The Efficiency of a DoParallel Algorithm and an FCA Network Graph Applied to Recommendation System

Phonexay Vilakone 1 and Doo-Soon Park 2,*

1 Department of Computer Sciences and Engineering, Soonchunhyang University, Soonchunhyang-Ro 22, Sinchang-Myeon, Asan-Si 31538, Chungcheongnam-Do, Korea; xayus@yahoo.com
2 Department of Computer Software Engineering, Soonchunhyang University, Soonchunhyang-Ro 22, Sinchang-Myeon, Asan-si 31538, Chungcheongnam-Do, Korea
* Correspondence: parkds@sch.ac.kr; Tel.: +82-41-530-4923

Received: 16 March 2020; Accepted: 22 April 2020; Published: 23 April 2020

Abstract: This article investigates the efficiency of a doParallel algorithm and a formal concept analysis (FCA) network graph applied to recommendation systems. It is the first article using the FCA method to create a network graph and apply this graph to improve the accuracy of a recommendation system. According to the fundamental knowledge about users who have similar feature information, they may like the same items. This idea was used to create an FCA network graph. In the proposed process, the k-clique method was used to divide this network graph into various communities. A combination of the k-nearest neighbor and the betweenness centrality methods was used to find a suitable community for a new user based on the feature information similarities between the new user and an existing user in each community. Finally, a data mining method created a list of items from suitable communities and recommended them to the new user. In essence, the execution in this article uses a doParallel algorithm as a mechanism in parallel processing technology. The result of the implementation is satisfactory. It proved that the proposed method could resolve the cold-start problem in a recommendation system and may overcome the vast time consumption when a huge dataset is involved.

Keywords: betweenness centrality; data mining; formal concept analysis; k-clique; k-nearest neighbor; parallel processing; recommendation system

1. Introduction

A recommendation system is a primary function in the field of business over the Internet or e-commerce [1–3]. It acts as a tool for filtering, prioritized and efficiently deliver relevant information that matches the requirement, interest and preference of their customer or users. In addition, this method can enhance the revenues of selling more product. It earns a lot of benefits to the business owner or service providers and users, especially in the field of online shopping environment, online movie, drop shipping, e-publishing, affiliate marketing, dating matching, search engine [4,5], etc. In the scientific library, it supports users to move beyond catalog searches. Inside the methodology, there are various methods, such as content-based filtering, collaborative filtering and hybrid filtering [6]. Collaborative filtering is the most powerful and widely used technique in the recommendation system [7–9]. Collaborative filtering can be divided into two categories: memory-based and model-based [10,11]. The sudden growth in the number of available digital information and the number of visitors to an e-commerce system has created a potential challenge of information overload, which handles timely access to items of interest on the Internet or execution consuming a lot of time from the large-scale of data. This issue has increased the demand for recommendation systems more than ever being in the past few years [12]. The cold-start problem is a well-known issue in the recommendation system. This
issue is related to recommendations for new items or novel users. In the case of new users, the system
does not have information about their preferences. Then it is hard to make recommendations for
them [13]. Currently, many researchers created various methods to solve the cold-start problem, which
is the primary issue in the recommender system, and almost all of these methods gave a satisfying
result. But to increase more accuracy in predicting the item to match with the requirement of the user,
this issue still needs more research and development [14].

Besides developing the new techniques to enhance the accuracy in the recommendation system,
many researchers were developed new methods to provide a result with high accuracy in various
research areas, such as formal concept analysis (FCA), social network analysis, data mining, similarity
measuring method, parallel processing method, etc. Hence, therefore, Some of their important findings
and proposed methods are briefly described below.

FCA is a method of data analysis that describes the correlation among a particular set of attributes
and a specific set of objects. It gained the attention of the researchers because the results using this
method were completely satisfied. Constantinos et al. [15] proposed new theory using conceptual
dographs of FCA method to discovered knowledge in online drug transactions, Cole et al. [16] used FCA
method to analyze an email collection, Zhang et al. [17] combined the FCA method and the attribute
partial order theory to create a visualization method of huge video/image dataset, Chen et al. [18]
proposed the relationship between FCA theory and rough set theory, Miroslav et al. [19] proposed the
reduction of FCA method based on modified Hamming distance, Liping et al. [20] used FCA method to
create the test suite reduction for mutation testing, Kumar et al. [21] used FCA method for performance
evaluation of employees of an organization, Amin et al. [22] tried to mining hypermethylated genes in
breast cancer tumor subtypes using FCA method, Nourhene et al. [23] used FCA method to detect
the community in the social network with node attributes, Christian et al. [24] proposed symptoms
investigation using FCA method for enhancing medical diagnoses, Hao et al. [25] proposed similarity
evaluation between graphs using FCA method, etc. However, only a few research articles related to
the FCA method are introduced to creating the network graph.

Nowadays, the users in the social network are increasing rapidly. According to this reason,
many researchers are tried to develop various new methods such as how to classify users into several
communities by using their personal information or create a technique that can pick up the cluster of
the users by using their connection [17–21]. Some work that concerning this research were presented
by Hao et al. [26], his work shown the opportunities and insights when social computing meets soft
computing, and Ryu et al. [27] predicted the unemployment rate using social media analysis. One of
those methodologies is the k-clique, it is typically used to analyze data of the users in the vast social
network and classify them into various groups that depend on their personal information or their
connections [25–29]. In the past few years, some researchers used the k-clique method as an algorithm
model in the recommendation system [2]. The result of this idea presented a high accuracy.

