Journal of Theoretical and Applied Information Technology
15th May 2018. Vol.96. No 09

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# AN UNDERSTANDING AND APPROACH SOLUTION FOR COLD START PROBLEM ASSOCIATED WITH RECOMMENDER SYSTEM: A LITERATURE REVIEW

**[1,2]HANAFI, [3]NANNA SURYANA, [4]ABD SAMAD BIN HASAN BASARI.**

[1]Department of Information Technology, University of Amikom Yogyakarta, Yogyakarta, Indonesia
[2,3,4]Faculty of Information and Communication Technology, University Teknikal Malaysia Melaka, Melaka, Malaysia.
Emails: [1]hanafi@amikom.ac.id, hanafiutem@gmail.com, [3]nsuryana@utem.edu.my, [4]abdsamad@utem.edu.my

**ABSTRACT**

In the ecommerce services, there is a very important tool that will determine the success to increase number of buying and selling in the marketing target, that is how the user in finding products that are suitable to be purchased. This tool is called recommender system. Recommender system is important tool for establishing an effective communication between users and retailers in ecommerce business. Effective and enjoyable communication to find the product is considered to have a significant impact that increase of marketing achievement. Recommender system established in the mid-90s. Based on technical approach, there are four types of recommender system namely Collaborative, Contents Based, Knowledge Based and Demographic filtering. Collaborative filtering is considered to be more superior than another two methods. It offers obviously advantages in terms of serendipity, novelty and accuracy. Although it has some benefit. However, there is a critical problem in collaborative filtering that called cold start. It is a major problem to which many researchers have paid much more attention to this particular research interest. In response to this particular problem the critical review and analysis on state of the art of the current technology, some possible solutions including approach, method and techniques used have been identified but they need further validation.

**Keywords -** *Recommender System, Recommendation System, Sparsity Data, Cold Start, Ecommerce Services.*

## 1. INTRODUCTION

Established in the beginning 20th century, the growth of Internet users has increased significantly; this implies that influent e-commerce company provides millions of items for millions of consumers at any time and any places. Choosing among in too big options is very difficult to be done for users. For example, what product needs to buy, what news to read, what movie should be watched, what music to listen, what advertising should be looked etc.

The growth of internet network is so vast and fast. In other words, it is influenced significantly by the number of internet users, It has led to the growth of so many social media applications. As one devastating consequences is the over flooded by unnecessary information. This is why the recommender system is very important in our living right now. The critical review in this research, has been conducted based on most recent related work. There is a need to understand about the state of the art of recommender system including proposed a way forward.

Refer [1] n the real business world and academic study, there are several motivations why we must needed a recommender system for e-commerce business:

- Increase quantity of products/items sell.
- Increase users/customers Fidelity.
- Increase the user service satisfaction.
- Sell more diverse items/product.
- Better understanding of what the users/customers want.

Many e-commerce business portals have employed recommender system machine to support business, for example Netflix, Amazon, YouTube, Facebook, Google, MovieLens, Last.fm, Alibaba, eBay and etc. According to reference [2][3] There are four primary technical goals of recommender system to implement in the system retailer namely;

a. **Relevant**: the frequent target of a recommender system is to serve information about items that suitable to users' interest. So, the product can easy to find by customers. It needed by users or customer to help them get especially product for online shopping.
b. **Novelty**: Recommender systems are very powerful to help a user find good products or items. The other important things the goals of recommender systems are to serve information

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

that has not seen in the past before. Repeated serving of popular information probably reduction in marketing target.

c. **Serendipity**: it means products recommended are unpredictable. The systems can serve better information, so users feel lucky to get the information. There no clear definition and no consensus to explain serendipity, but serendipity must have included three components are novel, relevant and unexpected.
d. **Diversity:** Or increasing recommendation diversity. It means diversity makes sure to users that the information shown is not repeated, so users are not bored.

### 1.1. Objective Research

Cold start is the number one problem in domain recommender system, this is very important problem should be addressing aim to determine accurate recommendation system. Refer to author [4] they have study review in hybrid recommender system perspective and found, in fact many researcher conduct study to develop hybridization aim to get way out from cold start problem. Reference by author [5] in this survey research who focus in mobile multimedia recommendation perspective, Refer Author in [6] in 2002 have conduct survey study that focus in strategy to engineering hybridization, he have proposed hybridization strategy that famous called Burk taxonomy hybridization. Refer Author in [7] interested to survey study in social media recommender system perspective, as we know social behavior influenced in many decision for recommendation, this study also influenced trend of social media application in our living. By author [8] they develop literature review that focus comprehensive classification technique to develop ecommerce recommender system, another author [9] interested to survey study about collaborative filtering point of view. Different angle of survey study have done by [10] they concern in recommender system based on mobile internet application point of view. A technical survey by [11] [12] that focus in engineering technical based on collaborative filtering. According by summarize literature review study by some researchers on above, there is no author concern on study about cold start, its important study by author opinion about how to technical approach, what important element to tacking this problem and what have been done and also evaluate for them also what the future challenge. Author [13] very concern do critique recommender system based deep learning technology.

### 1.2. Basic Method Recommender System

The basic idea of recommender system is recommending items by analyzing and learning user's profile, user's previous activities, and the kind of items available in the system. Many methods to approaches used by the e-commerce platform to recommend products to users' candidate. It has been identified that most essential and major strategies for recommender systems are content based, collaborative filtering, knowledge based and demographic filtering. Many practical industry and study also have been conducted as an attempt to search the possibility to mix between one to another called as hybrid methods. The most powerful and popular approach and widely implemented that named Collaborative filtering is the most efficient recommender systems [14]. Implemented recommender system was increase marketing target [15][16][17].

According reference [6] there are three algorithm models in which the development and implementation in real business world and research area, include:

a. **Content based.**
The system tries to learn to recommend items or product that is similar to the ones that the user liked in the past. The similarity of items will be calculated using the features contents related to the compared items. For example, if a user has rated a film that belongs to the comedy genre, then the system can learn to recommend other movies from this genre. Older Content Based recommendation procedures go for coordinating the properties of the client profile against the properties of the items. Much of the time, the properties of the item are essentially watchwords that are removed from the items portrayals. Semantic ordering methods speak to the item and user's profiles utilizing ideas rather than keyword.
b. **Collaborative Filtering.**
The first and most straightforward execution of this approach makes suggestions to the dynamic user based on item that different user with comparable tastes like before. The comparability in taste of two users is compute in view of the likeness in the rating history of the users. This is the motivation behind why alludes collaborative filtering as "individual to individual connection." Collaborative filtering is thought to be the most famous and generally executed method in recommender systems.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**c. Demographic Filtering**

Demographic Filtering uses demographic data such as age, gender, education, etc. for identifying categories of users. It does not suffer from the new user problem as is doesn't use ratings to provide recommendations. However, it is difficult today to collect enough demographic information that is needed because of online privacy concerns, limiting the utilization of demographic filtering. It is still combined with other recommenders as a reinforcing technique for better quality

**d. Knowledge Based**

Knowledge based uses knowledge about users and items to reason about what items meet the users' requirements and generate recommendations accordingly. A special type of knowledge based are constraint-based recommender system which are capable to recommend complex items that are rarely bought (i.e. cars or houses) and manifest important constrains for the user (price). It is not possible to successfully use collaborative filtering or content based in this domain of items as few user-system interaction data are available (people rarely buy houses)

### 1.3. Basic Type And Rating Determination

One of important thing to generate recommendation are user's activity for example user purchasing, and user clicks to product, items that were seen by user, sometimes, many e-commerce portal support available opinion comment for his/her product. Product interested by users include rating. Rating is representation of interest feel by users about products. It is indicator of degree of interesting user for product.

According to reference [1] [18] rating play role very important to generate recommendation. There are many kinds of rating type that was implemented in recommendation technique, for examples:

1. Numerical Rating such as the 1-5 start. Many large e-commerce companies using this model instance Amazon, Lazada group, mobile application provider iTunes, play store.
2. Continues rating. The rating is special on continuous scale. An example the system that implemented in Jetster joke recommendation engine take a value rating between -10 and 10.
3. Ordinal rating, such as "strongly agree, agree, neutral, disagree, strongly disagree.
4. Binary rating that model choice in which the user is simply asked to decide if a certain item is good or bad.
5. Unary rating can include that a user has observed or purchased an item, or otherwise rated the item positively.

