



A FCA-Based Concept Clustering Recommender System

G. Chemmalar Selvi^{1(✉)}, G. G. Lakshmi Priya¹,
and Rose Bindu Joseph²

¹ Vellore Institute of Technology, Vellore 632014, TN, India
chemmalar06@gmail.com

² Christ Academy Institute for Advanced Studies, Bengaluru 560083, India

Abstract. Recommender systems are information filtering software which is capable of resolving the recent issue of internet's information overload. The recommender system generate the recommendation more suitably based on the data gathered either implicitly like user profile, click information, web log history or explicitly like ratings (scale 1–5), likes, dislikes, feedbacks. The most important challenge to the recommender system is the growing number of online users making it a high dimensional data which leads to the data sparsity problem where the accuracy of recommendation depends on the availability of the data. In this paper, a new approach called formal concept analysis is employed to handle the high dimensional data and a FCA-based recommender algorithm, User-based concept clustering recommendation algorithm (UBCCRA) is proposed. The UBCCRA out performs by accurately generating the recommendation for the group-based users called cluster users. The experimental result is shown to prove the cluster recommendation with good result.

Keywords: Recommender system · Collaborative filtering · Clustering · Formal concept analysis · Sparsity

1 Introduction

The torrent increase in the enormous volume of digital information and the count of internet or online users have set a drastic confrontation with the information overload [1] which restricts the need of an hour to access to interest of certain items on the internet. There are few information retrieval systems like Google and YouTube who have targeted this problem and tried to provide likely solutions by finding, filtering, prioritizing and personalizing according to the online users. They simply associate the item contents with the user's profile and interest. This has increased the urge for recommender systems than even ever before.

Recommender Systems are known to be highly predictive and personalized software which is capable of interacting with the overwhelming data flowing in the internet. The Recommender Systems provide more suitable suggestions of items to the online users who are more of skeptical attitude when the choices are plenty around them [2]. They have the ability to predict whether a particular user would like an item or not based on the information acquired explicitly like ratings or implicitly like user

history, weblog or click depth [3]. Thus, they are essential to both service providers as well as to the users [4]. It has well decreased the transaction costs when identifying and preferring the items to buy and has also improved the strong decision making process and quality [5].

The application of faultless recommendation algorithms is very much essential for any recommender system to offer suitable and exact recommendation to its individual users. The recommendation algorithms are broadly classified into content-based [6], collaborative filtering-based [7] and hybrid-based algorithms [8]. In this paper, the application of collaborative filtering-based algorithm is presented by combining a new approach based on Formal Concept Analysis (FCA). In general, the people who buy similar items are formed as clusters which pave the way for performing most prominent and suitable recommendations of new items to be bought. Formal Concept Analysis (FCA) is a mathematical tool which is used for analyzing the large datasets and representation of complex information [9].

In recent years, the application of FCA over recommender system has started evolving. In the paper [10], the author has studied the friends' recommendations by presenting the CCG (concept context graph) which will represent the knowledge context preferably from the social interactions in order to recommend more friends. The concept similarity is computed and ranked among the matching concepts in the concept context graph of the target users for recommendation. The authors [11] discussed the exploration of FCA in the collaborative filtering-based algorithm by representing the relation between the users and items and have constructed two new algorithms for finding neighbours in a collaborative recommender. The algorithm has shown better result in finding out the neighbours without much loss in the result accuracy.

The rest of the paper is organized to explain more in detail about the application of new FCA-based approach over the collaborative filtering-based recommender algorithm. The first section of the paper provides the foundations and terminologies in FCA approach. In the second section, a new FCA-based collaborative filtering-based algorithm is proposed. In the third section, an algorithm to deal with recommender system using collaborative filtering-based is presented by illustrating with the example dataset. Finally, the conclusion of this paper provides the summary of the proposed work and its success in recommender systems.

2 An Overview of Formal Concept Analysis

Formal Concept Analysis (FCA) is a mathematical approach which defines the complex data representation and was first introduced by Wille and Ganter [12]. The one such property of FCA is to tackle uncertainty of data and it has been applied widely in the areas of fuzzy concept, interval-valued fuzzy concept, and rough set, approximate and triadic concept analysis [13–17]. The fundamental aspect of FCA are the formal context and concept hierarchy. The formal context consists of set of objects, set of attributes and binary relation among with objects and attributes. The followings describes the important terminologies in FCA.

Definition 1 Formal Context. A formal context is a triplet (P, Q, R) , where P and Q are two nonempty and finite sets, and R is a binary relation on P and Q (Refer Fig. 1).