Simultaneously, the data mining is a technique of analyzing hidden patterns and relationships in
data from vast amounts of data, and it automated using methods from statistics, machine learning,
pattern recognition and mathematics used in the recommendation system. The data mining method is
a trendy model in the field of the recommendation system. The technique of the data mining method,
such as a classification method, a clustering method and the prediction method. Those methods gave
a high-performance result. Moreover, another essential technology of the data mining method is an
association rule mining method, which is quite favorable for analyzing the frequency of the item in the
data transaction.

There are various techniques for measuring the similarity method among the user in the network
graph. These methods are widely discussed in many research articles. The commonly utilized
methods are the k-nearest neighbor and betweenness centrality method. The k-nearest neighbor
is a general algorithm that stores all existing cases and classifies new cases based on a similarity
measurement [30–33]. The betweenness centrality method is used to determine centrality in the
network graph based on close paths [34–36]. The result of using both of these methods is quite satisfied.
The doParallel mechanism is a parallel backend for the for–each package in the Parallel processing methodology. It provides a tool required to execute for–each loops in parallel processing. The for–each loops in parallel processing need to be used in conjunction with a container such as doParallel package instead of running code in parallel processing. The user needs to register a parallel backend to use. Differently, the for–each loop will perform tasks continuously, even when the %dopar% operator used. The doParallel uses multicore functionality on systems and snow component on Windows. Note that the multicore component only runs tasks on a single computer, not a cluster of machines. However, the snow function can use to run on a group using Unix operating systems such as Windows or even a combination. It is useless to use doParallel and parallel on a machine with a single processor with a single core. To improve the speed, it must work on devices with multiple processors, multiple cores or both. The for–each mechanism will issue a warning that it is working in sequence if there is no parallel backend registration. However, this warning will only be issued once.

Therefore, to increase the accuracy further compared to the recent results in the recommendation system, to solve the problem of cold-start and the construction of new technique, instead of using the standard technology in the recommendation system, the problem of execution consuming a lot of time to access to the large-scale dataset, the efficiency of a doParallel algorithm and FCA network graph applied to the recommendation system will be examined in this article.

The result of the implementation is satisfactory and proved that the proposed method could solve the cold-start problem in the recommender system and execution, consuming a lot of time in a large-scale dataset is solved.

The organizing of the article is:

• Section 1, this section gives a brief introduction of the methods, some of their related research, problem statement and contribution of this experiment. The methods used in this experiment are Recommendation System, formal concept analysis, social network analysis, data mining methodology, Similarity Measuring method, Parallel Processing Algorithm.

• Section 2, this section gives a brief explanation of the related methods applied to our experiment. Besides that, this section also provides depth with detail of the workflow of the proposed method.

• Section 3, this section displays the result of each process from the workflow of the proposed method. The detail of other methods mentioned in this article and their results from the experiment also introduced. Moreover, this section provided a brief explanation of the evaluation and the experimental findings from a comparison of the proposed method with the existing methods.

• Section 4, this section provides a discussion of the work from the author to the reader, the conclusion of the experiment and the future perspectives.

Problem Statement and Contribution

The recommendation techniques are widely used in e-business, for example, Netflix’s movie recommendation, Amazon’s book recommendation, Pendura’s music recommendations, Yahoo’s news recommendation, etc. One of the major challenges in the effectiveness of the recommendation system is how to recommend products to the new users, and the long time of the execution in the large-scale dataset. The cold-start problem occurs when new users enter the system, or new items are added to the catalog. For instance, if the data information of the existing users and the rating of the new user items are unavailable, then the new users must rate enough things so that the system can save their settings correctly and gives reliable recommendations to the new users accordingly. Similarly, when new items are added to the system, they must be ranked by many users before being able to recommend users with similar interests to those that get ranked. The cold-start problem can be solved in many ways, including asking the user at the beginning to rate some items, asking the users explicitly to state their taste in aggregate, and suggesting items to the user based on the collected demographic information. In the case of the scalability problem, the rate of growth of both the number of items and the number of users, it becomes difficult for a typical recommender to process such large-scale data. For example, Amazon.com recommends more than 18 million items to more than 20 million customers.
The scalability can be solved with different techniques, including clustering, reducing dimensionality and Bayesian network.

The main contribution of this article is summarized as follows. This article proposed the efficiency of a doParallel algorithm and the FCA network graph applied to the recommendation system. The main objective is to solve the cold-start problem, which is the primary issue in the recommendation system, and to solve the execution consuming a lot of time during access to the item from the massive scale of the dataset. The process of this article is to start using the personalized information of the users in the movie lenses dataset to create the FCA network graph. Then, the FCA network graph was divided into various communities by using the technique of the $k$-clique method. The personalized information of a new user applied to calculate the similarity with the user in the diverse community. This process is a combination of the $k$-nearest neighbor method and the betweenness centrality method to assign suitable communities to a new user. First, the $k$-nearest neighbor method is used for finding suitable communities for a new user. Second, the betweenness centrality method is used as the same purpose as the $k$-nearest neighbor method. Finally, the suitable communities, which generated by the $k$-nearest neighbor method and betweenness centrality method is gathered together and assigned to a new user. The association rule mining method as a mechanism of the data mining methodology is used for listing the name of the movies from appropriate communities. This list of the movies was recommended to a new user—the execution in this experiment used doParallel algorithm, which is a mechanism in parallel processing technology. This algorithm is used to split the proposed method algorithm into the various parts for the execution of them at the same time using the several cores of CPU. This idea can solve the problem of the execution consuming a lot of time to access the huge dataset.

To assess the performance. The proposed method is compared to the two most successful methods used in the experiment: the collaborative filtering method and the maximal clique method. The workflow of the maximal clique method is: it divided users into various communities by using users’ personalized information. Then, the new user will be assigned to a suitable community with the help of the combination method from the betweenness centrality method and the $k$-nearest neighbor method. The data mining method helps to find the movie from a suitable community. The resultant movies will be recommended to the new users. The collaborative filtering method follows the same as the process of the maximal clique method. Exceptionally in process two used the collaborative filtering method to divided users in the FCA network graph. Then, another process will handle the same method with implementation in the proposed method. Compared to the existing three methods, our proposed method gave more accurate and completely satisfying results.