### 1.4. Factor Causes Cold Start Problem

From depth study, it becomes apparent that Cold start and sparsity data are main problem factor effecting in collaborative filtering based. These causes of problem is added new users or new items also both of them [8]. Recommender system based on collaborative filtering will obtain recommendation with better accuracy when available good information about rating is related to hunching and feeling of users who interested in items/products.

Accuracy degree of recommender system was leveraged not only availability of information, but also by quality of information. In case there is not enough information to build recommendation in a recommender system has been identified as sparsity data. Cold start is familiar problem in recommender system. In these cases, extreme sparsity data, in the other hand when there is no rating that gave the users to an item. Impact of cold start problem is no recommendation will be obtained in the system. The impact of sparsity data is the result of suggestion is not accurate. As we show in Table 3 as below, table is filled in number of rating and sign of "?", That mean user has no giving a rating from a movie. This implies that users' needs an effort to predict a rating score that unpredicted.

In the rating matrix only, a small percentage of elements get values. Even the most popular items may have only a few ratings. For example, in a large Netflix rating dataset provided for Netflix Prize competition [19] there are about 100 million ratings given by over 480,000 users to about 18,000 movies. There is only around 1% of rating matrix elements receiving ratings. With a sparse rating matrix, it is very challenging to estimate the relationship between items and users and make effective recommendation. Another well-known problem for collaborative filtering approach is the cold start problem, which can happen on new users or new items also both of them.

Cold start problem can be divided into complete cold start problem and incomplete cold start problem by whether number of rating records is zero or not. Generally, the sparsity of ratings for cold start items is higher than 85% [20], and the sparsity of rating for complete cold start items is 100 %. Illustration of incomplete rating matrix shown in below.

Journal of Theoretical and Applied Information Technology
15th May 2018. Vol.96. No 09

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 1. Illustration of cold start classification*

| | (not cold start) | | | | (incomplete cold start) | | | | (cold start) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *i1* | *i2* | *i3* | | *i1* | *i2* | *i3* | | *i1* | *i2* | *i3* | |
| *u1* | | | | - | | | | - | | | | - |
| *u2* | | | | ** | | | | - | | | | - |
| *u3* | | | | - | | | | - | | | | - |
| *u4* | | | | ** | | | | _ | | | | _ |
| *u5* | | | | - | | | | * | | | | - |
| *u6* | | | | ** | | | | - | | | | - |
| *u7* | | | | - | | | | - | | | | - |
| *u8* | | | | ** | | | | - | | | | - |
| *u9* | | | | ** | | | | - | | | | - |
| *u10* | | | | * | | | | - | | | | - |

### 1.5. Evolution Of Collaborative Filtering

Collaborative filtering technique makes mixture of the rating that served by many users to make product prediction of users needed. The major challenge for design collaborative filtering is how to solve rating matrices are sparse. Rating is the important thing as representation of user interest for a product or item.

The main idea of collaborative filtering technique is rating as representation of the feeling of users to items can be imputed because rating s analytic is having highly related between various user and item. Most of the technique of collaborative filtering are focus on influence inter user correlation and inter item correlation for predict recommendation process.

According to figure 3 on above, we show the variants of collaborative filtering. The development of recommendation fully influenced of disadvantage of older model that was assembled. For example, born the model based collaborative filtering consider to addressing of disadvantaged on mayor problem on memory based that have serious problem in scalability. The emergence of model based mainly based to enhance disadvantages of memory based.

#### 1.5.1. Memory Based

In the earliest collaborative filtering technique about in the middle 90s, the most popular method is memory based also famous as Neighborhood based, this technique very powerful to predict rating. Memory based is the method use prediction based on statistical method. These is several statistical methods that used Cosine, Spearman, Pearson.

Collaborative filtering models use the collective power of user's interest given by rating items to produce recommendation. The fundamental test in planning collaborative filtering techniques is that the underline rating matrices are inadequate. There are two variants of techniques are normally utilized powered in collaborative filtering, which are mentions to as memory-based methods and model-based methods.

Memory-based methods: Memory based methods are equaled as neighborhood based collaborative filtering algorithm. These were among oldest collaborative filtering algorithms; the basic ideas were the rating of user and item mixing are predicted on the basis of their neighborhoods. These will be explaining in two ways:

#### A. User-Based Collaborative Filtering

User based method are an effort to get recommendation that defined in order to identify similar users to the target users for whom the value rating prediction will be calculated. In order to determine the neighborhood of the target user *i*, her similarity to all the other users is calculated. A similarly function need to be defined between the rating special by users. Similarity computation is very difficult because different users may have different scale.

One of sample method to measure the similarity $sim(u, v)$ between the rating vector of two user that mention on above is called Person Correlation Coefficient.

$$\mu_u = \frac{\sum_{k \in I_u} r_{uk}}{|I_u|} \quad \forall u \in \{1 \ldots\ldots m\}$$

Next, Person Correlation Coefficient between the rows (users) *u* and *v* are explaining bellow:

$$Sim(u,v) = Pearson\,(u,v) = \frac{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u) \cdot (r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u)^2} \cdot \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk} - \mu_v)^2}}$$

For this situation, the evaluations given by similarly invested user of an objective user A are utilized as a part of request to make the propose for A. Along these lines, the essential thought is to decide user, who are like the objective user A, and prescribe evaluations for the surreptitiously appraisals of A by registering weighted midpoints of the appraisals of this recommender system.

#### B. Item-Based Collaborative Filtering

So as to make the rating prediction for target item B by user A, the initial step is to decide a set S of item that are most like target item B. The appraisals in item set S, which are determined by An, are utilized to anticipate whether the user A will like item B. Basic compute to implement neighborhood methods are used similarity formulation explained on bellow:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- If a is the active user for whom we seek recommendations, u another user and i and j two items, we will denote:
- I(a), I(u) and I(a&u) = I(a) ∩ I(u) the sets of items consumed by a, u, both a and u respectively.
- U(i), U(j) and U(i&j) = U(i) ∩ U(j) the set of users who consumed i, j and both i and j, respectively.
- $\vec{l}(u)$ the line of matrix R for user u and $\vec{c}(i)$ its column for item i.
- $\bar{l}(u)$ the average of $\vec{l}(u)$ (average rating given by *u* or average number of items consumed by *u*) and $\bar{c}(i)$ (*i's* average rating or number of users who consumed *i*).

$$\bar{l}(u)=\frac{1}{c}\sum_{i=1}^{c} r_{ui} \quad \bar{c}(i)=\frac{1}{L}\sum_{u=1}^{L} r_{ui}$$

The similarity between user's *a* and *u* can be defined through many similarity measures, for example Cosine (COS), Pearson correlation coefficient (PCC), Person's Correlation (COR) or Adjust Cosine (ACOS), Constrained Pearson's Correlation (CPC), Proximity, Impact, Popularity (PIP) and Spearman's Rank Correlation bellow respectively:

*Table 2. Method often used similarity measure*

| Measures | Formulation |
|---|---|
| Person's Correlation (COR) | $\text{Sim}(u_x,u_y)=\frac{\sum_{h=1}^{n'}(r_{u_x,i_h}-\overline{r_{u_x}})(r_{u_y,i_h}-\overline{r_{u_y}})}{\sqrt{\sum_{h=1}^{n'}(r_{u_x,i_h}-\overline{r_{u_x}})^2}\sqrt{\sum_{h=1}^{n'}(r_{u_y,i_h}-\overline{r_{u_y}})^2}}$ <br> where $r_{u,i}$ is the rating of the item *i* by user *u*, $\overline{r_u}$ is the average rating of user *u* for all the co rated items, and $n'$ is the number of items co rated by both users |
| Cosine (COS) | $\text{Sim}(u_x,u_y)=\frac{\sum_{h=1}^{n'} r_{u_x,i_h} r_{u_y,i_h}}{\sqrt{\sum_{h=1}^{n'} r_{u_x,i_h}^2}\sqrt{\sum_{h=1}^{n'} r_{u_y,i_h}^2}}$ <br> where $r_{u,i}$ is the rating of the item *i* by user *u* and $n'$ is the number of items co rated by both users |
| Adjust Cosine (ACOS) | $\text{Sim}(u_x,u_y)=\frac{\sum_{j=1}^{m'}(r_{u_j,i_x}-\overline{r_{u_j}})(r_{u_j,i_y}-\overline{r_{u_j}})}{\sqrt{\sum_{j=1}^{m'}(r_{u_j,i_x}-\overline{r_{u_j}})^2}\sqrt{\sum_{j=1}^{m'}(r_{u_j,i_y}-\overline{r_{u_j}})^2}}$ <br> where $r_{u,i}$ the rating of the item *i* by user *u*, $r_u$ is the average rating of user *u* for all the items rated by the user, and $m'$ is the number of users who rated both of the items |
| Constrained Pearson's Correlation (CPC) | $\text{Sim}(u_x,u_y)=\frac{\sum_{h=1}^{n'}(r_{u_x,i_h}-r_{med})(r_{u_y,i_h}-r_{med})}{\sqrt{\sum_{h=1}^{n'}(r_{u_x,i_h}-r_{med})^2}\sqrt{\sum_{h=1}^{n'}(r_{u_y,i_h}-r_{med})^2}}$ <br> where $r_{u,i}$ is the rating of the item *i* by user *u*, $r_{med}$ is the median value in the rating scale (e.g. 3 in the rating scale of 5), and $n'$ is the number of items co-rated by both users |
| Spearman's Rank Correlation | $\text{Sim}(u_x,u_y)=1-\frac{6\sum_{h=0}^{n'} d_h^2}{n'(n'^2-1)}$ <br> Where $d_h$ is the difference in the ranks for item *h* by the two users and $n'$ is the number of items co-rated by both users |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | |
|---|---|
| Proximity, Impact, Popularity (PIP) | Sim $(u_i, u_j) = \sum_{k \epsilon C_{i,j}} PIP(r_{i,k}, r_{j,k})$<br>where $r_{i,k}$ and $r_{j,k}$ are the ratings of item *k* by user *i* and *j*, respectively, $C_{i,j}$ is the set of co-rated items by user $u_i$ and $u_j$, and PIP $(r_{i,k}, r_{j,k})$ is the PIP score for the two ratings $r_{i,k}$ and $r_{j,k}$. For any two ratings $r_1$ and $r_2$, PIP $(r_1, r_2)$ = Proximity $(r_1, r_2)\cdot$ Impact $(r_1, r_2)\cdot$ Popularity $(r_1, r_2)$. |

### 1.5.2. Model Based

Memory based having crucial problem in scalability. It is a reason why many researchers have been tried to another strategy mostly focus in model based to improve time to compute in collaborative filtering to much more efficient for the time and computer resources. However, this method raises many benefit, there is also have some disadvantages rise too. For example, major problem in this method are cold start and sparsity data. Both of problem will happen when new user and news item actually new come in table matrix of collaborative filtering. While sparsity data will happen when data rating collected by user to an item is not enough. This condition will have impact the result of recommendation will not accurate.

In model-based techniques, machine learning and data mining strategies are used as a part of the setting of model based. In situations where the model is parameterized, the parameters of this model are found out inside the setting of an improvement structure. A few cases of such model-based techniques incorporate Decision Trees, Rule Based Models, Bayesian method and Neural Network and etc.

### 1.6. Evaluation For Recommender System

Accuracy metrics are utilized to assess either the prediction precision of evaluating the rating of specific user item mixed the precision of the top-k ranking predicted by a recommender system. Commonly, the ratings of a set R of sections in the rating matrix are covered up, and the precision is assessed over these covered up entries. Distinctive classes of strategies are utilized for the two cases.

Reference [10]Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), which is defined by Equation 1, Equation 2, equation 3 for recall, equation 4 for precision and equation 5 for *F1*.

*Table 3. Evaluation method often use*

| No. | Method | Formulation approach |
|---|---|---|
| 1 | Mean Absolute Error (MAE) | $MAE = \frac{\sum_{(u,i) \in R_{test}} \lvert R_{u,i} - R'_{u,i} \rvert}{\lvert R_{test} \rvert}$<br>Where $\lvert R_{test} \rvert$ represents the number of ratings in test set. $R_{u,i}$ Is the predicted rating for user u on item I and $R_{u,i}$ is the actual rating in test set A lower MAE or RMSE represents a high accuracy. Rating in test set A lower MAE or RMSE represents a higher predictive accuracy. |
| 2 | Root Mean Square Error (RMSE) | $RMSE = \sqrt{\frac{\sum_{(u,i) \in R_{test}} \lvert R_{u,i} - R'_{u,i} \rvert}{\lvert R_{test} \rvert}}$ |
| 3 | Recall | $recall = \frac{\sum_u L(N,u)}{\sum_u L(u)}$<br>Recall and Precision: These are the most popular metric for evaluating information retrieval system. Where *U* is the number of |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | |
|---|---|---|
| | | users. The upper limit of precision is 1, which means all of the items in recommendation list are relevant. |
| 4 | Precision | $$precission = \frac{\sum_u L(N,u)}{UN}$$ |
| 5 | *F1* | $$F_1 = \frac{2 * recall * precission}{recall + precission}$$ Sometime both of them (recall and precision) have conflict, the formulation approach to solve them between recall and precision are mention on above. |

## 2. CRYTICAL AND ANALYS STRATEGIC

The research questions (RQ) were specified to keep the review focused. The research questions addressed by this literature review. In order to achieve the goal of this particular research, an intensive literature survey has been carried out using as systematic way as possible. The search of relevant literature and to keep the concentration on the underlying interest has been based on a set of research questions (RQ). The formulated research questions are as follows:

RQ1. What the important component to develop a solution to handling cold start problem?

RQ2. How do the existing approaches perform to handle cold start problem?

RQ3. What dataset were used to conduct the research in cold start problem solution.

RQ4. What application domain were researches in cold start problem use?

RQ5. How to measure the perform of achievement to addressing cold start problem?

### 2.1. Collecting Source Of Paper

In this research, the authors use the source of the main digital library for the field of computer science that is Springer, IEEE Explore, ACM, Scopus, Sciencedirect. The table of digital library source from internet are listed in table 2 shown in below.

*Table 4. Source from Primary Digital Library*

| No | Sources | websites |
|---|---|---|
| 1 | Springer | http://link.springer.com |
| 2 | Sciencedirect | http://www.sciencedirect.com |
| 3 | IEEE | http://ieeexplore.ieee.org |
| 4 | ACM | http://dl.acm.org |
| 5 | Scopus | http://scopus.com |

### 2.2. Selecting Paper

In the selection of papers from the main source of digital libraries is to take only from the title that implicitly mentions the ecommerce recommender system topic especial in cold start problem. beyond the context it will be eliminated and here are the findings of the search by taking a span of eleven years from 2006 until 20016. The final result of selecting paper shown in table 5 on below. Total selecting papers result mostly 63 related paper from major digital library in computer science.

*Table 5. Result selecting paper*

| Digital Sources | Detail Selection |
|---|---|
| **IEEE Explore** | **38** |
| **ACM Digital Library** | **15** |
| **Scopus** | **9** |
| **Sciencedirect** | **6** |
| **Springerlink** | **4** |
| **Total** | **72** |

Result accumulation paper by year shown in graphic 2 in below. According the graphic, there is increasing significant growth number of research in recommender system field year by year.

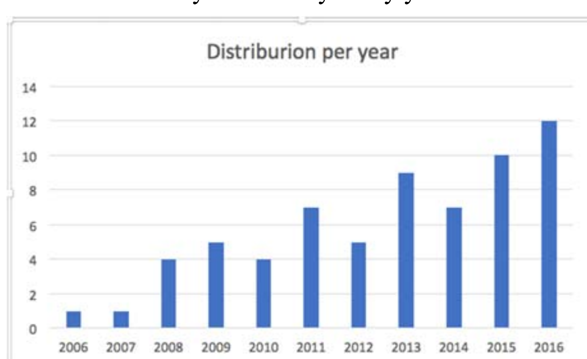


*Figure 1. Distribution paper by year*

## 3. RESULT AND DISCUSSION

The search results from the digital library are shown figure 2 on below. The distribution of cold start research related research shows that from year by year has increased. It means that the researcher's interest to conduct study in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

recommender system field, especially to solve cold start problem have increasing significantly. The author on reference [4] found when cold start problem is the most major problem in recommender system research field. this problem is very serious and again linked problems that will appear in the future is associated with big data which increasing too much information will being challenge for the future.