R	q ₁	q ₂	q ₃	q ₄	q ₅
p ₁	X	X		X	X
p ₂	X	X	X		
p ₃				X	
p ₄	X	X	X		

Fig. 1. Formal context M

Here, P and Q are the set of objects and set of attributes, respectively where $(p, q) \in R$ indicate that the object p has the attribute q and $(p, q) \notin R$ indicate that the object p does not have the attribute q .

Definition 2 Concept-Forming Operators. For every formal context (P, Q, R) , the concept-forming operators $\uparrow: 2^P \rightarrow 2^Q$ and $\downarrow: 2^Q \rightarrow 2^P$ for every $A \subseteq P$ and $B \subseteq Q$ as:

$$A^\uparrow = \{q \in Q \mid \text{for each } p \in A : (p, q) \in R\}$$

$$B^\downarrow = \{p \in P \mid \text{for each } q \in B : (p, q) \in R\}$$

The operator \uparrow assigns the subsets of Q to subsets of P . A^\uparrow is the set of all attributes shared by all the objects in A . Analogously, the operator \downarrow assigns the subsets of p to subsets of Q . B^\downarrow is the set of all objects shared by all the attributes in B .

Definition 3 Formal Concept. Formal concept is a pair (A, B) for any formal context (P, Q, R) such that $A \subseteq P$ and $B \subseteq Q$ where $A^\uparrow = B$ and $B^\downarrow = A$.

Here (A, B) is the formal concept if and only if set A consist of only those objects that are shared by all the attributes shared from the set B which is called extent of the concept. Also, set B consist of only those attributes that are shared from the set A which is called intent of the concept.

The corresponding formal context M represented in Fig. 1, contain the following formal concepts:

$$C_0 = (\{\}, \{q_1, q_2, q_3, q_4\}), C_1 = (\{p_1\}, \{q_1, q_2, q_4, q_5\}), C_2 = (\{p_2, p_4\}, \{q_1, q_2, q_3\}),$$

$$C_3 = (\{p_1, p_3\}, \{q_4\}), C_4 = (\{p_1, p_2, p_4\}, \{q_1, q_2\}), C_5 = (\{p_1, p_2, p_3, p_4\}, \{\})$$

Thus, the concept-forming operators can be used to define formally the concept's extent as $\text{Ext}(P, Q, R) = \{B^\downarrow \mid B \in Q\}$ and concept's intent as $\text{Int}(P, Q, R) = \{A^\uparrow \mid A \in P\}$.

Definition 4 Subconcept-Superconcept. For the given formal context (P, Q, R) , if (A_1, B_1) and (A_2, B_2) are the formal concepts then it holds a partial order relation \leq such as: $(A_1, B_1) \leq (A_2, B_2)$ if and only if $A_1 \subseteq A_2$ and $B_2 \subseteq B_1$.

This partial order relation formally describes the subconcept - superconcept relationship which can be represented by drawing the line or Hasse diagram called concept lattice.

Definition 5 Concept Lattice. The concept lattice is defined as the hierarchical structure of partial-order relation \leq denoted by $\beta(P, Q, R)$ such that $\{(A, B) \in 2^P \times 2^Q \mid A^\uparrow = B \text{ and } B^\downarrow = A\}$. The concept lattice for the corresponding formal context M is represented in the Fig. 2.

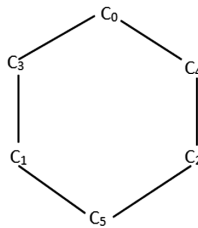


Fig. 2. Concept lattice

Thus, the FCA has been well explored in the following areas: text mining, web mining, software mining, bioinformatics, chemistry, medicine, ontology engineering, and so on [9]. In recent years, the application of FCA to recommender system has a massive attention and is still growing area in the research topics. In the next section, the paper discusses about the application of such FCA to recommender system which uses collaborative filtering technique.

3 Application of FCA with Collaborative Filtering-Based Recommender System

Collaborative filtering (CF) in recommender systems use the tabular data structure called as user-item rating matrix. The intuition behind working this algorithm is purely based on the ratings that are given by the active users when they bought or purchased the items. But in the real-world, it is rare to observe the ratings in the user-item rating matrix. Hence the nature of the user-item rating matrix becomes sparse which is the most challenge faced by the recommender system in today service. Generally, the CF technique is classified as Memory-based methods which predict the ratings of the active user's by matching the similar users with each other's [18] and Model-based methods which aim at predicting the ratings of the active users from the kind of pattern generated from the user-item rating matrix [19].