2. Materials and Methods

2.1. Materials

This section gives a brief description of the definitions and how to apply the related methods to the experiment. The related methods are the formal concept analysis method, the $k$-clique method, $k$-nearest neighbor method, betweenness centrality method, doParallel mechanism and association rule mining.

2.1.1. Formal Concept Analysis (FCA) Network Graph

FCA is a theory of the data analysis that describes the relationship between a particular set of attributes and a specific set of objects. It generates two types of output from input data. The first is the conceptual lattice, which is a collection of formal concepts in data. Formal concept is arranged according to sub-concepts and relationships of super ideas. The second output is a set of attributes called attribute implications. An attribute implication is describing dependencies between attributes of objects, which are observed in consecutive points in times. \cite{15-25}.

The process to create an FCA network graph is, at first, the concept lattice is generated from the formal context. The attribute implications function or association rules function used to create a
relationship table that describes as the relationship among the user. Finally, this table used for creating a network graph.

2.1.2. The k-Clique Method

This is a technique in the social network analysis used for analyzing information in the sophisticated community structures from big social networks. Typically, k-clique is defined as a cluster of users that connected in big social networks graph. The k-clique technique picks up the communities from network graphs that have a value of k cliques. This community refers to sub-graphs of k users that exclusively connected. Figure 1 shows an example of the k-clique diagram. In which \( k = 3 \) is referring to a triangle or at \( k = 4 \) is referring to a rectangle and this rectangle consists of 3 triangles [28].

![Figure 1. Example of the graph with k-clique.](image)

In Figure 1, “a1, a2, a3, a4, a5, a6, a7 and a8” is a group of 8-cliques; “a2, b1, b2 and a 6” is a group of 4-cliques; “b1, b2 and b3” is a group of 3-cliques; “a6, b2, c1, c2 and a8” is a group of 5-cliques; “d1, a8 and a7” is a group of 3-cliques and “c2, a8 and d1” is a group of 3-cliques. Inside the graph, there are still various k-cliques. Inside the graph, there are have various k-cliques.

2.1.3. DoParallel Algorithm

It provides an excellent and efficient parallel programming platform for multi-processor/multi-core computers running operating systems. The doParallel algorithm package acts as an interface between the for–each and the parallel package of R2.14.0 and newer. It offers the tools needed to process a for–each loop at the same time. It is straightforward to install and very easy to use. Instantaneously, an average R programmer can start implementing parallel programs, without any previous experience in parallel computing. In this article, this algorithm is used to split the proposed method algorithm into various parts. Simultaneously begin their execution using the multi-core CPU. This strategy can overcome the massive time consumption during the implementation of an algorithm.

The step of the applied doParallel algorithm is:

- Step 1: Create a cluster of workers.
- Step 2: Register the cluster.
- Step 3: Processes the for-loop in parallel.
- Step 4: Stop the cluster.

2.1.4. Betweenness Centrality Method

It is a measure of the center point in the graph along the shortest paths for every pair of vertices in the connected graph. There is at least one shortest path between those vertices, regardless of the
number of road edges that passed the vertices. The betweenness centrality of each vertex is the number of shortest paths that pass through the vertex. It is a standard indicator of the center. It applies to a variety of problems in network methods, including issues related to social networks, scientific cooperation, transportation and biology [34–36]. The formula of betweenness centrality method of a node $v$ is described below:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

(1)

where $\sigma_{st}$ is the shortest paths number from node $s$ to node $t$ and $\sigma_{st}(v)$ is the value of those paths that pass through $v$.

2.1.5. k-Nearest Neighbor Algorithm

It is one of the most straightforward classification algorithms and widely used as learning algorithms. A $k$-nearest neighbor is a parameter-less learning algorithm. The objective of this algorithm is to use a database of data points separated into several classes to predict the classification of new sample points. The role of a $k$-nearest neighbor Algorithm is based on feature similarities. Figure 2 shows an example of how to classify a given data point by using the closely out-of-sample features that resemble the training set determines. It can be used for classification—the output is a class membership (predicts a class—a discrete value). The objects were classified according to the majority votes of the neighboring to assign general categories among the $k$-nearest neighbor method. It can also use for regression—output is the value for the object (constant value prediction). This value is the median of the importance of its $k$-nearest neighbor method.

![Figure 2. Finding the cluster.](image)

2.1.6. Dataset

The dataset used in this experiment is a movie lenses dataset [37]. It was divided into two parts, part one is a training dataset with include 80% of the users, and part two is the test dataset, which includes 20% of the users. There are 100,000 ratings from 943 users of 1684 movies. Simple demographic data for users included age, gender and occupation—details of the data set are given below.

The users in the test dataset are a random list of users. The number of users in the test dataset is 143, each user data consisting of three types of personal data: age, gender and occupation. The users in the experimental dataset are the remaining users after random selection from the movie lenses dataset. The number of users in the innovative dataset is 800 users.
2.1.7. Mean Absolute Percentage Error (MAPE) Method

MAPE is widely used in the field of statistics structure as predicting the accuracy of the predictive way; the formula of MAPE is shown below [38]. The execution of the experimental part of this article is completed. A list of the recommended movies generated from the experiment and the list of the movies rated by users from the movie lenses dataset is used for prediction accuracy. The mean absolute percentage error (MAPE) is a method used to estimate the output of the proposed method.

\[
\text{MAPE} = \frac{100}{t} \sum_{n} \left| \frac{H_n - G_n}{A_n} \right|
\]

(2)

"G_n" is the variable of predicting a value and "H_n" is the variable of real value.

