Serendipity in Recommender Systems

Ashishkumar Patel [#1], Kiran Amin[#2]

[#1] Research Scholar, C U Shah University, Wadhwan City, Gujarat, India
[#1] Computer Engineering Department, LDRP ITR, KSV University, Gandhinagar, Gujarat, India
[#2] Professor, U V Patel College of Engineering, Gapat University, Kherva, Gujarat, India
[1] ashish.ldrp@gmail.com
[2] kiran.amin@ganpatuniversity.ac.in

Abstract—**Today the growth of information technology, social networking and commercial websites are heavily increased. Because of it the information available for users is increased. Because of this information overload users cannot select best items from available various items. To solve this problem Recommender System (RS) required to explicit preferences and monitoring of implicit behaviour of user. The traditional RS recommends the items to the user based on their choice. But after some time user become bored with same kind of recommendations again and again. To provide surprise in the presented items another king of qualitative parameter required namely - "Serendipity". Serendipity refers to the "phenomenon of spontaneously understanding unexpected things, including time, location, people, and contents". This paper is detailed survey about serendipity in various aspects in study, equations, components, related works and measures of serendipity in RS.**

Keyword - Serendipity, Recommender System, Information Overload, Item Preferences

I. INTRODUCTION

The Recommender Systems recommends items, which might be useful for users. However, these items are not interesting if user is already familiar with the same product. Therefore, the novelty and serendipity in the recommended list are essential to make the recommendation interesting [12]. In current situation because of explosion of items in social networks and ecommerce, the research in serendipitous recommendation is essential. The word "Serendipity" comes from Old Persian tale whose name is "The Three Princes of Serendip". This tale was translated and publish by Cristoforo Armeno in 1557. Then after almost two hundred years, Horace Walpole use the word "Serendipity" means that to make the discovery by accident. The use of this terminology in the field of science by Foster and Ford in 2003.

However, in RS research, serendipity means to recommend unexpected, unforeseeable or surprising items [12]. The recommended items to the user are serendipitous if items are based on user's test, are unknown and they are difficult to search it.

Traditional RS only focus on the accuracy for recommending items. Recent studies suggests that beyond the accuracy other factors also should be considered for recommendation. Accuracy is one of the most required criteria but to present variety in the recommended list is also important concern [22]. Researcher says that instead of presenting the list of items in the recommendation with same properties and contents to list the items with different test and properties will more satisfies the users [17].

Before submitting your final paper, check that the format conforms to this template. Specifically, check the appearance of the title and author block, the appearance of section headings, document margins, column width, column spacing and other features.

II. COMPONENTS OF SERENDIPITY

The serendipity in computational concept is defined and use by Alison Pease et.al. [20]. Components of serendipity [18, 4] are as below:

Prepared Mind: It represent the background domain knowledge or knowledge base of the system. In recommender system, the prepared mind is user's experience about the item. The perception of the user in twitter for the sender [23]. It also represent the goal and unsolved problems.

Serendipity Trigger: It is the process, which proceeds the discovery, which will lead to the serendipity. It includes the event, occurrence of phenomena etc.

Bridge: It connects the trigger to the result. It includes heuristic techniques or analogical reasoning, genetic algorithms, automata theory etc. Some techniques are more creative than others, which may useful for accuracy of the RS but for serendipity oriented RS if the less similar items are added with the high similar items which lead to the surprise.

Result: This is the final output of RS or the new discovery. It includes a new process, problem, hypothesis, product, artifact, object etc. If the result is unknown, un1expected, coincidental then it will be definitely serendipitous.

III. STEPS FOR SERENDIPITOUS RECOMMENDATION

For serendipitous discovery, there is three steps process define by Jordanous (2012). Steps for serendipitous recommendations [9] are as below:

Step 1: Definition of Serendipity for given context. Discovery of serendipitous recommendation can be based on

a) Background knowledge previously uninteresting things, which trigger as an interesting due to chance.

b) The processing background knowledge and reevaluated trigger using analogical or reasoning techniques, we obtain a new things, which will be win-win for both RS provider and RS users.

Step 2: Specify standards to measure serendipity.

The measuring of serendipity is totally depends on the system, types of data and expectations of the users. The different ways to measure the serendipity in literature is described in [5, 13, 21].

Step 3: Testing and evaluating the generated results with standards defined in above step.