The novelty of this paper is applying a new and efficient mathematical approach called formal concept analysis into the field of recommender system which can effectively solve the most challenge faced by the recommender system called data sparsity and high-dimensional problem [20]. The general schema of user-based recommendation is used to build the presented system model [21, 27–30]. The input data to any recommender system is the user-item rating matrix where users are represented in the rows and items are represented in the columns and the values are represented in the corresponding cells of the matrix. Such a data representation is facilitated by the formal context in the FCA. The formal context $M = (P, Q, R)$ denotes that P containing the set of users = {Allen, Bob, Tom, Alice}, Q containing the set of items (movies) = { m_1, m_2, m_3, m_4, m_5 } and R gives the relation between user and item (Fig. 3).

R	m_1	m_2	m_3	m_4	m_5
Allen	5	3		2	4
Bob	3	1	4		
Tom				5	
Alice	2	5	3		

Fig. 3. Formal context M representing user-item rating matrix

The paper proposes a new clustering algorithm based on the formal concept analysis known as User-based concept clustering recommendation algorithm (UBC-CRA). The idea behind this proposed algorithm is to cluster the similar or like-minded users based on the similar items purchased by these users. Based on the number of clusters formed, the similarity measure like most popularly used Pearson or cosine similarity measures can be induced to find the deepness of the similarity between each clusters [22–26]. Finally, the ranking method like Top-N Recommendation is used to recommend the items to the target users in the clusters. The User-based concept clustering recommendation algorithm (UBCCRA) is given below:

Algorithm 1 (User-based concept clustering recommendation algorithm (UBC-CRA))

Input: User-Item rating matrix
 Output: Top-N Recommendation
 Processing Steps:

1. Take the user-item rating matrix M and transform it as formal context $M' = (U, I, R)$ where U comprise of set of active users, I comprise of set of items and R stands for relation existing between user U and item I . If the user U has purchased an item I , $(u, i) \in R$ then mark it 1: $(u, i) \notin R$ else 0. Thus the dimension of the formal context M' is $U \times I$.
2. Generate the formal concepts from the formal context M' by using:
 $Tu^{\uparrow} = \{i \in I \mid \text{for each } u \in Tu : (u, i) \in R\}$
 $Ti^{\downarrow} = \{u \in U \mid \text{for each } i \in Ti : (u, i) \in R\}$
3. Form the clusters of similar users purchased similar items by removing the redundant formal concepts (Tu, Ti) : $Tu \subseteq U$ and $Ti \subseteq I$ where $Tu^{\uparrow} = Ti$ and $Ti^{\downarrow} = Tu$
4. Calculate the weighted concept cluster similarity $Sim_{\hat{w}}$ by applying the similarity measure Jaccard Index as shown below:

$$Sim_{\hat{w}}(CC1, CC2) = \hat{w} * S(Tu1, Ti1) + (1 - \hat{w}) * S(Tu2, Ti2)$$

where the weight \hat{w} is from $0 \leq \hat{w} \leq 1$, concept clusters $CC1 = (Tu1, Ti1)$, $CC2 = (Tu2, Ti2)$ and S is the Jaccard Similarity Index computed as:

$$Jaccard\ Similarity\ Index, S = \frac{|Tu1 \cap Tu2|}{|Ti1 \cup Ti2|}$$

5. Use the ranking method and select the top most concept cluster to perform the cluster-based recommendation.

The above presented algorithm performs the User-based concept clustering recommendation. The main contribution of this algorithm is to reduce the high dimensional data and simplify the complex data representation while analyzing the data. The above algorithm can be modified to perform Item-based concept clustering recommendation by transposing the user-item rating matrix where items are represented as rows and users are represented as columns. In the next section explains the User-based concept clustering recommendation algorithm (UBCCRA) by experimenting with a small dataset.