2.2. Design of the Proposed Method

This section describes in detail the scenario of the proposed method. The main idea of this article is to solve the cold-start problem and solve the problem of execution that consumes a lot of time on a large-scale dataset. The cold-start issue concerns the personalized recommendations for users with no or few histories (new users). Hence, providing advice to users with little history becomes a difficult problem for collaborative filtering models because their learning and predictive ability are limited. According to the ordinary idea of the user who has the same feature information or attribute or personalized information may like the same items. Figure 3 shows the detail of the proposed method. Process 1 used the personalized information of the user in the movie lenses dataset, which divides into two parts as the training part and testing part. The information in the training data used to create the table of a formal context. After the formal context generated, then the Lattice Miner 2.0 application was used to create the concept lattice. The concept lattice presents the line diagram and obtainable by collecting all objects from the concept node to the lattice infimum, and its intent is available by visiting all attributes from the concept node to the lattice supremum. Later, the FCA network graph will create via the help of the association rule function in the Lattice Miner 2.0 that use for checking the relationship between objects (users) in the concept lattice.

In process 2, After the FCA network graph was generated from process 1. The \(k\)-clique method, which is the potential mechanism of the social network analysis for detecting the community, is used to identify and divide the users in the FCA network graph into various communities. Later, the data were stored in the dataset.

In process 3, once a new user joins the system, their personalized information used to find suitable communities by the comparison of the similarity of the new user to the communities stored in the dataset and user's personalized information from the movie lenses dataset. In this step:
In process 2, after the FCA network graph was generated from process 1. The \(k\)-clique method, which is the potential mechanism of the social network analysis for detecting the community, is used to identify and divide the users in the FCA network graph into various communities. Later, the data were stored in the dataset.

In process 3, once a new user joins the system, their personalized information used to find suitable communities by the comparison of the similarity of the new user to the communities stored in the dataset and user’s personalized information from the movie lenses dataset. In this step:

- First, the \(k\)-nearest neighbor method was employed to find the matching community for the new users, as shown in Figure 4a.
- Second, the betweenness centrality method finds the matching community for the new users, as shown in Figure 4b.
- Finally, the combination of the \(k\)-nearest neighbor method and the betweenness centrality method was used to assign suitable communities to the new users, as shown in Figure 4c.

![Figure 4. Finding the appropriate community for the new user.](image_url)

In process 4, when the new user knew their suitable communities. The data from the movie lenses dataset and data from suitable communities is used to create a list of recommended movies. In this step, the association rule mining method is one of the various powerful mechanisms in the data mining method. This mechanism is used to calculate and find the movie from the suitable communities of the new users. Finally, the list of recommended movies from the previous process was recommended to the new users.

In this article, the below algorithm is used to split the proposed method algorithm into the various parts, then executed them at the same time in the multiple cores of CPU. This idea can solve the problem of the execution of consuming a lot of time.

The algorithm used to detect the community using the \(k\)-clique method based on doParallel processing is [Algorithm 1].

**Algorithm 1.** Detects the community using a \(k\)-clique algorithm based on doParallel processing

```
Algorithm Detects the community
Input: FCA graph
Output: A list of community
Step 1: initialize TD, Comu
Step 2: TD ← FCA graph
Step 3: cl ← makeCluster(5)
Step 4: registerDoParallel(cl)
Step 5: ls ← foreach(icount(iters)) %dopar% {
   Step 6: to.ls ← kCliqueFun()
   Step 7: to.ls ← summary(to.ls)
Step 8: stopCluster(cl)
Step 9: Comu ← ls
```
Algorithm 1 shows the procedure of detection of the community from the FCA network graph based on using the k-clique method and doParallel processing. The procedure begins by importing the FCA network graph, and consequently, the list of the communities get generated. The procedure in detail is; Step 1, initially the declaration of variable “TD” and “Comu”. Step 2, the value from the FCA graph assigned to variable “TD”. Step 3, the library package of parallel processing was installed and then assigned the number of a worker to a cluster (“c1”). Step 4, registration of a cluster to the doParallel processing. Step 5, initialization of a for–each loop based on operator “%dopar” to variable “to.ls”. Step 6, called the k-clique function and its execution, then assigned the result to variable “ls”. Step 7, summary the result of “to.ls” to new “to.ls”. Step 8, after the cluster gets completed the execution, then we need to stop this execution. Eventually, step 9, assigned the value from “ls” to “Comu”.

3. Results

This section shows a brief definition and how to apply the related methods in this article. The results obtained from the experiment of each method through the proposed method workflow from process 1 to process 5 are described below.

3.1. Formal Context Result

This part presents the step to create a formal context and its result from the experiment. According to a brief description in the previous section, a formal context is composed of the set from objects and attributes that describes a domain in the ideal case. A table can represent a formal context with rows and columns corresponding to objects and attributes, and the character “X” describes the relationship between object and attribute.

