, affecting the difference in the variable. Thus, prediction and recommendation process was improved with equation (8).

$$P_{(new)u,i} = P_{(old)u,i} + 5 * \gamma(R_{u,i} - P_{(old)u,i}) \tag{8}$$

From equation (8), it is the improvement on recommend items from the research on prediction and recommending items by determining new item recommendation $P_{(new)u,i}$ from $P_{(old)u,i}$. The item rating acquired from the prediction of the research on prediction and recommending items 5 is the heaviest weight of item rating dataset $R_{u,i} -$

$P_{(old)u,i}$ is the difference value between real rating and predicted rating and γ modified value for error reduction. In this research, the improvement on the equation for recommend item was done by randomly select the users from active users and use the result of item prediction in the improvement of preference prediction and recommend the items to users as seen in the next image.

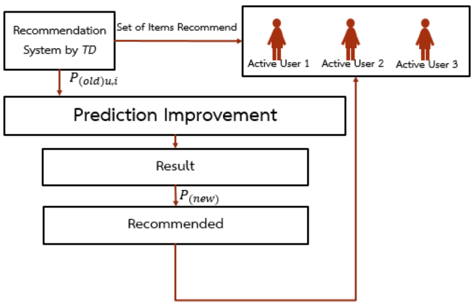

Fig. 5 Improvement on Item Recommend.

From fig. 5, showing the item recommend improvement process by using the prediction result in the improvement process to acquire new item rating for recommending to the users and test the performance of item recommendation.

D. Evaluation

From the aforementioned research improvement method in the test and evaluation of the research, it is a process to test and evaluate results with MAE and RMSE evaluation on item prediction and recommendation and using similarity algorithm COR, COS, JACC, and DIS for testing the improvement of item recommend and recommendation and evaluate the performance of item recommend and recommendation from the improvement of item rating.

IV. RESULT AND DISCUSSION

From the research on item rating and recommendation prediction with old equipment and statistic dataset, the sum of others' rating (WS) was weighted from equation.

$$P_{u_a,i_a} = \bar{r}_{u_a} + \frac{\sum_{h=1}^{n}(r_{u_h,i_a} - \bar{r}_{u_h}) \cdot sim(u_a,u_h)}{\sum_{h=1}^{n}|sim(u_a,u_h)|} \qquad (9)$$

$sim(u_a, u_h)$ which is a similarity value used in predicting preference ranking of the active users.

$$\mathrm{sim}(u_a, u_h) = [\mathrm{COR}(u_a, u_h), \mathrm{COS}(u_a, u_h), \mathrm{JACC}(u_a, u_h), \mathrm{DIS}(u_a, u_h)]$$

The similarity value used in the prediction of item rating with the active user uses the traditional data and statistic data and after the test and evaluation, then evaluate MAE and RMSE on the result of recommend item.

TABLE I
MAE OF PREDICTION IN TD AND SD DATASET

Similarity	Traditional Data (MAE)	Statistics Data (MAE)	|*TD-SD*| (MAE)
COR	0.857178322	1.043255773	0.18607745
COS	0.857225429	1.043223156	0.18599773
JACC	0.872280905	1.049827599	0.17754669
DIS	0.858297143	1.040316747	0.18201960

TABLE II
RMSE OF PREDICTION IN TD AND SD DATASET

Similarity	Traditional Data (RMSE)	Statistics Data (RMSE)	|*TD-SD*| (RMSE)
COR	1.118590972	1.451273272	0.33268230
COS	1.118606444	1.451237410	0.33263097
JACC	1.130071765	1.455747542	0.32567578
DIS	1.120553540	1.446934450	0.32638091

From Table I and Table II, the resulted MAE and RMSE from the error value test from using active user were used to find each similarity. If considering on the error created from each user, it would see that the Euclidean Distance (DIS) similarity gives lower error value of each user than other methods. The research tested γ modification while setting γ in each level as seen in table III from randomly selecting three users from the test data to use in the next modification.