The publication distribution of research results in this field is shown in figure 3, where the majority of researchers publish in EEEI. there are more than 38 papers published in journals and conferences in ieee. second place is occupied by ACM. ACM is very concentrated in organizing researchers in this field and has a special specialty conference field called recsys. Researchers as many as 15 people published the results of his research at ACM. The third sequence is Scopus, then sciencedirect also member of Elsevier publisher group, and the last one is Springer.

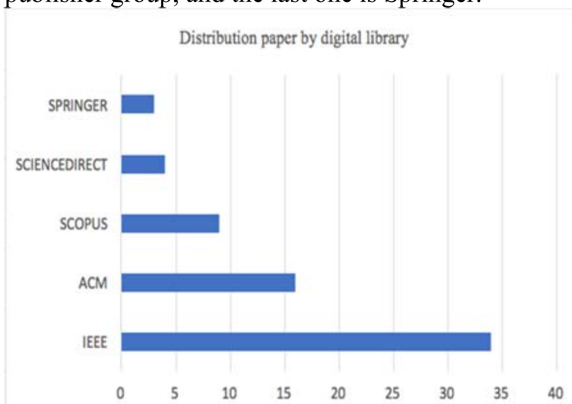


*Figure 2. Distribution paper by major digital library*

#### 3.1.1. RQ1: What the important component to develop a solution to handling cold start problem?

According to the observations from the review on above, there are two very important things how to develop recommender system that compromise with cold start problem that have been conducting by the researchers. The summaries of literature review result are shown on table 6, table 7 and table 8;

1. **Information:** enhances information from existing (user information and items feature) and b. exploring from another source such social media behavior, expression from comment, items metadata, location and etc. The summary review shown table 6 on below.
2. **Technique:** improves technical aspect for instance new engineering from existing method with machine learning and data mining. The summary review shown table 8 on below.
3. **Mathematical engineering**: Improves mathematical engineering such matrix factorization variant. The summary review shown table 8 on below.
4. **Hybridization method**: Hybrid is essential thing to refine a disadvantage method. The most popular method to conduct hybridization is by taxonomy that was created by Burk [6].

Trend of several researchers has been interesting to exploits information from social network due limitation of availability information come from users and items. There is so much information in social media applications, it is not easy to generate valid information from a lot of information available in social media and appropriate to embedding in users and items matrix and this become major challenge in recommender system research field. Improving cold start problem thru social media reached 30% have conduct study. It has become the second majority before enhance method to improve cold start problem by exploits users and items feature. Several researchers are still trying to reengineer machine learning method to solve the problem for instance conduct hybridization and majority conduct hybridization method to increase the performance result, some other uses new method like deep learning machine [21]. The group who choose this option reached 62%. The last one is reengineering mathematical method [22] and statistical method [23]. Complete result shown on figure 4.

Refer Figure 3, complete selecting study by comprehensive keyword, this is separated year of publication are shown on below. Author consider involving by years publication, technical approach to solve the problem, tools to measuring study result by metric approach, dataset often uses by researcher. All of them are shown in one graphical figure on below.

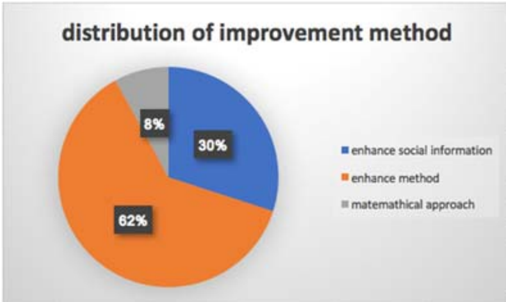


*Figure 4. Distribution of improvement method*

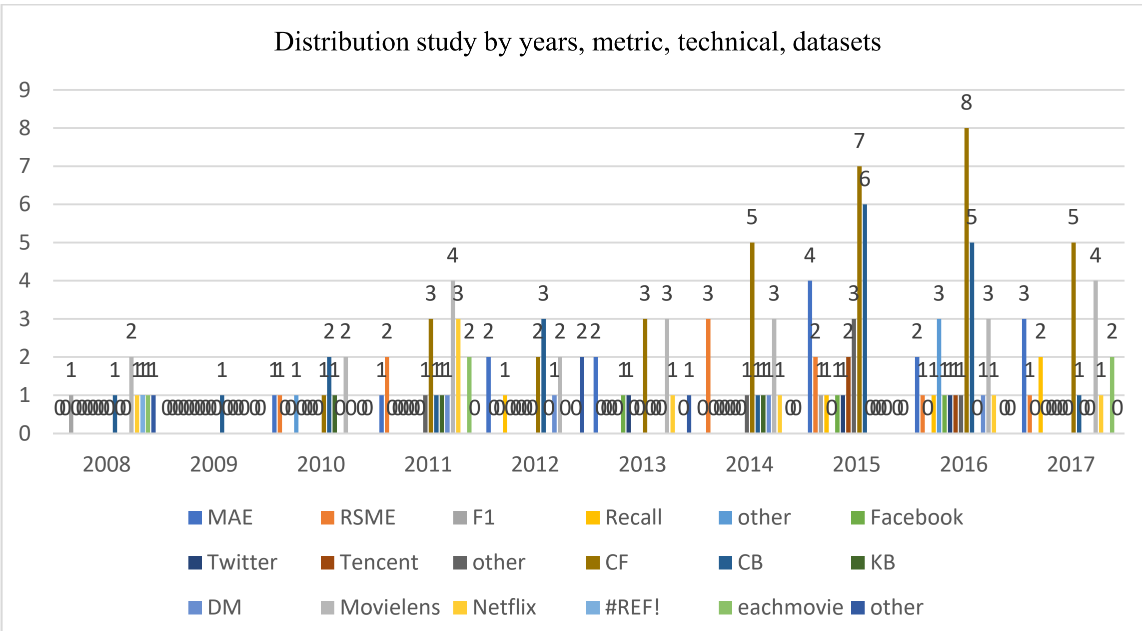


*Figure 5 Distribution study by years, metric, technical, datasets*

Detailed technical explanation for graphic figure on above is defined in table 6. In this table describe detail technical algorithm, result of metric measuring, source of reference, detailed dataset have been used by researcher. This is result of classification based on social network information to enhance information from external factor.

*Table 7. Enhance information from social network*

| No | Information Source | Algorithm | Single/hybrid | Evaluation | Ref. |
|---|---|---|---|---|---|
| 1 | Facebook (Private Data) | -Stochastic Expectation Maximization (EM) + -Non-Negative Matrix Factorization (NNMF) | -Stochastic -Collaborative Filtering [CF] | 99.99% $[MAE_r]$, 99.99% $[MAE_u]$, 100% [RMAE] | [24] |
| 2 | Twitter + App iTunes | -LDA - Matrix Factorization | -Collaborative Filtering [CF] -Content Based [CB] | Data sparse 98% improve up 30% | [25] |
| 3 | Private data | Social graph | -Jaccard | - | [26] |