In this experiment, the user’s personalized information from movie lenses dataset was used to create a formal context data. In a formal context, the data column refers to the user’s attributes, such as age, gender and occupation. The data row refers to the user’s ID. The detail of the formal context is shown in Table 1. The character “A1 to A30” applies to the user’s personalized information. For example, A1 refers to the value of age less than 20 years old, A2 refers to the value of age less than 30 years old, A3 refers to the value of age less than 40 years old, A8 refers to the value of maturity higher than 80 years old, A9 refers to gender as female, A10 refers to gender as male, A11 refers to occupation as admin and A30 refers to rule as other jobs. The character “obj 1 to obj 800” refers to the user id in the training dataset. The character “X” refers to the relationship between object and attribute.

| Object | A1 | A2 | A3 | ... | A8 | A9 | A10 | A11 | A12 | A13 | ... | A30 |
|--------|----|----|----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|
| Object 1 | X  |    |    |     |    |    | X   |     |     |     |     |     |     |
| Object 2 |    | X  | X  |     | X  |     |     |     |     |     |     |     |     |
| Object 3 | X  |    |    |     |    |    |     | X   |     |     |     |     |     |
| Object 4 |    | X  |    |     |    |    | X   |     |     |     |     |     |     |
| Object 5 | X  |    |    |     |    |    |     |     | X   |     |     |     |     |
| ...    |    |    |    |     |    |    |     |     |     | X   | X   |     |     |
| Object 800 | X  |    |    |     |    |    |     |     |     |     |     |     |     |

3.2. Concept Lattice Result

This part presents the step to create the concept lattice and its result from the experiment. Referring to a brief description in the previous section, concept lattice presents as a line diagram that originated from the formal context. The objects are shown inside as blue nodes drawn below some attribute nodes, and the attributes inside gray nodes drawn above some object nodes. The nodes distributed a way that the extent of a concept. It is obtainable by collecting all objects from the concept node to the lattice infimum, and it is intent available by visiting all attributes from the concept node to the lattice
supremum. The labeling can also be explained utilizing the notions of object concept and attribute concept. Figure 5 shows the concept lattice created from a formal context, as described in Table 1.

3.3. FCA Network Graph Result

The previous section was given a brief explanation of the step to create the FCA network graph. This section will present on how to create the FCA network graph. Once the concept lattice is generated, then the attribute implications function or association rules with the best support as a tool in Lattice Minor 2.0 Application is used to create a relationship table among the user. Later on, this table was used for creating a network graph. The result of the FCA network graph is shown in Figure 6. This graph also shows the relationship between the user after import from the concept lattice.

3.4. Communities Result

This part shows community detection from the FCA network graph and its results from the experiment. Once the FCA network graph was created, the k-clique method helps to classify the users in the FCA network graph into the various communities, as shown in Table 2. The character “k” refers
3.4. Communities Result

This part shows community detection from the FCA network graph and its results from the experiment. Once the FCA network graph was created, the k-clique method helps to classify the users in the FCA network graph into the various communities, as shown in Table 2. The character “k” refers to the number of users in each community. The number of users “k” in the community starts from 3 to 8 when the value of “k” is higher than 8; the community is not found. In Table 2, when the number of users in the community equals 3, then the amount of the community is 176. When a number of the users equals 4, then the number of the community is 135. The number of users equals 5, then the value of the community is 110. The number of users equals 6, then the amount of community is 101. Finally, when the value of the user equals 7, then the number of the community is 83, and the number of the community is 75, then the number of users in the community is equal to 8.

| #user | Number of Community |
|-------|---------------------|
| k = 3 | 176                 |
| k = 4 | 135                 |
| k = 5 | 110                 |
| k = 6 | 101                 |
| k = 7 | 83                  |
| k = 8 | 75                  |

3.5. Suitable Communities Result

This section shows the matching of a suitable community to a new user and its result from the experiment. The result of the suitable community to the new users are shown in Tables 3–5. Table 3 presents a suitable community for a new user with the help of the k-nearest neighbor method. The character “C” refers to the community, and the number display at the back of character “C” is the ID of the community. The role “u1 to u143” refers to the user id of the new user login. The new user can assign to the various communities if their personalized information similarity to the user in various groups.

| #user | Suitable Community    |
|-------|-----------------------|
| u1    | C1, C2, C5, C8        |
| u2    | C56, C100             |
| u3    | C80                   |
| u4    | C50, C60, C76         |
| u5    | C23, C56              |
| u6    | C123, C130, C145, C170|
| …    | …                     |
| u143  | C100, C102            |

Table 4. Suitable communities from the betweenness centrality method.

| #user | Suitable Community    |
|-------|-----------------------|
| u1    | C123, C145            |
| u2    | C56, C100, C150, C164 |
| u3    | C80, C90, C123        |
| u4    | C50, C60, C76         |
| u5    | C23, C56              |
| u6    | C123, C130, C145, C170|
| …    | …                     |
| u143  | C100, C102, C150, C175|
Table 5. Suitable communities from the kNN betweenness centrality method.

| #user | Suitable Community          |
|-------|-----------------------------|
| u1    | C1, C2, C5, C8, C123, C145  |
| u2    | C56, C100, C150, C164       |
| u3    | C80, C90, C123              |
| u4    | C23, C50, C56, C60, C76     |
| u5    | C23, C56                    |
| u6    | C123, C130, C145, C170      |
| ...   | ...                         |
| u143  | C100, C102, C150, C175      |

Table 4 shows the result of the suitable community to the new user with the help of the betweenness centrality method. The meaning of the character “C” and “u1 to u143” refer to the same purpose, as explained above.

Table 5 shows the result of the suitable community to the new user with a combination of a suitable community with the help of the k-nearest neighbor method and a suitable community with the help of the betweenness centrality method.

3.6. List of the Movie Result

This part shows the movie list generation from a suitable community, which is assigned to a new user as a result of the experiment. After the new user was assigned to suitable communities. Then all the movies may become recommended movies to the new user, as shown in Table 6.