TABLE III
RELATIONSHIP BETWEEN γ AND MAE

γ	User 1 (MAE)	User 2 (MAE)	User3 (MAE)
0.15	0.148598359	0.522447144	0.633570509
0.16	0.118878687	0.417957715	0.506856407
0.17	0.089159015	0.313468287	0.380142305
0.18	0.059439344	0.208978858	0.253428204
0.19	0.029719672	0.104489429	0.126714102
0.195	0.014859836	0.052244714	0.063357051
0.198	0.005943934	0.020897886	0.02534282
0.199	0.002971967	0.010448943	0.01267141
0.19999	2.97197E-08	1.04489E-07	1.26714E-07
0.2	**0**	**1.16069E-16**	**2.13129E-17**
0.201	0.002971967	0.010448943	0.01267141

From Table III, showing MAE from randomly selecting three users in item recommendation by modifying γ to reduce error produced from the recommendation. The appropriate γ for the improvement is at 0.2. Therefore, γ is 0.2 and compare MAE value from the modification on item recommend and MAE from the existing approach as seen in next table.

TABLE IV
MAE AND RMSE FROM THE MODIFLYING γ ON ITEM PREDICTION AND RECOMMENDATION

Similarity	Existing Approach	Propose Approach
DIS (MAE)	1.739488017	4.57939E-17
DIS (RMSE)	2.069314298	1.74581E-14

From Table IV, MAE and RMSE from the modification on item rating and item recommendation prediction allow us to know that after the modification get the minimization then recommend item was improved. Mean Absolute Error (MAE) compared to exiting approach was decrease 99.9 %.

V. CONCLUSIONS

The aim of this research was to improve Improved Items Recommendation for Memory-based Collaborative Filtering Technique. Currently, recommendation system was created to make it popular among the users by recommending the items with the most likelihood to be matched with the needs of the customer. The popular technique used for creating a recommendation system by using the history is Memory-based Collaborative Filtering Technique, which is a process with 3 steps including similarity computation, neighbor selection, and prediction and recommend. The creation of this technique has its model performance evaluation by the accuracy value, in which the error from recommendation will be calculated. The evaluation usually uses Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in the process. In this research, we propose equation for decrease MAE in equation (8) and modifying γ is [0, 1) and from the equation, γ is equal to 0.2. to reduce MAE compared to exiting approach was decrease 99.9 %.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). “Recommender systems survey”. Knowl.-Based Syst., Vol. 46, pp. 109-132.

[2] Lu, J., Wu, D., Mao, M., Wang, W., and Zhang, G. (2015). “Recommender system application developments: A survey”. Decis. Support Syst., Vol. 74, pp. 12-32.

[3] Niu, J., Wang, L., Liu, X., and Yu, S. (2016). “FUIR: Fusing user and item information to deal with data sparsity by

using side information in recommendation systems". J. Netw. Comput. Appl., Vol. 70, pp. 41-50.

[4] Kant, S. and Mahara, T. (2018). "Merging user and item based collaborative filtering to alleviate data sparsity". Int. J. Syst. Assur. Eng. Manag., Vol. 9(1), pp. 173-179.

[5] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). "Item-based Collaborative Filtering Recommendation Algorithms". in Proceedings of the 10th International Conference on World Wide Web, New York, NY, USA, pp. 285-295.

[6] Su, X. and Khoshgoftaar, T.M. (2009). "A Survey of Collaborative Filtering Techniques". Adv Artif Intell, Vol. 2009, pp. 4:2-4:2.

[7] Patra, B.K., Launonen, R., Ollikainen, V., and Nandi, S. (2015). "A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data". Knowl.-Based Syst., Vol. 82, pp. 163-177.

[8] da Silva, E.Q., Camilo-Junior, C.G., Pascoal, L.M.L., and Rosa, T.C. (2016). "An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering". Expert Syst. Appl., Vol. 53, pp. 204-218.

[9] Tatiya, R.V. and Vaidya, P.A.S. (2014). "A Survey of Recommendation Algorithms". IOSR J. Comput. Eng., Vol. 16(6), pp. 16-19.

[10] Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. (2004). "Evaluating Collaborative Filtering Recommender Systems". ACM Trans Inf Syst, Vol. 22(1), pp. 5-53.

[11] Willmott, C.J. and Matsuura, K. (2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance". Clim. Res., Vol. 30(1), pp. 79-82.

[12] Ordonez, C. (2003). "Clustering Binary Data Streams with K-means". in Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, New York, NY, USA, pp. 12-19.

[13] Sculley, D. (2010). "Web-scale K-means Clustering". in Proceedings of the 19th International Conference on World Wide Web, New York, NY, USA, pp. 1177-1178.