| | | | | | |
|---|---|---|---|---|---|
| 4 | Twitter | -Genetic Algorithm<br>-Support Vector Regression<br>-text sentiment analysis | -Collaborative Filtering [CF]<br>-Content Based [CB] | - | [27] |
| 5 | Facebook | - Interview<br>-Decision Tree<br>-Person Correlation Coefficient [PCC] | -KNN<br>-Interview | [F-measure, threshold 3=0.3532 threshold 4=0.3482] [Precision, threshold 3=0.75 threshold 4=0.54] [recall threshold 3=0.231 threshold 4=0.257] | [28] |
| 6 | Tencent QQ | -Similarity PCC | -KNN | RSME=0.800 | [29] |
| 7 | Imhonet (Social network from Russia) | -Principal Modularity Maximization (PMM) + K-NN | -KNN | Graphical friend relationship | [30] |
| 8 | Flixter & Douban | -Discriminative selection<br>-KNN, similarity user rating | -KNN | - | [31] |
| 9 | Synthetic social network & Movielens | Nearest Neighborhood | -KNN | MAE less item high 0.87<br>MAE more item high 0.78 | [16] |
| 10 | Twitter | Corpus+ TF-IDF + tweet metadata | -Completion matrix | 770 users=72.67%<br>3500 users=53% | [32] |
| 11 | Douban | Integrates event content, organizer, location and user social relation | -Bayesian<br>-Poison Factorization | Beijing=446.6s[WMLFM] 11.7s[CBPF]<br>Shanghai=192.8s[WMLFM] 4.8s[CBPF] | [33] |
| 12 | Tencent Weibo | Incorporate Users Tags, Items keyword in the matrix | -Content Based [CB<br>-Collaborative filtering [CF] | 25% MAE= 0.2013<br>RSME= 0.3352 | [34] |
| 13 | Tencent Weibo | Incorporate Users Tags, Items keyword, social network in the matrix | -KNN<br>-Social regularization<br>-Content Based [CB] +<br>-Collaborative filtering [CF] | 50% MAE= 0.1622<br>50% Recall@10x= 0.3398<br><br>75% MAE= 0.1412<br>75% Recall@10x= 0.3625 | [35] |
| 14 | Foursquare | Location based social networks (LBSN) | -user frequency<br>-location frequency<br>-Similarity between both | S.LF. UF= 23.53<br>PSMM= 1.23<br>SHM= 5.58<br>CF= 23.59<br>gSCorr= 24.13 | [36] |
| 15 | Local (Indonesia) food commerce delivery | -Items based<br>-Users based<br>-Location preference | -Similarity mix between 3 matrixes | - | [37] |

Journal of Theoretical and Applied Information Technology
15th May 2018. Vol.96. No 09

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 8. Shown this is result of classification based on internal information to enhance information from internal factor. It also an essential factor to enrich information as a stage to handling cold start problem on recommender system.

*Table 9. Enhance users and items information*

| No | Information Source | Algorithm | Single/Hybrid | Evaluation | Ref. |
|---|---|---|---|---|---|
| 1 | DBpedia metadata | -Graph based<br>-Extract external meta data,<br>Joined user rating. | -Hybrid Graph<br>-generic matrix factorization | - | [28] |
| 2 | Movielens | -Collaborative filtering<br>-Content similarity<br>-Popularity prediction<br>-mix 3 matrixes | -Similarity users KNN | -Item MAE=0.9582 RSME=1.2321<br>-User MAE=0.8532 RSME=1.1014<br>-This method MAE=0.8085 RSME=0.9370 | [38] |
| 3 | Flixter, Netflix | Develop 3 corpus a) rating b) classification c) learning classification | -Selective naive Bayes classifiers | Netflix RSME= 0.862<br>Flixter RSME= 0.898 | [39] |
| 4 | Movielens & Netflix | Clustering user info and item info and mix between both become bi clustering | -PCC, K-means | RSME 45%=0.7390<br>RSME 60%=0.7218<br>RSME 75%=0.7202 | [40] |
| 5 | Design method | Associate rule and clustering technique | -Fuzzy | No result was explained | [41] |
| 6 | Movielens, Eachmovie, Netflix | Interview process to generate user profile. | -decision tree<br>-functional matrix factorization | MovieLens RSME=0.9098<br>Eachmovie RSME=1.2226<br>Netflix RSME=0.9703 | [42] |
| 7 | Movielens | Joining multi domain uses tags | -Wilcoxon rank-sum test | MAE UserItemTag 100%=0.800 | [43] |
| 8 | Movielens, Bookcrossing, FilmTrust | | -Bayesian similarity measure | Compare with, PIP, COS, PCC, MSD, SM. | [44] |
| 9 | - | Integrate content information about domain items into collaborative filter/ handle cold start item. | -Association Rule Mining with fuzzy | - | [45] |
| 10 | - | Combine users and item with Apache Mahout | -Content based<br>-Collaborative based | - | [46] |
| 11 | Movielens | Cluster rating matrix and content feature | -K-means,<br>-Decision tree,<br>-PCC. | MAE n=0, [3.1502]<br>MAE n=1, [0.8251] | [47] |
| 12 | Movielens, Douban | Dual discriminative selection | | MAP Items CS=50% 0.35<br>MAP Users CS=50% 0.3 | [48] |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | | | | |
|---|---|---|---|---|---|
| 13 | Movielens | Item feature based on actor and director | -Jaccard<br>-Cosine<br>-Spearman Rank. | Recall: [Average]<br>Jaccard: 0.241<br>MF: 0.2544<br>Cosine: 0.259<br>Pearson: 0.2536 | [49] |
| 14 | Movielens, Eachmovie, Netflix | Interview and build decision tree construct mixed matrix factorization | -Content based<br>-Collaborative | RSME Movielens=0.9098<br>RSME Eachmovie=1.2226<br>RSME Netflix=0.9703 | [50] |
| 15 | Movielens | incorporate user/item metadata into the LFM, | Majorization - Minimization technique | 1M [Datasets]<br>MAE Users CS: 0.7082<br>MAE Items CS: 0.7176<br>RSME Items CS:0.8984<br>RSME Users CS:0.9148 | [51] |
| 16 | Book crossing | Item feature (item taxonomy) | -Similarity measure KNN | F1 test<br>25%=0.005 | [52] |
| 17 | Netflix | -Extract content feature<br>-Hybrid CF with Time aware with TimeSVD++ | -Hybrid Deep learning<br>-TimeSVD++ | RSME 100=1.049<br>RSME 200=1.070<br>RSME 300=1.048 | [53] |
| 18 | Movielens | Transfer learning with Tradaboost, Cluster user-based item interest. | - | - | [54] |
| 19 | Netflix | Explores item content features learned from a deep learning transfers into TimeSVD++ | -Pearson Correlation,<br>-SDAE Deep learning,<br>-TimeSVD++ | RSME test<br>K100:1.049<br>K200:1.070<br>K300:1.048<br>K400:1.040 | [21] |
| 20 | Movielens | using clustering and also filtering to relieve cold start problem. | Hybrid method | MAE: Coverage F-measure<br>10:0.825 92.40% 0.850<br>20:0.808 91.80% 0.854 | [55] |
| 21 | Movielens | Clustering based users interest and demographic and construct decision tree structure. | | MAE test<br>MAE 0=1.8344<br>MAE 1=0.8522 | [56] |
| 22 | MovieLens | Parallel method use demographic, content information, previous rates. | -Demographic based<br>-Content based<br>-Hybrid | Precision test<br>10=81.23 | [57] |
| 22 | Movielens, Netflix, Jetster | Proposed new method similarity measure PIP | | | [23] |
| 23 | Movielens | User demographic, Users classification, neighborhood similarity. | -Demographic based<br>-Collaborative based | - | [58] |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | | | | |
|---|---|---|---|---|---|
| 24 | Comoda | Content based & demographic | -weighted<br>-switching hybrid | - | [59] |
| 25 | Facebook, Wikipedia | Cross domain | -Personality based matrix factor<br>-use active learning, based cross domain. | - | [60] |
| 26 | Netflix & Movielens | Enhance Users and item profile | -Enhance profile expansion uses query expansion techniques | - | [61] |
| 27 | - | Extract music content feature uses deep learning | -Content based<br>-Collaborative based | RMSE test<br>PMF= [0.0109] [0.0110]<br>HLDBN= [0.0132] [0.0131] | [62] |
| 28 | TV1 | Items feature, IPTV Provider | -Content based | - | [63] |
| 29 | Movielens | semi-supervised co-training algorithm | -Explore context aware by matrix | RMSE test<br>Users based KNN=1.0250<br>Items based KNN=1.0733<br>CSELL1=0.9013<br>CSELL2=0.8987<br>CSELL3=0.9012 | [64] |
| 31 | Twitter | critiquing-based | -Content based | | [65] |

Reference table 10. this is result of classification based on enhance from mathematical approach. It also an essential factor to reduce dimensional reduction. This method very popular after competition that had been held by Netflix to improve accuracy.