Table 6. List of the movies from suitable communities.

| #user | Movie in the Community           |
|-------|----------------------------------|
| u1    | M1, M2, M5, M80, M100, M200, M1500, ..., M1600 |
| u2    | M1, M23, M35, M50, M150, M400, M1300, ..., M1230 |
| u3    | M11, M23, M511, M830, M1200, M240, M1500, ..., M1550 |
| u4    | M211, M232, M675, M830, M1200, M1100, M11500, ..., M1630 |
| u5    | M321, M434, M531, M830, M1100, M1200, M1301, ..., M1320 |
| u6    | M14, M22, M53, M820, M1300, M1330, M1420, ..., M1450 |
| ...   | ...                             |
| u81   | M1, M23, M35, M50, M150, M400, M1300, ..., M1230 |
| u82   | M11, M23, M511, M830, M1200, M240, M1500, ..., M1550 |
| u83   | M211, M232, M675, M830, M1200, M1100, M11500, ..., M1630 |
| ...   | ...                             |
| u143  | M241, M242, M532, M840, M1100, M1200, M1501, ..., M1550 |

Table 6 display the ID of the movie, which may be later recommended to the new user. The character “M” and the number at the end of it refer to the movie id. The amount of the recommended movie depends on the movie in a suitable community, and also rely on the movie watched by the user in that community.

3.7. Recommended Movie Result

This part presents the movie list generation, which may become the recommended movie to a new user, and its result from the experiment. In this section, the association rule mining as the mechanism in data mining is used to find the proper movie to recommend to the user. Table 7 shows the results of this process. The character “M” and its number are referred to as the id of the movie and the character “u1 to u143” are refers to the ID of the new user.
Table 7. List of the recommended movies.

| #user | Movie ID                      |
|-------|-------------------------------|
| u1    | M1, M2, M5, M80, M100       |
| u2    | M1, M23, M35, M50, M150     |
| u3    | M11, M23, M511, M830, M1200 |
| u4    | M211, M232, M675, M830, M1200|
| u5    | M321, M434, M531, M830, M1100|
| u6    | M14, M22, M53, M820, M1300  |
| ...   | ...                           |
| u47   | M211, M232, M675, M830, M1200|
| u48   | M321, M434, M531, M830, M1100|
| u49   | M14, M22, M53, M820, M1300  |
| u50   | M321, M434, M531, M830, M1100|
| u51   | M14, M22, M53, M820, M1300  |
| u52   | M321, M434, M531, M830, M1100|
| u54   | M31, M34, M531, M80, M1100  |
| ...   | ...                           |
| u143  | M241, M242, M532, M840       |

3.8. Evaluation of the Proposed Method Result

This section shows the evaluation result of all the methods used in the experiment, such as the collaborative filtering method, maximal clique method and the proposed method. For example, in Table 7, “u1” is an ID of a 35 years old male new user, and he is an engineer by profession. Alternately, “u80” is an ID of a 24 years old female new user, and she is a student. The system recommended movie IDs to “u1” as m24, m242, m532, m821 and m1222, which refers to Toy Story (1995), Twelve Monkeys (1995), Dead Man Walking (1995), Antonia’s Line (1995) and From Dusk Till Dawn (1996) movies, respectively. The mean absolute percentage error used for evaluating the output of the experiment, as shown in Figures 7 and 8.

![Figure 7. Mean absolute percentage error (MAPE) results using the proposed method.](image-url)
Figure 7 shows the result of the proposed method, which displays various values depending on the value of k. The value of k is the number of users in each community. In essence, the community was detected from the FCA network graph with the help of the social network analysis methodology, such as the k-clique method. The number of users in each community begins from the 3 people until the maximum number of users in the community, which can be detected by the k-clique method. In this experiment, the maximum number of users in the community detected by the k-clique method is equal to 8 people. For the detail of the MAPE result in this proposed method is; when the amounts of users in the community are equal to 3 people, this means the result of the prediction or incorrect value or the MAPE result value will be 16.71%. When we checked the number of users in the community equals to 4 people, we found that the result value of the prediction or MAPE value is 15.21%. After that, when the amounts of users in the community are equal to 5 people, the result of the prediction or incorrect value or the MAPE result value becomes 15.15%. Later on, we found that the result value of the prediction or inaccurate value is equal to 13.87% when the number of a user is 6 users. However, if the number of users in the community equals 7 users, then the result value of the prediction or MAPE value is equal to 12.55%. Finally, when the maximum number of the user in the community was detected at the 8 people, we found that the value of prediction is equal to 10.27%. Figure 7 shows that the best outcome of the proposed method is 10.27%, while the value of k = 8.

The MAPE’s results of the collaborative filtering method, and maximal cliques method used to evaluate the preciseness of the proposed method. If the MAPE output value result becomes a low value, that means our proposed method is useful. First of all, the mean absolute percentage error for the proposed method will calculate. In brief, The result of the proposed method is displayed in various values, which depend on the value of k. The value of k is a number of users in each community. The result value of this proposed method is 10.27%, for k = 8 people. Once completion of the calculation of the result value of the proposed method. Comparatively, the result value of the collaborative filtering method is used to calculate, and the result was 18.88%. The workflow and calculates done for MAPE result of the collaborative filtering method was using the same way to the experimental of the proposed method. Eventually, the result value of the maximal clique approach was calculated as 27.55%. The experimental approach of this method was similar to the maximal clique method.

Figure 8 shows a comparison of all the methods used in this experiment. The best outcome of the MAPE is 10.27% from the proposed method, followed by the experiment of using collaborative filtering with the MAPE is 18.88%, and the last satisfaction outcome belongs from the MAPE using maximal clique method. As a matter of clarification, the workflow of the collaborative filtering method...
in this implement is; the collaborative filtering classified users into several groups by using user personal information. After that, the new user will be assigned to a suitable group with the help of the combination method from the k-nearest neighbor method and betweenness centrality method. Later on, the data mining method helps to find the movies from an appropriate group. These movies will become the movie recommend to the new user. In the case of the maximal clique method, this method follows the process in Figure 3. Exceptionally in process two, this process used the maximal clique method to classify users in the FCA network graph. Then, for another operation such as process three, four and five were handled by the same method with implementation in the proposed method.