4 Experimental Results and Discussion

In this section, we have considered a small dataset which consist of 12 users and 8 movies in the movie recommendation system. The User-based concept clustering recommendation algorithm (UBCCRA) is applied to this small dataset to perform

movie recommendation to the group-based users with the similar interest of preference. The user-item rating matrix M is converted to formal context M' and represented as below (Figs. 4 and 5):

5	3	2	0	1	0	0	0
1	4	0	0	0	1	0	3
0	5	0	0	4	0	3	0
3	2	0	0	0	2	0	1
5	3	2	0	1	0	0	0
0	2	0	0	4	0	2	0
0	5	0	0	2	0	5	0
0	0	0	0	0	0	4	0
4	3	5	0	5	0	0	0
0	0	0	0	0	0	2	0
2	1	3	0	5	0	0	0
3	2	0	0	0	3	0	5

Fig. 4. User-Item rating matrix M

x	x	x		x			
x	x				x		x
	x			x		x	
x	x				x		x
x	x	x		x			
	x			x		x	
	x			x		x	
							x
x	x	x		x			
							x
x	x	x		x			
x	x				x		x

Fig. 5. Formal context M' of User-Item rating matrix M

From the formal context M' , the clusters are formed by generating the formal concepts. The formal concepts are generated by matching the similar kind of items purchased by the similar kind of active users. From the generated formal concepts, the redundant formal concepts are obtained by applying the sub-super concept partial ordering relationship. Hence, the partial ordering gives the actual formal concepts which form the cluster with similar users and similar items without loss of data. Initially, there were 4096 formal concepts which are redundant in nature was constructed, of which only 9 formal concepts was considered as concept clusters after applying partial ordering. In the Fig. 6 the concept clusters are illustrated for the given dataset of size of 12×8 matrix.

CC_i	(Tu_i, Ti_i)
CC_0	($\{\}$, $\{1, 2, 3, 4, 5, 6, 7, 8\}$)
CC_1	($\{1, 5, 9, 11\}$, $\{1, 2, 3, 5\}$)
CC_2	($\{2, 4, 12\}$, $\{1, 2, 6, 8\}$)
CC_3	($\{3, 6, 7\}$, $\{2, 5, 7\}$)
CC_4	($\{3, 6, 7, 8, 10\}$, $\{7\}$)
CC_5	($\{1, 3, 5, 6, 7, 9, 11\}$, $\{2, 5\}$)
CC_6	($\{1, 2, 4, 5, 9, 11, 12\}$, $\{1, 2\}$)
CC_7	($\{1, 2, 3, 4, 5, 6, 7, 9, 11, 12\}$, $\{2\}$)
CC_8	($\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, $\{\}$)

Fig. 6. Concept clusters of formal context M'

For the concept clusters CC_i , the weighted concept cluster similarity Sim_w is applied which is used to find the deepness of similarity among the users and items between each clusters. Let the similarity between the concept clusters $CC1$ and $CC5$ be $\text{Sim}_w(CC1, CC5) = 0.57$, where the weight is assumed to be 0.5 since the rating may be biased based on the feedback of recommendation. Finally the calculated similarities among the every cluster are ranked in the descending order and the top most concept with high similarity value is considered based on which the recommendation is generated to every of these clusters. Hence, the cluster-based recommendation is completed which might suitably fit the cluster users since the recommendation is generated based on the similar list of user preferences in the target cluster.

5 Conclusion

Collaborative filtering is the most popular recommender algorithm where there is an automatic generation of recommendation for the target users based on the similar taste of interest or preferences of the other active users. But the lack of user preferences or ratings boils down the accuracy of recommendation. In this paper, a new FCA-based concept clustering algorithm is proposed which is able to resolve the high dimensional data and thereby identifies the clusters of users with the similar user's interest or similar user's preference. The formal context is constructed for the given user-item rating matrix which has well addressed the high dimensional data. The method of generating the clusters is based on finding the non-redundant formal concept clusters. The deepness of similarity exhibited by each cluster is predicted by applying the weighted concept similarity and finally the top most similarity score is utilized to perform the recommendation for every cluster. In the future work, per-user based recommender algorithm can be developed using the FCA method.

References

1. Konstan, J.A., Riedl, J.: Recommender systems: from algorithms to user experience. *User Model. User-Adapt. Interact.* **22**, 101–123 (2012)
2. Pan, C., Li, W.: Research paper recommendation with topic analysis. In: *Computer Design and Applications*, pp. V4–V264. IEEE (2010)
3. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys 2011*, pp. 57–164. ACM, New York (2011)
4. Hu, R., Pu, P.: Potential acceptance issues of personality-ASED recommender systems. In: *Proceedings of ACM Conference on Recommender Systems, RecSys 2009*, New York City, NY, USA, p. 225, October 2009
5. Pathak, B., Garfinkel, R., Gopal, R., Venkatesan, R., Yin, F.: Empirical analysis of the impact of recommender systems on sales. *J Manag. Inform. Syst.* **27**(2), 159–188 (2010)
6. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *The Adaptive Web. LNCS*, vol. 4321, pp. 325–341. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_10

7. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. In: *Advances in Artificial Intelligence* (2009)
8. Kilani, Y., Otoom, A.F., Alsarhan, A., Almaayah, M.: A genetic algorithms-based hybrid recommender system of matrix factorization and neighborhood-based techniques. *J. Comput. Sci.* **28**, 78–93 (2018)
9. Poelmans, J., Ignatov, D.I., Kuznetsov, S.O., Dedene, G.: Formal concept analysis in knowledge processing: a survey on applications. *Expert Syst. Appl.* **40**(16), 6538–6560 (2013)
10. Zhang, W., Du, Y.J., Song, W.: Recommender system with formal concept analysis (2015)
11. du Boucher-Ryan, P., Bridge, D.: Collaborative recommending using formal concept analysis. In: *bramer, m, Coenen, F., Allen, T. (eds.) SGA1 2005*, pp. 205–218. Springer, London (2006). https://doi.org/10.1007/978-1-84628-226-3_16
12. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin (1999). <https://doi.org/10.1007/978-3-642-59830-2>
13. Doerfel, S., Jäschke, R., Stumme, G.: Publication analysis of the formal concept analysis community. In: *Domenach, F., Ignatov, Dmitry I., Poelmans, J. (eds.) ICFA 2012. LNCS (LNAI)*, vol. 7278, pp. 77–95. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29892-9_12
14. Belohlavek, R., Vychodil, V.: Formal concept analysis and linguistic hedges. *Int. J. Gen Syst.* **41**(5), 503–532 (2012)
15. Ignatov D.I., Gnatyshak, D.V., Kuznetsov, S.O., Mirkin, B.G.: Triadic formal concept analysis and triclustering: searching for optimal patterns. *Mach. Learn.* (2015). <https://doi.org/10.1007/s10994-015-5487-y>
16. Yao, Y.: A comparative study of formal concept analysis and rough set theory in data analysis. In: *Tsumoto, S., Słowiński, R., Komorowski, J., Grzymała-Busse, Jerzy W. (eds.) RSCTC 2004. LNCS (LNAI)*, vol. 3066, pp. 59–68. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25929-9_6
17. Djouadi, Y.: Extended Galois derivation operators for information retrieval based on fuzzy formal concept lattice. In: *Benferhat, S., Grant, J. (eds.) SUM 2011. LNCS (LNAI)*, vol. 6929, pp. 346–358. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23963-2_27
18. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In: *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pp. 175–186. ACM, October 1994
19. Si, L., Jin, R.: Flexible mixture model for collaborative filtering. In: *Proceedings of the 20th International Conference on Machine Learning, ICML 2003*, pp. 704–711 (2003)
20. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. *Knowl.-Based Syst.* **46**, 109–132 (2013)
21. Ignatov, D.I., Nenova, E., Konstantinova, N., Konstantinov, A.V.: Boolean matrix factorisation for collaborative filtering: an FCA-based approach. In: *Agre, G., Hitzler, P., Krisnadhi, A.A., Kuznetsov, S.O. (eds.) AIMS 2014, LNCS*, vol. 8722, pp. 47–58. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10554-3_5
22. Formica, A.: Concept similarity in formal concept analysis: an information content approach. *Knowl.-Based Syst.* **21**(1), 80–87 (2008)
23. Tadrat, J., Boonjing, V., Pattaraintakorn, P.: A new similarity measure in formal concept analysis for case-based reasoning. *Expert Syst. Appl.* **39**(1), 967–972 (2012)
24. Wang, L., Liu, X.: A new model of evaluating concept similarity. *Knowl.-Based Syst.* **21**(8), 842–846 (2008)

25. He, H., Hai, H., Rujing, W.: FCA-based web user profile mining for topics of interest. In: 2007 IEEE International Conference on Integration Technology, pp. 778–782. IEEE, March 2007
26. Lin, Z.C., Zhu, G.J.: A concept similarity algorithm based on FCA. *Comput. Technol. Dev.* **9**, 112–114 (2008)
27. Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J.: Item-based collaborative filtering recommendation algorithms. *WWW* **1**, 285–295 (2001)
28. Ma, H., King, I., Lyu, M.R.: Effective missing data prediction for collaborative filtering. In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 39–46. ACM, July 2007
29. Wang, J., De Vries, A.P., Reinders, M.J.: Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 501–508. ACM, August 2006
30. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. *Commun. ACM* **35**(12), 61–71 (1992)