MapRat: Meaningful Explanation, Interactive Exploration and Geo-Visualization of Collaborative Ratings *

Saravanan Thirumuruganathan‡, Mahashweta Das‡, Shrikant Desai‡,
Sihem Amer-Yahia†, Gautam Das‡,†, Cong Yu††
‡University of Texas at Arlington; †Qatar Computing Research Institute; ††Google Research
‡{ saravanan.thirumuruganathan@mavs, mahashweta.das@mavs, shrikant.desai@mavs, gdas@cse}.uta.edu, † {syahia, gdas}@qf.org.qa, ††congyu@google.com

ABSTRACT

Collaborative rating sites such as IMDB and Yelp have become rich resources that users consult to form judgments about and choose from among competing items. Most of these sites either provide a plethora of information for users to interpret all by themselves or a simple overall aggregate information. Such aggregates (e.g., average rating over all users who have rated an item, aggregates along pre-defined dimensions, etc.) can not help a user quickly decide the desirability of an item. In this paper, we build a system MapRat that allows a user to explore multiple carefully chosen aggregate analytic details over a set of user demographics that meaningfully explain the ratings associated with item(s) of interest. MapRat allows a user to systematically explore, visualize and understand user rating patterns of input item(s) so as to make an informed decision quickly. In the demo, participants are invited to explore collaborative movie ratings for popular movies.

1. INTRODUCTION

Collaborative rating sites such as IMDB[1], Yelp[2], and Amazon[3] have become an integral part of how users make informed decisions about items. Such sites consist of items (e.g., movies, restaurants, e-commerce products, etc.) and an active community that provides feedback in the form of ratings, reviews, tags, etc. For example, a bakery in San Francisco, *Golden Gate Bakery* received more than thousand ratings, and popular restaurants routinely exceed that number by many factors. This huge explosion in user feedback causes a significant cognitive overload of information on users, who usually want to quickly form a judgment without investing a lot of time. Many collaborative sites today provide some sort of filters to reduce information overload. For example, Amazon shows the distribution of user star ratings. However, the filters are either too general to convey meaningful information or too detailed so that users feel overwhelmed with information. Some movie rating sites like IMDB provide aggregates over a pre-defined set of user demographics (gender, age group and country). Though this is an improvement over a single aggregate value or rating distribution, these pre-defined aggregates do not necessarily provide information ideal for decision making.

In other words, collaborative sites do not address the difficulty users encounter when deciding whether an item is desirable and a user is typically left on her own to make the best use of available information: she either trust an overall aggregate that distills thousands of rating to a single numeric value or spend valuable time examining individual reviews. We envision that appropriate leveraging of the rich meta data associated with users and items in collaborative sites can be useful in explaining the item ratings to the users. Collaborative rating sites typically contain user profile information about the reviewers (or, raters) of items. The same is however not true for most of its users. The absence of profile information and explicit preferences of users, who visit the sites for building their opinion, makes it difficult to provide effective assistance to the users. A rational approach is to use a neighbor style explanation [1] that utilizes reviewer profile information to generate explanations, while allowing a user to select the best explanation herself. Given the lack of user context and information, generating explanations based on how different reviewer sub-populations rated an item seems the most reasonable approach to us.

There exist prior systems such as OIC Weave[4] that can provide visualization of ratings along different demographic attributes. However, such systems do not provide any automatic and interactive exploration of the rating information. Though Weave allows exploration along a single dimension, it fails to identify more granular reviewer sub-populations that can potentially have interesting rating patterns.

In [2], we introduced a framework that leverages meta data associated with ratings to automatically provide meaningful interpretations of ratings associated with input item(s). It quickly identifies a small set of good groups (i.e., user sub-populations) based on user demographics (*age, gender, location* and *occupation*) that succinctly explains user rating patterns for input item(s). In this work, we go beyond [2] and build a system *MapRat* that not only helps users make better decisions by providing meaningful explanations of item ratings in collaborative rating sites, but also supports interactive explorations and appealing geo-visualization of the explanations retrieved. We focus on two main tasks: *Similarity Mining*

*The work of Saravanan Thirumuruganathan, Mahashweta Das, Shrikant Desai and Gautam Das is partially supported by NSF grants 0812601, 0915834, 1018865, a NHARP grant from the Texas Higher Education Coordinating Board, and grants from Microsoft Research and Nokia Research.

[1]`http://www.imdb.com/`

[2]`http://www.yelp.com/`

[3]`http://www.amazon.com/`

[4]`http://oicweave.org/`

Proceedings of the VLDB Endowment, Vol. 5, No. 12

(SM) which identifies groups of reviewer sharing similar ratings on item(s) and *Diversity Mining* (DM) which identifies groups of reviewer sharing dissimilar ratings on item(s).