3.9. Speedup of the Execution Consuming Time from Large-Scale Dataset Result

This section shows the result of the execution time consuming while the performance of the algorithm using the sequential and doParallel mechanism of all the methods implemented in this article. The environment material of the implementation in this article is Windows 10 Pro 64-bit for Operating System, Intel® Core™ i5 CPU for a processor and 24 GB for the RAM. First, all method used in this experimental is execution used the sequential algorithm. Second, all method used in this innovative is performance used the doParallel algorithm. In essence, 5 cluster workers were registered, then %dopar% is used for running of the doParallel algorithm of the method implemented in this article. The comparison result of using a sequential and doParallel algorithm showed in Table 8.

| Method          | Sequential | doParallel |
|-----------------|------------|------------|
| Proposed method | 48 h       | 3 h        |

Table 8 shows the comparison result of using a sequential and doParallel algorithm. The sequential algorithm of the proposed take 48 h to complete the execution, and the result of the time-consuming of using doParallel algorithm of the proposed method take 3 h to achieve the performance. The comparison shows that the outcome using doParallel algorithm is completely satisfied such as the timely execution of the proposed method is decreased from 48 h to 3 h.

4. Conclusions

4.1. Discussion

Since the recommendation system is essential in the business over the Internet, online marketing is providing a lot of benefits to the business owner for providing the movies to the consumer through the Internet or online entertainment. The number of movies has increased to become more overcrowded; finding a video that users are looking for through the existing technologies is very hard. For this reason, users need a system that can recommend their movie needs and the best technology about these; it is a recommendation system. However, the collaborative filtering method is a powerful method using in the recommendation system to predict user needs, as this method helps to satisfy their predictions. Inside the collaborative filtering mechanisms still have some issues like a cold-start problem, which is related to the new user login without having any prior Internet searching information or without their history issue before.

Today, many researchers are interested in developing various methods to improve accuracy, instead of collaborative filtering methodology. Therefore, to improve the accuracy in the system, we present the FCA network graph created from the personalized information of the user in the system to solve the issue of cold-start problems and offer the doParallel algorithm to solve the problem of execution consuming a long time to access to the vast dataset.

In this manuscript, we proposed the efficiency of a doParallel algorithm and the FCA network graph applied to the recommendation system, which has shown the best accuracy of the movie
recommendation system. However, to evaluate the performance, collaborative filtering methods, the maximal clique methods, and the proposed methods are used to assess the movie lenses dataset.

The experiment result of MAPE in Figure 8 shows the proposed method provides the result of MAPE with more accuracy than the existing method, the accuracy result from the proposed method is 89.73%, from the collaborative filtering method is 81.12% and from the maximal clique method is 72.45%. These results confirm that the proposed method is powerful and useful in the area of recommendation systems. The proposed method has shown that it can solve the cold-start problem and timely execution issue as well. The comparison shows that the appropriate performance of a collaborative filtering algorithm decreased from 24 h to 1 h, the maximal clique algorithm reduces from 29 h to 1 h and 20 min and the proposed method decreased from 48 h to 3 h.

4.2. Future Research Direction

Since this article only covers the movie lens dataset. The data were collected by the GroupLens Research Project at the University of Minnesota. This dataset consists of 100,000 ratings from 943 users on 1682 movies; each user has rated at least 20 movies. Therefore, they may not provide full of valuable information to the recommendation system, but movie lens 1-M dataset, 10-M dataset, 20-M dataset and other kinds of the dataset should be considered. Another challenge from this article, is that researchers need to consider modification of the algorithm of the proposed method, which is complicated and difficult to apply in real world situations. Ultimately, this proposed method may pave the way for future opportunities.

Author Contributions: Conceptualization, P.V. and D.-S.P.; methodology, P.V.; writing—original draft preparation, D.-S.P.; writing—review and editing, P.V.; visualization, D.-S.P.; supervision. All authors have read and agreed to the manuscript published version.

Funding: This research was funded by Soonchunhyang University and the National Research Foundation of Korea (No. NRF-2020R1A2B5B01002134).

Acknowledgments: The authors would like to thank the Soonchunhyang University and the National Research Foundation of Korea for financially supporting this research under contract No. NRF-2020R1A2B5B01002134.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jeong, W.H.; Kim, S.J.; Park, D.S.; Kwak, J. Performance improvement of movie recommendation system based on personal propensity and secure collaborative filtering. J. Inf. Process. Syst. 2017, 13, 157–172.
2. Vilakone, P.; Park, D.S.; Xinchang, K.; Hao, F. An efficient movie recommendation algorithm based on improved k-clique. Hum. -Cent. Comput. Inf. Sci. 2018, 8, 1–15. [CrossRef]
3. Xingchang, K.; Vilakone, P.; Park, D.S. Movie recommendation algorithm using social network analysis to alleviate the cold-start problem. J. Inf. Process. Syst. 2019, 15, 616–631.
4. Isinkaye, F.O.; Folajimi, Y.O.; Ojokoh, B.A. Recommendation systems: Principles, methods, and evaluation. Egypt. Inform. J. 2015, 16, 261–273. [CrossRef]
5. Vilakone, P.; Xinchang, K.; Park, D.S. Personalized movie recommendation system combining data mining with the k-clique method. J. Inf. Process. Syst. 2019, 15, 1141–1155.
6. Hao, F.; Park, D.S.; Yin, X.Y.; Wang, X.M.; Vilakone, P. A location-sensitive over-the-counter medicines recommender based on tensor decomposition. J. Supercomput. 2019, 75, 1953–1970. [CrossRef]
7. Hao, F.; Pei, Z.; Park, D.S.; Vilakone, P.; Seo, H.S. Mobile cloud services recommendation: A soft set-based approach. J. Ambient Intell. Humiz. Comput. 2018, 9, 1235–1243. [CrossRef]
8. Konstan, J.A.; Riedl, J. Recommender systems: From algorithms to user experience. User Model User-Adapt Interact 2012, 22, 101–123. [CrossRef]
9. Pan, C.; Li, W. Research paper recommendation with topic analysis. Comput. Des. Appl. IEEE 2010, 4, V4-264–V4-268.
10. 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’11), Chicago, IL, USA, 23–27 October 2011; ACM: New York, NY, USA, 2011; pp. 157–164.