*Table 11. Enhance mathematical method*

| No | Source | Method | Single/hybrid | Evaluation | Ref. |
|---|---|---|---|---|---|
| 1 | Movielens | Factorization with decoupled completion & transduction | | - | [66] |
| 2 | Movielens & Eachmovie | Complete consideration users and items information | -regression approach based on profiles | - | [67] |
| 3 | Synthetic Dataset | Non-negative matrix factorization | -Enhance matrix factorization | - | [22] |
| 4 | Movielens, Netflix, Jetster | Proposed new method similarity measure PIP | -Statistical approach | - | [23] |

#### 3.1.2. RQ2: How do the existing approaches perform to handle cold start problem?

According the results of point 2.3.1, most researchers attempt to solve cold start problems into three parts: exploits information from external factors such as behavior in social media, optimizing information of users and information of items and then improving the mathematical approach. Some of researcher conduct hybridization between three important components on above.

Social media becomes some very interesting components to solve the cold start problem because in this decade the development of social media is very fast like Facebook, twitter, Instagram. Those are social media applications that very popular in the world with the number of users reaching in billions. Based on this social media information could be extracted interest information about group or person who interested for products or services such books, films, television shows and music and etc. [32] [25] [28].

There are also researchers who dig data from public data base collections such as imdb, Wikipedia, TripAdvisor [68]. Database is used to explore the public interest in a product. This public database is a more contextual representation in solving cold start problems. Some researchers analyze public opinion using text mining as a representation of the interest and satisfaction of a service. It is useful to get information of person's or group's interest in a product or service.

The first of the most majority researcher tried to solve cold start problem with enhance item information and user information to improve quality of information. There is several information suitable to build recommender systems for instance items metadata, users demographics, user opinion or comments for a movie, example researcher using imdb opinion. Several researchers conduct hybridization of one method to another to bringing a novel method and algorithm to increase the performance of cold start problem.

The last one of some method to increase performance to solve cold start is uses enhance mathematical method, for instance non-negative matrix factorization [22]. Indeed, this method cannot increase performance significantly, this method has function to support for other method increase performance for example in accuracy point of view.

#### 3.1.3. RQ3: What datasets were used to conduct the research in cold start problem solution

There are two types of datasets: public datasets and private data sets. Private datasets belong to private companies and they are not distributed as public datasets. Public data sets belong to private companies and they distributed to public. According based on figure 5, majority study in ecommerce recommender system field uses public dataset until 71%, and another researcher usage private datasets 29%. Private datasets in here have explanation as researcher generate the data from private company who have rule limited public access, also they have capturing the user behavior come from social network who they have unlimited access to explore the information for example group activity, group interest, user profile and activity [24] .

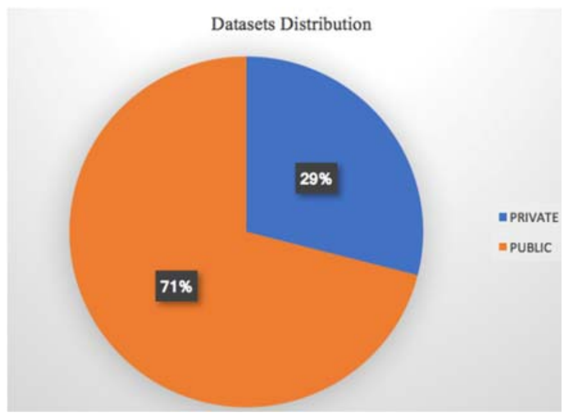


*Figure 6. Distribution type of datasets*

Distribution public dataset are shown in figure 6 on below. Majority researcher use Movielens datasets more than 31%. Movielens very popular for research in recommender system due this popularity is to a certain degree, a reflection of the incredible rate of growth of personalization and recommendation research, where datasets such as these have substantial value in exploring and validating ideas[69]. The second one majority use Netflix datasets more than 11%. The benefit uses public dataset is author can be compared with another author result in same field area of research. The most dataset was used Movielens and Netflix, those shown on figure 6. These is character datasets consist MovieLens and Netflix on below, according:

1. **Netflix Prize Dataset:** In October 2006 Netflix released a large movie rating dataset and challenged the data mining, machine learning

Journal of Theoretical and Applied Information Technology
15th May 2018. Vol.96. No 09


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and computer science communities to develop systems that could beat the accuracy of Cinematics by certain amounts [19]. The largest of the available datasets, this set of movie ratings consists of 100 million ratings, by 480, 189 users who have rated one or more of the 17, 770 movies on a 1−5 star scale [70].

2. **MovieLens Dataset:** There are currently three movie-rating datasets available from the Group Lens website6. The first includes 100, 000 ratings of 1, 682 movies by 943 users; the second has 1 million ratings for 3, 900 movies by 6, 040 users; the last includes 10 million ratings and 100, 000 tags for 10, 681 movies by 71, 567 users. As with the Netflix dataset, items are rated on a 1−5 star scale. In this thesis, we have used the first two datasets since the third dataset was only released in 2009. The MovieLens datasets, first released in 1998, describe people's expressed preferences for movies. These preferences take the form of <user, item, rating, timestamp> tuples, each the result of a person expressing a preference (a 0-5 star rating) for a movie at a particular time [69].

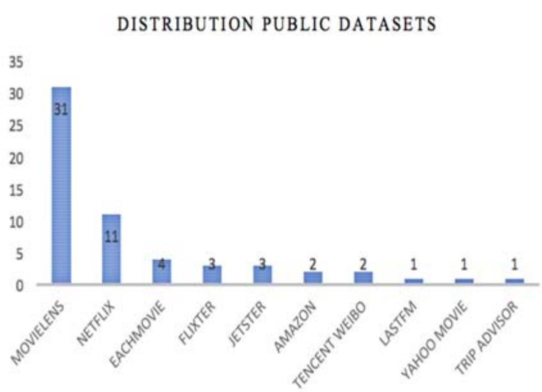


*Figure 7. Distribution of public datasets*

### 3.1.4. RQ4: What application domain was used by researcher to research in cold start problem?

According to the figure 7 on above, the result literature review, majority application object to conduct the research are movie application, then web content, book store, multi-product and etc.

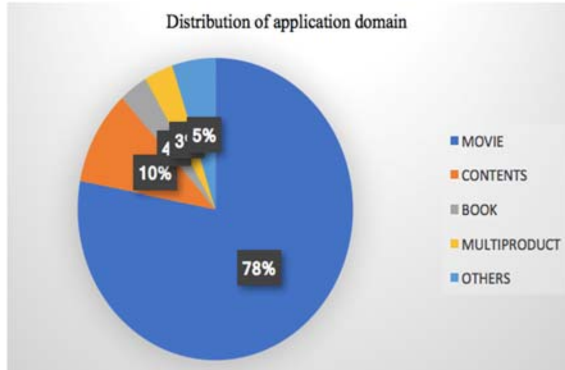


*Figure 8. Application domain*

### 3.1.5. RQ5: How to measure the perform of achievement the result of study?

There are several ways to measure the outcome of the development of methods to solving cold start problems. Here are some of the methods used by the researchers shown in the figure 8. the method most often used by the majority of researchers is MAE (mean absolute error), second order is RMSE (root mean square error) the next is recall and precision. Detail explanation of formula have explained in table 2 on above.

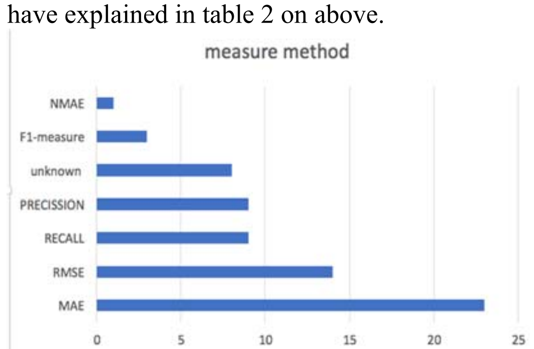


*Figure 9. Measure method*

## 4. LIMITATION

Online shopping often called ecommerce have been applied in many domains of online business applications, certainly, there are many unique technical methods to handling cold start problems based on application domain point of view. This study ignores the uniqueness of handling with domain types of online shopping applications. Subsequent research, researchers will try to uniquely resolve unique approaches based on application domains.

Another limitation on this research, answering research questions in this research there is unambiguous answer, because in this research there is some evidence that cannot be compared apple to apple, this is caused by some researchers who sometimes do not include the test results clearly and firmly, also caused types of data sets that are very numerous and varied, metric measurements and selection of technical method by the researchers. However, it can generally be answered by statistical means although this is sometimes biased. The answer statistically, for example, the use of social media to be involved from year to year is increasing, according to the writer's understanding can be interpreted opportunities and challenges in the future involves social media as a source of information that can be utilized optimally to build a more perfect recommender system.