Consider the movie *The Twilight Saga : Eclipse* which has close to $100,000$ ratings in IMDB. Though the average rating of all reviewers is 4.8 on a scale of 10, we find that *female reviewers under 18* and *female reviewers above 45* love the movie and give very high ratings (SM). Again, *male reviewers under 18* and *female reviewers under 18* consistently disagree on their ratings for the movie: the former group hates it while the latter loves it (DM). Thus DM allows users to understand the rating patterns for controversial items. Though all the examples so far involve a single item, both the mining tasks can be applied to a set of items with some common features, such as *all movies directed by Woody Allen*.

Our system MapRat produces meaningful explanations of rating patterns for input item(s) and then visualizes the interesting results over a map. MapRat supports sophisticated mining of item ratings by allowing an user to explore the rating patterns across three dimensions : *user demographics, geographic location* and *time*. Rating explanations based on user demographics allow a user to quickly decide if she likes the item by considering the aggregate rating of the result group with which she identifies the most; exploration along one or more user dimensions results in explanations which serves a user's personalized needs. Our system heavily utilizes the geographical aspect of reviewer sub-population since it provides a convenient anchor on which the explanations can be visualized on; exploration along geographic location allows identification and visualization of rating trends over a map. While geographical location is our primary dimension for visualization, MapRat also provides convenient mechanisms for users to explore along other dimensions. Exploration along time facilitates understanding of the evolution of rating patterns over time.

The main technical challenge in achieving the objective of our system is how to efficiently select user groups that best explains the ratings for input item(s) from thousands of potential candidates. The amount of available user feedback in terms of ratings is humongous and the Similarity Mining and Diversity Mining problems are proved to be NP-hard in [2]. Visualization of the rating interpretations to cater to users' cognitive needs and aid further interactive explorations posses additional challenges.

2. MAPRAT DESIGN

We introduce the data model and the architecture, as well as provide a brief discussion of our mining tasks and approaches.

2.1 Data Model

The data model of MapRat largely adheres to that described in [2]. A collaborative rating site $\mathcal{D}$ is modeled as a triple $\langle \mathcal{I}, \mathcal{U}, \mathcal{R} \rangle$, representing the sets of items, reviewers and ratings respectively. Each rating $r \in \mathcal{R}$ is itself a triple $\langle i, u, s \rangle$ where $i \in \mathcal{I}, u \in \mathcal{U}$, and $s \in [1, 5]$ is the integer rating that reviewer u has assigned to item i. Additionally, both $\mathcal{U}$ and $\mathcal{I}$ are associated with attributes denoted by $\mathcal{U}_\mathcal{A}$ and $\mathcal{I}_\mathcal{A}$ respectively. The review attributes are typically *Age, Gender, Occupation, ZipCode*. Item attributes depend upon the collaborative rating site. For example, item attributes for movies in IMDB can be *Title, Genre, Actor* and *Director*.

The notion of *group* is defined based on data cube [3]. Informally, a group is the set of rating tuples describable by a set of attribute value pairs belonging to reviewers, items or both. Since in our system, we intend to interpret ratings using reviewer attributes, groups are defined using a set of attribute value pairs describing reviewers. Intuitively, a group is a set of ratings that can described using a subset of reviewer attributes. For example, the group $\{\langle \texttt{location}, \texttt{ca} \rangle, \langle \texttt{occupation}, \texttt{student} \rangle\}$ consists of ratings by student reviewers in California.

2.2 Rating Mining: Similarity and Diversity

The task of meaningful explanation of item ratings boils down to identifying *good* groups. The essential characteristics of a good group are : each group should be easily understandable by a user; the groups should together cover a significant proportion of available ratings; and ratings within each group should be as consistent as possible.

Given an item (or set of items) I, the primary goal of MapRat is to generate meaningful explanations for the ratings R_I associated with I. It is not a single interpretation that is interesting to an end user. Hence, we define a family of interpretations that can generate succinct explanations for the ratings. Each of the interpretation focuses on some aspect of rating behavior in which groups of user agree or disagree over their ratings. The user is provided with rating interpretations through a set of groups each of which are describable using the review attributes and cover a reasonable fraction of associated item ratings.

Similarity Mining (SM) : Given a set of items, this sub-problem generates interpretations by identifying reviewer groups which are describable by their attributes that have very similar ratings for the items. SM is most useful in identifying reviewer preferences. Additionally, a user can choose the reviewer group she most identifies with and choose their aggregate rating. This value has a higher utility to a user than the aggregate for the entire set of reviewers.