11. Viana, P.; Pinto, J.P. A collaborative approach for semantic time-based video annotation using gamification. Hum. -Cent. Comput. Inf. Sci. 2017, 7, 13. [CrossRef]

12. Pazzani, M.J. A framework for collaborative, content-based and demographic filtering. Artif. Int. Rev. 1999, 13, 393–408. [CrossRef]

13. Bojnordi, E.; Moradi, P. A novel collaborative filtering model based on the combination of correlation method with matrix completion technique. In Proceedings of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP), IEEE, Shiraz, Fars, Iran, 2–3 May 2012.

14. Blerina, L.; Kostas, K.; Statthes, H. Facing the cold start problem in recommender systems. Expert Syst. Appl. 2014, 44, 2065–2073.

15. Constantinou, O.; Babak, A.; Petra, S.B. Discovering Knowledge in Online Drug Transactions Using Conceptual Graphs and Formal Concept Analysis. In Proceedings of the European Intelligence and Security Informatics Conference (EISIC), Uppsala, Sweden, 17–19 August 2016.

16. Cole, R.; Eklund, P. Analyzing an Email Collection Using Formal Concept Analysis. Princ. Data Min. Knowl. Discovery. Pkdd 1999, 1704, 309–315.

17. Zhang, Y.Y.; Wang, X.M.; Wang, J.K.; Liu, J.B. A visualization method of massive video/image databases based on formal concept analysis and the attribute partial order theory. In Proceedings of the 6th International Congress on Image and Signal Processing (CISP), Hangzhou, China, 16–18 December 2013; p. 1.

18. Chen, X.Y.; Li, G.Y. The Relation between Formal Concept Analysis and Rough Set Theory. In Proceedings of the International Conference on Computational Intelligence and Communication Networks, Bhopal, India, 14–16 November 2014.

19. Miroslav, S.; Peter, B.K. Formal Concept Analysis Reduction Method Based on Modified Hamming Distance. In Proceedings of the World Symposium on Digital Intelligence for Systems and Machines (DISA), Kosice, Slovakia, 23–25 August 2018.

20. Li, L.P.; Gao, H.H. Test suite reduction for mutation testing based on formal concept analysis. In Proceedings of the IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD), Takamatsu, Japan, 1–3 June 2015.

21. Kumar, C.A.; Sumangali, K. Performance evaluation of employees of an organization using formal concept analysis. In Proceedings of the International Conference on Pattern Recognition, Informatics, and Medical Engineering (PRIME-2012), Salem, Tamilnadu, India, 21–23 March 2012.

22. Islam, I.A.; Samar, K.K.; Aboul, E.H.; Hesham, A.H. Formal concept analysis for mining hypermethylated genes in breast cancer tumor subtypes. In Proceedings of the 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, India, 27–29 November 2012.

23. Ryu, P.M. Predicting the unemployment rate using social media analysis. J. Inf. Process. Syst. 2018, 14, 904–915.

24. Hao, F.; Sim, D.S.; Park, D.S.; Seo, H.S. Similarity evolution between graphs: A formal concept analysis approach. J. Inf. Process. Syst. 2017, 13, 1158–1167.

25. Hao, F.; Park, D.S.; Pei, Z. When social computing meets soft computing: Opportunities and insights. Hum. -Cent. Comput. Inf. Sci. 2018, 8, 8. [CrossRef]

26. Christian, S.; Diana, S.; Diana, T. Symptoms investigation by means of formal concept analysis for enhancing medical diagnoses. In Proceedings of the 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 21–23 September 2017.
31. Okfalisa; Ikbal, G.; Mustakim.; Reza, N.G.I. Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. In Proceedings of the 2nd International conferences on Information Technology, Information Systems, and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 1–2 November 2017.

32. Kholid, F.; Moh, Z.; Anita, S.N.; Ira, P. Application of K-Nearest Neighbor Algorithm For Puzzle Game of Human Body’s System Learning on Virtual Mannequin. In Proceedings of the International Conference on Applied Science and Technology (iCAST), Manado, Indonesia, 26–27 October 2018.

33. Faza, G.; Maman, A.; Aji, G.P. Prediction of Smartphone Charging using K-Nearest Neighbor Machine Learning. In Proceedings of the Third International Conference on Informatics and Computing (ICIC), Palembang, Indonesia, 17–18 October 2018.

34. Freeman, L. A set of measures of centrality based on betweenness. *Sociometry* **1977**, *40*, 35–41. [CrossRef]

35. Mantrach, A. The sum-over-paths covariance kernel: A novel covariance measure between nodes of a directed graph. *Ieee Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1112–1126. [CrossRef]

36. Dolev, S.; Elovici, Y.; Puzis, R. Routing betweenness centrality. *J. Acm.* **2010**, *57*, 25:1–25:27. [CrossRef]

37. Harper, F.M.; Joseph, A.K. The MovieLens Datasets: History and Context. *Acm Trans. Interact. Intell. Syst.* **2015**, *5*, 4. [CrossRef]

38. Kim, S.; Kim, H.Y. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **2016**, *32*, 669–679. [CrossRef]