## 5. CONCLUSION

In all of research that we were conducted, behavior in social media application cannot be integrated in ecommerce application in real industry yet in some popular social media application research could be integrated in data and synthetic applications. But, there are two researches that successful integrated social media behavior into ecommerce recommender system that is Tencent QQ and Weibo, it is easier to do because the ownership of both applications is in the same corporation. In the other hand, it easy to incorporate between social media behavior and real ecommerce industry because there are in the same platform and same corporation.

In this literature review we conducted to analyzed from 70 major study consist journal and conference proceeding with focus in cold start problem. We have investigated the most interested problem to serve better recommendation with solving cold start problem. We also investigate several studies tried to enhance main information source, modification from existing method, enhance user profile and item profile, improve mathematical engineering, kind of application domain, several datasets may they used, kind of evaluation method and opportunity and challenge in the future research.

We have classified the method to solve the cold start problem divided into three strategies included exploring external information use like social media environment, location-based movement, the second enhance existing information user information and item information into new technical method also modified from existing method and the third is reengineering mathematical method. We found that hybrid method was used in mainly study, some of them combine knowledge based with collaborative filtering, also content based and collaborative filtering, there is some of them combine demographic, content based and collaborative filtering. Collaborative filtering become the most majority involve with another method.

We have presented the classification method that used in cold start problem, this method is very common to used similarity measure event user similarity, item similarity, behavior similarity, environments similarity in social network. This method very common called K-NN in specific technical. KNN classification technique is favorite established to develop neighborhood in recommender system based collaborative filtering. We detected the different hybridization consideration that related in taxonomy by Burk and conclude that the most method to hybridization is weight technique.

In evaluation method, accuracy is still becoming the most critical target to consideration. The research dominated involved comparison with same method and also consideration prediction metric for measure the result. The most popular way to evaluation metric is RME (root mean error) and RMSE (root mean square error). Although this evaluation method was used majority used by researcher, there is disadvantage contextual consideration point of view. Evaluation metric just involve completion matric and accuracy without contextual by customer or user. Another method that more credible method is based on feed back or survey. It is highly suggested to measure user satisfaction of recommendation result. We also presented what public datasets commonly used in cold start problem study. The most favorite datasets are MovieLens.

We also found in last 5 years, enhance information from social network was interest several authors to explore, social media become something beautiful to exploits, because in there is available to much information. Some author tried to detected similarity behavior properties in social media users, this is inspired from homophily theory. We regard in the future research, social media become the opportunity and big challenge, it is not only to solve some of the problem in recommender system for instance cold start problem, but also to increase contextual point of view in recommender system.

## 6. ACKNOWLEDGMENT

We would say to thank for those support our literature review research. The organization that support our research is University Teknikal Malaysia Melaka, Melaka, Malaysia, University Amikom of Yogyakarta, Yogyakarta, Indonesia and also funded by PT. Time Excelindo, Yogyakarta,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the Holding Company in Indonesia. I never forget, my deep thank for double blind reviewer who have given critique and suggestion to make improvement for this paper review.

**REFERENCE:**

[1] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, vol. 54. 2015.

[2] L. Candillier, K. Jack, F. Fessant, and F. Meyer, *Recommender systems*, vol. 40, no. 3. 1997.

[3] D. Kotkov, S. Wang, and J. Veijalainen, "A survey of serendipity in recommender systems," *Knowledge-Based Syst.*, vol. 111, no. August, pp. 180–192, 2016.

[4] E. Çano and M. Morisio, "Hybrid recommender systems: A systematic literature review," *Intell. Data Anal.*, vol. 21, no. 6, pp. 1487–1524, 2017.

[5] F. Xia, N. Y. Asabere, A. M. Ahmed, J. Li, and X. Kong, "Mobile multimedia recommendation in smart communities: A survey," *IEEE Access*, vol. 1, pp. 606–624, 2013.

[6] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *CEUR Workshop Proc.*, vol. 1621, no. November, pp. 36–43, 2002.

[7] D. Bernardes, M. Diaby, E. Viennet, W. Labs, and R. Moncey, "A Social Formalism and Survey for Recommender Systems," vol. 16, no. 2, pp. 20–36.

[8] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 10059–10072, 2012.

[9] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Adv. Artif. Intell.*, vol. 2009, no. Section 3, pp. 1–19, 2009.

[10] Z. Yang, B. Wu, K. Zheng, X. Wang, and L. Lei, "A survey of collaborative filtering-based recommender systems for mobile internet applications," *IEEE Access*, vol. 4, no. c, pp. 3273–3287, 2016.

[11] E. Karydi and K. Margaritis, "Parallel and Distributed Collaborative Filtering: A Survey," *ACM Comput. Surv. Artic.*, vol. 49, no. 37, 2016.

[12] W. Pan, "A survey of transfer learning for collaborative recommendation with auxiliary data," *Neurocomputing*, vol. 177, pp. 447–453, 2016.

[13] L. Zheng, "A Survey and Critique of Deep Learning on Recommender Systems," no. September, p. 31, 2016.

[14] S. Chen, S. Owusu, and L. Zhou, "Social Network Based Recommendation Systems: A Short Survey," *2013 Int. Conf. Soc. Comput.*, pp. 882–885, 2013.

[15] U. Nadine, H. Cao, and J. Deng, "Competitive Recommendation Algorithm for E-commerce," no. 1, pp. 1539–1542, 2016.

[16] P. S. Thilagam, "Alleviating Data Sparsity and Cold Start in Recommender Systems using Social Behaviour," 2016.

[17] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Syst.*, vol. 46, pp. 109–132, 2013.

[18] F. Ricci, L. Rokach, and B. Saphira, *Recommender Systems Handbook*, 2nd ed. Springer Berlin Heidelberg.

[19] J. Bennett and S. Lanning, "The Netflix Prize," *KDD Cup Work.*, pp. 3–6, 2007.

[20] E.-P. L. Chenyi Zhang, Ke Wang, Hongkun Yu, Jianling Sun, "Latent Factor Transition for Dynamic Collaborative Filtering," *Icdm*, pp. 452–460, 2014.

[21] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative Filtering and Deep Learning Based Hybrid Recommendation for Cold Start Problem," *2016 IEEE 14th Intl Conf Dependable, Auton. Secur. Comput. 14th Intl Conf Pervasive Intell. Comput. 2nd Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr.*, pp. 874–877, 2016.

[22] U. Ocepek, J. Rugelj, and Z. Bosni??, "Improving matrix factorization recommendations for examples in cold start," *Expert Syst. Appl.*, vol. 42, no. 19, pp. 6784–6794, 2015.

[23] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem," *Inf. Sci. (Ny).*, vol. 178, no. 1, pp. 37–51, 2008.

[24] Y. Xie *et al.*, "Elver: Recommending Facebook pages in cold start situation without content features," *Proc. - 2013 IEEE Int. Conf. Big Data, Big Data 2013*, pp. 475–479, 2013.

[25] J. Lin, "Addressing Cold-Start in App Recommendation : Latent User Models Constructed from Twitter Followers Categories and Subject Descriptors," 2013.

[26] A. Hannech, M. Adda, H. Mcheick, and C. Science, "Cold-start recommendation strategy based on social graphs," 2016.

[27] D. H. Alahmadi and X. J. Zeng, "Twitter-based

recommender system to address cold-start: A genetic algorithm based trust modelling and probabilistic sentiment analysis," *Proc. - Int. Conf. Tools with Artif. Intell. ICTAI*, vol. 2016–Janua, pp. 1045–1052, 2016.

[28] P. Bedi, C. Sharma, P. Vashisth, D. Goel, and M. Dhanda, "Handling cold start problem in Recommender Systems by using Interaction Based Social Proximity factor," *2015 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2015*, pp. 1987–1993, 2015.

[29] L. J. M. Dali and Q. Zhiguang, "Cold-Start mastered: LebiD1," *Proc. - 17th IEEE Int. Conf. Comput. Sci. Eng. CSE 2014, Jointly with 13th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC 2014, 13th Int. Symp. Pervasive Syst.* , pp. 181–184, 2015.