IV. RELATED WORKS ON SERENDIPITY

Folksonomy Based

The conventional classifications are top down approach but the folksonomy is the bottom up classification technique [7]. This technique also uses the tag information on the items assigned by the users. In folksonomy-based classification, the items are classified based on tags on items. In this approach, also the standard books are serendipitously recommended.

In this system, users have to assign at least two tags on book. Based on the tags assigned by the user on the book, the impression of user regarding the book is created which represents the feelings regarding the book. In this method using "concept" regarding book, system finds the useful and unexpected book for the user and recommends top-n books to the users.

www Access Log Based

In this research, the authors used the access log information of web server to find the serendipitous recommendation [14]. They uses the session information and create groups based on co-occurrence of items and then select appropriate group by degree of serendipity.

In this RS there are two phases, one is for pre-processing and other is for recommendation. In the pre-processing phase, the session based data extracted from server logs are converted into item based data. For this operation, the inverted index concept is used. This structure of data is used to access efficient information of the specific item. The second phase is of the recommendation list generation. In this phase, the items are divided into groups based on session of the users. Initially the groups are searched which are similar to active session. Then recommended items are extracted from source of items that are in the same session as the group. If these recommended items belongs to different groups as compared to the selected within the same session then they are considered as serendipitous recommendations [14].

Social Serendipity or Serendipity in Social Networks (SNS)

The social information for a user contains many of the things like friends, like, dislike, group, timestamp etc. In this study, authors proposed method to find "surprising" relationships between two users who attend an event [19]. General idea is that, when people attend an event like conference, seminar, social event etc. then they were not aware about other people who have similar kind of interest. Authors proposed an approach which find such types of people using Flickr photo pool. They utilize the Flickr tags to find similar users, i.e. two find two group members who have been in similar place but they are not friend. They identified the similar users by using geo-tagged photos.

In another approach for serendipitous recommendation the authors uses the item access record and interactive information from user's social networking websites [19, 25]. In this approach, the serendipity of recommendation decided based on friendship information, recency and frequency of the data. This approach also uses user's self, his/her friends' and database frequency.

Using Bridging Score (BRS)

In this research author discovers serendipitous items by using utility function for "surprise me" query [21] Authors proposed an algorithm which calculates the bridging score which is the connectivity to other groups in the graph as shown in figure. It uses graph mining technique to calculate the similarity between nodes. If any dataset is mapped in the graph then this technique can be used for any types of dataset.

In order to enable serendipitous discoveries, this paper estimates the utility function for a "surprise me" query. The paper proposes an algorithm to calculate the bridging score. The algorithm uses graph mining technique to calculate similarity between nodes. The bridging score can be calculated based on connectivity of

users in group one with other group as shown in figure which illustrates the difference in recommendations of traditional approaches versus the BRS heuristic. This approach is a hybrid approach that combines the conventional RS and the connectivity information of a user with other groups. By using this approach author solved the problem of cold start, over specialization and sparsity.

Using General Unexpectedness by Questionnaires

In this study Akiyama et. al. finds serendipitous item after creating "concept" about the item. In this research, the concept is created based on the questionnaires [2]. The questionnaires are created based on the actual data and baseline assumptions.

Using Content Based Recommender System

Iaquinta et. al. finds the serendipity in digital library [8]. They uses the strategy discovered and used by Toms and Elaine G. to find serendipitous recommendation [24] as below:

1. Role of chance or 'blind luck', which authors implemented via a random information node generator.

2. Pasteur principle ("chance favors the prepared mind"), implemented via a user profile.

3. Anomalies and exceptions, partially implemented via poor similarity measures.

4. Reasoning by analogy.

In this work they propose an architecture for content based RSs that implements the third option described as above to provide serendipitous recommendations.

V. MEASURING SERENDIPITY

In our research, we found following different ways to measure the Serendipity that depends on types of dataset. The serendipity measures totally depends in its context. Apart from Accuracy, the different metrics used, defined and proposed by different authors are usefulness, unexpectedness, novelty, coverage and serendipity.

Unexpectedness and Serendipity

"Unexpectedness" means something hard to predict. To unexpectedness can be generated by unlikely combination of features [2]. Murkani et. al. and Mouzi Ge at. al. proposed approach to measure unexpectedness [12, 13]. They defines an unexpected set of recommendations UNEXP as:

$$UNEXP = RS \setminus PM \qquad (1)$$

Where RS means recommendation generated by traditional recommendations and PM means recommendation generated Primitive Prediction Model (PPM).

Here the unexpected items are those items which are which are in RS but does not in PM. Based on this the serendipity is defined as:

$$SRDP = \frac{|\ UNEXP \cap USEFUL\ |}{|\ N\ |} \qquad (2)$$

Where USEFUL = Set of "useful" items

N = Total no. of recommendations

Precision, Recall, F1 and ROC

The important and interesting recommender systems does not depends only the accuracy of recommended items based on user preferences but how much the recommended items are serendipitous [13]. In the field of information retrieval, the Precision and Recall are two metrics that are used to measure the accuracy of recommendation [19]. The precision and recall based on confusion matrix are as:

$$Precision = TP / (TP + FP) \qquad (3)$$

$$Recall = TP / (TP + FN) \qquad (4)$$

Where TP = True Positive, FP = False Positive, FN = False Negative and TN = True Negative. Based on precision and recall the F-Measure [19] can be termed as:

$$F\text{-}Measure = 2 * (Precision * Recall) / (Precision + Recall) \qquad (5)$$

In the term of serendipity focus is always on unexpected items. To fulfil it if we plot the curve of FP to FN then it gives information of not preferred but recommended items. This curve gives information about unexpected recommended items which leads to serendipity [7].

Personalized Ranking

In most of the accuracy based recommender systems try to maximize the Area Under ROC Curve (AUC). Because of this biasing towards preferences unexpected items are not appear in recommendation and in top list also. To solve this problem Qiuxia et. al. proposed a new measure termed as Serendipitous AUC (SAUC) [11].

VI. COMMERCIAL UTILIZATION OF SERENDIPITY

The most of the commercial web portals implemented and utilizes the concepts of serendipity including Google, Facebook, Amazon and eBay [6].

eBay, the online shopping portal with Discover project. It provides serendipity by removing popular items in user's list to recommend interesting and surprised items. The algorithm uses the user's interaction with items i.e. clicks on item, return on the same item, forward that item to his/her friends, text analysis of the item descriptions for finding serendipitous items.

Amazon uses the offline activity of visiting library, weather user open more contents of book for information to find unexpected discovery. It uses Statistically Improbable Phrases (SIPs), means most distinctive contents in the book. It extrapolate the information from discover distinctive contents which leads to additional unexpected contents.

Another services developed by Clever Sense [15] which is acquired by Google. In this service, the user interest and preferences are used to learn the serendipity based on user behavior in popular social networking websites.

To find the interesting people around the user in social network is created by Glancee [3], which is later acquired by Facebook.

VII. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have explored the concept of serendipity in RS in all ways. Serendipity is an important parameter beyond the accuracy of any RS for user satisfaction. Serendipity can occurs when some inspiration metaphor presents. However, it happens accidently. Because of incomplete knowledge, improper combinations and underestimation also leads to the serendipity. To provide serendipitous recommendations low similar items or the items available at long tails should be included in top listing. More measures on serendipity are tested and proposed in future. By adding different types of information available from social networking portals are more promising area in research of serendipitous recommendations.