[30] S. Sahebi and W. W. Cohen, "Community-based recommendations: a solution to the cold start problem," *Work. Recomm. Syst. Soc. Web (RSWEB), held conjunction with ACM RecSys'11*, 2011.

[31] S. Qiu, J. Cheng, X. Zhang, B. Niu, and H. Lu, "Community discovering guided cold-start recommendation: A discriminative approach," *Proc. - IEEE Int. Conf. Multimed. Expo*, vol. 2014–Septe, no. Septmber, 2014.

[32] M. Moh, "Using Social Media Presence for Alleviating Cold Start Problems in Privacy Protection," pp. 11–17, 2016.

[33] W. Zhang and J. Wang, "A Collective Bayesian Poisson Factorization Model for Cold-start Local Event Recommendation," *SIGKDD 2015 Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1455–1464, 2015.

[34] K. Ji and H. Shen, "Addressing cold-start: Scalable recommendation with tags and keywords," *Knowledge-Based Syst.*, vol. 83, no. 1, pp. 42–50, 2015.

[35] K. Ji and H. Shen, "Jointly modeling content, social network and ratings for explainable and cold-start recommendation," *Neurocomputing*, vol. 218, pp. 1–12, 2016.

[36] H. Gao, J. Tang, and H. Liu, "Addressing the cold-start problem in location recommendation using geo-social correlations," *Data Min. Knowl. Discov.*, vol. 29, no. 2, pp. 299–323, 2014.

[37] N. Suryana, A. Samad, and B. I. N. Hasan, "PAPER SURVEY AND EXAMPLE OF COLLABORATIVE FILTERING IMPLEMENTATION IN RECOMMENDER," vol. 95, no. 16, 2017.

[38] J.-H. Wang and Y.-H. Chen, "A Distributed Hybrid Recommendation Framework to Address the New-User Cold-Start Problem," *2015 IEEE 12th Intl Conf Ubiquitous Intell. Comput. 2015 IEEE 12th Intl Conf Auton. Trust. Comput. 2015 IEEE 15th Intl Conf Scalable Comput. Commun. Its Assoc. Work.*, pp. 1686–1691, 2015.

[39] D. Poirier, F. Fessant, and I. Tellier, "Reducing the cold-start problem in content recommendation through opinion classification," *Proc. - 2010 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2010*, vol. 1, pp. 204–207, 2010.

[40] D. Zhang, C. H. Hsu, M. Chen, Q. Chen, N. Xiong, and J. Lloret, "Cold-start recommendation using Bi-clustering and fusion for large-scale social recommender systems," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 2, pp. 239–250, 2014.

[41] H. Sobhanam and a. K. Mariappan, "Addressing cold start problem in recommender systems using association rules and clustering technique," *2013 Int. Conf. Comput. Commun. Informatics*, pp. 1–5, 2013.

[42] K. Zhou, S.-H. Yang, and H. Zha, "Functional matrix factorizations for cold-start recommendation," *Proc. 34th Int. ACM SIGIR Conf. Res. Dev. Inf. - SIGIR '11*, p. 315, 2011.

[43] M. Enrich, M. Braunhofer, and F. Ricci, "Cold-Start Management with Cross-Domain Collaborative Filtering and Tags," pp. 101–112, 2013.

[44] G. Guo, "Integrating Trust and Similarity to Ameliorate the Data Sparsity and Cold Start for Recommender Systems."

[45] S. C. F. Chan, "Applying Cross-Level Association Rule Mining to Cold-Start Recommendations Applying Cross-Level Association Rule Mining to Cold-Start Recommendations," no. December 2007, 2014.

[46] M. Sarumathi and S. Singarani, "Systematic Approach for Cold Start Issues in Recommendations System," 2016.

[47] D. Sun, Z. Luo, and F. Zhang, "A novel approach for collaborative filtering to alleviate the new item cold-start problem," *11th Int. Symp. Commun. Inf. Technol. Isc. 2011*, no. Iscit, pp. 402–406, 2011.

[48] X. Zhang, J. Cheng, S. Qiu, G. Zhu, and H. Lu, "DualDS: A dual discriminative rating elicitation framework for cold start recommendation," *Knowledge-Based Syst.*, vol. 73, pp. 161–172, 2015.

[49] P. Yi, C. Yang, X. Zhou, C. Li, and A. C.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Filtering, "A Movie Cold-Start Recommendation Method Optimized Similarity Measure."

[50] K. Zhou, S. Yang, and H. Zha, "Functional Matrix Factorizations for Cold-Start Recommendation," 2011.

[51] A. Gogna and A. Majumdar, "A comprehensive recommender system model: Improving accuracy for both warm and cold start users," *IEEE Access*, vol. 3, pp. 2803–2813, 2015.

[52] L. Weng, Y. Xu, Y. Li, and R. Nayak, "Exploiting Item Taxonomy for Solving Cold-start Problem in Recommendation Making," pp. 113–120, 2008.

[53] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items," *Expert Syst. Appl.*, vol. 69, pp. 1339–1351, 2017.

[54] "A Two-Stage Cross-Domain Recommendation for Cold Start Problem in Cyber Physical Systems."

[55] C. Huang and J. Yin, "Effective association clusters filtering to cold-start recommendations," *Proc. - 2010 7th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2010*, vol. 5, no. Fskd, pp. 2461–2464, 2010.

[56] D. Sun, C. Li, and Z. Luo, "A Content-Enhanced Approach for Cold-Start Problem in Collaborative Filtering," pp. 4501–4504, 2011.

[57] J. Basiri, "Alleviating the Cold-Start Problem of Recommender Systems Using a New Hybrid Approach," pp. 962–967, 2010.

[58] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems," *Expert Syst. Appl.*, vol. 41, no. 4 PART 2, pp. 2065–2073, 2014.

[59] V. N. Zhao, M. Moh, and T. S. Moh, "Contextual-Aware Hybrid Recommender System for Mixed Cold-Start Problems in Privacy Protection," *Proc. - 2nd IEEE Int. Conf. Big Data Secur. Cloud, IEEE BigDataSecurity 2016, 2nd IEEE Int. Conf. High Perform. Smart Comput. IEEE HPSC 2016 IEEE Int. Conf. Intell. Data S*, pp. 400–405, 2016.

[60] I. FernÃ¡ndez-TobÃas, M. Braunhofer, M. Elahi, F. Ricci, and I. Cantador, "Alleviating the new user problem in collaborative filtering by exploiting personality information," *User Model. User-Adapted Interact.*, vol. 26, no. 2–3, pp. 221–255, 2016.

[61] V. Formoso, D. Fernández, F. Cacheda, and V. Carneiro, "Using profile expansion techniques to alleviate the new user problem," *Inf. Process. Manag.*, vol. 49, no. 3, pp. 659–672, 2013.

[62] X. Wang and Y. Wang, "Improving Content-based and Hybrid Music Recommendation using Deep Learning," pp. 627–636.

[63] C. Paper, P. C. Politecnico, V. Recommendation, M. Content, F. View, and P. Cremonesi, "Analysis of cold-start recommendations in IPTV systems Analysis of Cold-Start Recommendations in IPTV Systems," no. January 2009, 2017.

[64] M. Zhang, J. Tang, X. Zhang, and X. Xue, "Addressing cold start in recommender systems," *Proc. 37th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '14*, pp. 73–82, 2014.

[65] Y. Salem, J. Hong, and W. Liu, "CSFinder: A cold-start friend finder in large-scale social networks," *Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015*, pp. 687–696, 2015.

[66] I. Barjasteh, R. Forsati, D. Ross, A. H. Esfahanian, and H. Radha, "Cold-Start Recommendation with Provable Guarantees: A Decoupled Approach," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 6, pp. 1462–1474, 2016.

[67] S.-T. Park and W. Chu, "Pairwise Preference Regression for Cold-start Recommendation," *Proc. Third ACM Conf. Recomm. Syst.*, vol. 37, pp. 21–28, 2009.

[68] L. Peska and P. Vojtas, "Using Linked Open Data in Recommender Systems," *Proc. 5th Int. Conf. Web Intell. Min. Semant.*, p. 17:1--17:6, 2015.

[69] F. M. Harper and J. A. Konstan, "The MovieLens Datasets: History and Context," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, p. 19:1--19:19, 2015.

[70] N. K. Lathia, "Evaluating Collaborative Filtering Over Time," 2010.