REFERENCES

[1] Adamopoulos, Panagiotis, and Alexander Tuzhilin (2014). "On unexpectedness in recommender systems: Or how to better expect the unexpected." ACM Transactions on Intelligent Systems and Technology (TIST) 5.4 (2014): 54.
[2] Akiyama, Takayuki, Kiyohiro Obara, and Masaaki Tanizaki (2010). "Proposal and evaluation of serendipitous recommendation method using general unexpectedness." Proc. Workshop on the Practical Use of Recommender Systems, Algorithms and Technologies.
[3] Andrea Vaccari, Alberto Tretti, Gabriel Grise (2010). Glancee Inc., www.glancee.com.
[4] Corneli, J., Pease, A., Colton, S., Jordanous, A., & Guckelsberger, C. (2014). Modelling serendipity in a computational context. arXiv preprint arXiv:1411.0440.
[5] Werner Creixell, Juan Carlos Losada, Tom Arredondo, Patricio Olivares, Rosa Maria Benito (2012). "Serendipity in social networks." NHM 7.3 (2012): 363-371.
[6] Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, Cataldo Musto (2015). "An investigation on the serendipity problem in recommender systems." Information Processing &Management 51.5 (2015): 695-717.
[7] Hisaaki Yamaba, Michihito Tanoue, Kayoko Takatsuka, Naonobu Okazaki, Shigeyuki Tomita (2013). On a Serendipity-oriented Recommender System based on Folksonomy and its Evaluation, Procedia Computer Science, Volume 22, 2013, Pages 276-284, ISSN 1877-0509, http://dx.doi.org/10.1016/j.procs.2013.09.104.
[8] Iaquinta, L., de Gemmis, M., Lops, P., Semeraro, G., Filannino, M., & Molino, P. (2008). Introducing serendipity in a content-based recommender system. In 8th International conference on hybrid intelligent systems (pp. 168–173). IEEE Computer Society.
[9] Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. Cognitive Computation 4(3):246–279.
[10] José Campos, A. Dias de Figueiredo (2002), Programming for Serendipity, AAAI Technical Report FS-02-01, American Association for Artificial Intelligence (www.aaai.org)
[11] Qiuxia Lu , Tianqi Chen ; Weinan Zhang ; Diyi Yang ; Yong Yu (2012). "Serendipitous Personalized Ranking for Top-N Recommendation." Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01. IEEE Computer Society.
[12] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In Proceedings of the fourth ACM conference on Recommender systems. ACM, New York, NY, USA, 257-260. DOI=10.1145/1864708.1864761 http://doi.acm.org/10.1145/1864708.1864761
[13] Murakami, Tomoko, Koichiro Mori, and Ryohei Orihara (2008). Metrics for evaluating the serendipity of recommendation lists. New frontiers in artificial intelligence. Springer Berlin Heidelberg, 2008. 40-46.
[14] Nakashima, Toshiharu, Kenji Nakamura, and Shigeru Oyanagi (2014). Serendipitous Recommendation Method based on WWW Access Log. Proceedings of the International MultiConference of Engineers and Computer Scientists. Vol. 1.
[15] Nima Asgharbeygi, Babak Pahlavan (2011). Clever Sense offers "Alfred," a mobile app delivering personalized recommendations for nearby restaurants, coffee shops, and nightclubs, http://www.thecleversense.com
[16] Oku, K., Hattori, F. (2013). Fusion-based recommender system for serendipity-oriented recommendations. Journal of Japan Society for Fuzzy Theory and Intelligent Informatics 25.1 (2013): 524-539.
[17] Oku, Kenta, and Fumio Hattori (2011). Fusion-based recommender system for improving serendipity. Proceedings of the Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), at the 5th ACM International Conference on Recommender Systems.
[18] Oku, Kenta, and Fumio Hattori (2012). User Evaluation of Fusion-based Approach for Serendipity-oriented Recommender System. Workshop on Recommendation Utility Evaluation: Beyond RMSE (RUE 2011).
[19] Passant, Alexander; Mulvany, Ian; Mika, Peter; Maisonneauve, Nicholas; Löser, Alexander; Cattuto, Ciro; Bizer, Chris; Bauckhage, Christian and Alani, Harith (2008). Mining for Social Serendipity. In: Seminar on Social Web Communities, 21-26 September 2008, Dagstuhl.

[20] Pease, Alison, Simon Colton, Ramin Ramezani, John Charnley and Kate Reed (2013). A discussion on serendipity in creative systems.Proceedings of the Fourth International Conference on Computational Creativity.
[21] Rajani, N. F. Exploring Accuracy with Serendipitous Discoveries in Recommender Systems.
[22] Sean M. McNee, John Riedl, Joseph A. Konstan (2006). Accurate is not always good: How Accuracy Metrics have hurt Recommender Systems, Extended Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems.
[23] Sun, Tao, Ming Zhang, and Qiaozhu Mei.(2013). Unexpected Relevance: An Empirical Study of Serendipity in Retweets. ICWSM. 2013.
[24] Toms, Elaine G. (2000). Serendipitous Information Retrieval. DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries.
[25] Yu-Shian Chiu, Kuei-Hong Lin, Jia-Sin Chen (2011). A Social Network-based Serendipity Recommender System, International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)

AUTHOR PROFILE

Mr. Ashishkumar B. Patel is working as an Assistant Professor at LDRP Institute of Technology and Research, a constituent institute under KadiSarvaVishwavidyalaya from last 11 years. His area of research is Data Mining in which he is pursue Ph.D. in Recommender System at C U Shah University, Wadhwan, Gujarat.

Dr. Kiran Amin is working as a Professor and In-charge Principal at U V Patel College of Engineering of Ganpat University, Kherva, Gujarat. Dr. Amin has more than 20 Years of experience in Academic. His area of research are Data Mining, Artificial Intelligence and Computer Networks. In his academic career he published more than 20 papers in reputed International Journals/Conferences like IEEE, Springer and ACM.