Diversity Mining (DM) : Given a set of items, this sub-problem generates interpretations by identifying a set of meaningfully labeled reviewer groups that consistently disagree on these items. DM is most useful in identifying reviewer response towards controversial items.

These two mining tasks collectively retrieve meaningful explanations of user rating behaviors that cover the most frequently used notions of interestingness. Each of the sub-problems is modeled as an optimization problem. We include constraints that ensure that each of the returned groups are meaningfully labeled and collectively cover a significant fraction of ratings. Additionally, we limit the number of such chosen groups to be small enough, not to overwhelm a user. The objective function depends upon the individual sub-problems and the optimization problems are solved using Randomized Hill Exploration (RHE) algorithm [2].

2.3 Architecture

There are two major components in the MapRat system : Rating Mining and Visualization.

Rating Mining : This module accepts a set of items I from the front-end and collects all the corresponding rating tuples R_I. The set of groups that has at least one rating tuple in R_I are then constructed. The next step is to cast the problem as an optimization task corresponding to each of the two sub-problems : Similarity Mining and Diversity Mining. For each of the two sub-problems, the RHE algorithm is employed to retrieve the best set of reviewer groups that provide meaningful rating interpretations. Besides returning explanations, our system also provides visualization of the review groups. The location of the reviewer is a convenient and natural attribute to anchor the visualization. Such location based visualization allows for rapid scanning of the explanations, highlight geographical trends in rating patterns (if any) and also provides a mechanism to overlay explanations from different interpretations. Using a combination of aggressive data pre-processing, result pre-

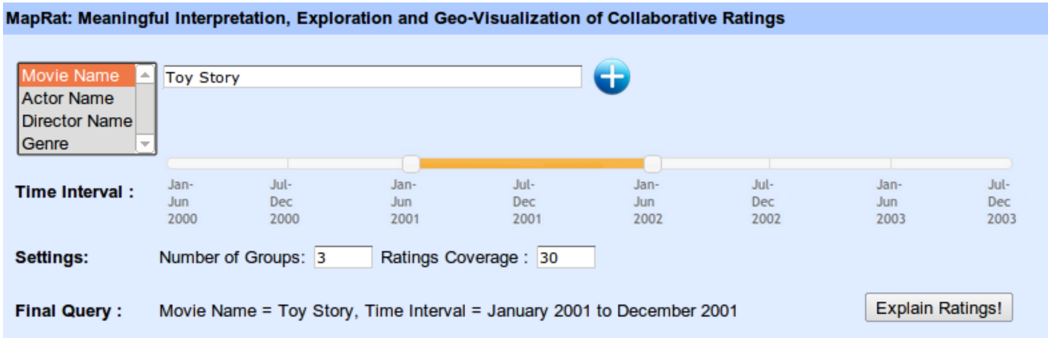

Figure 1: Primary User Interface of MapRat.

computation and caching techniques, the latency of MapRat is minimized.

Visualization : This module is responsible for displaying the rating interpretations over a map such that a user can get a fast overview of the rating trends over geographic regions. Each of the group always specify a geo-condition and hence it is always visualizable on the map. The set of groups that are generated from each of the sub-problems (SM and DM) is considered as rating interpretation *object*. Each set of such objects are then rendered as a Choropleth map [4] using the average group rating for shading. Dark red corresponds to lowest rating while dark green denotes the highest and the intermediate values are represented by the red-green gradient. Each group is also annotated with icons that identify the attribute value pairs used to define it. The set of these Choropleth maps form an *exploration*. Such an exploration is formed from the same set of input rating tuples R_I and constraints, but provide different perspective in terms of meaningful rating interpretations. Collectively, the two different visualizations provide a comprehensive insight into reviewer rating patterns. In addition, the system also allows a user to drill deeper and view lower level aggregate statistics. For example, if the original geo condition was over a state, the drill down provides city level statistics. Finally, navigation over time dimension allows a user to understand the evolution of the reviewer rating pattern over a period of time.

3. USER INTERFACE AND DEMO

The MapRat system can work on any collaborative rating site that provides data as descried in Section 2.1. For the purpose of the demo, we use Million rating data set from MovieLens [5]. It contains around one Million ratings over 3900 movies by 6040 MovieLens users. The set of user attributes $\mathcal{U}_{\mathcal{A}}$ consists of age, gender, occupation and zip-code. The set of item attributes $\mathcal{I}_{\mathcal{A}}$ consist of the movie title and genre. We integrate the MovieLens data with information available from IMDB, in order to include additional item attributes such as actors and directors.

3.1 User Interface

The MapRat system consists of a web based front-end that allows a user to enter one or more items. The primary UI for entering is shown in Figure 1. A user can enter a conjunctive or disjunctive query by entering one or more attribute value pairs. Possible attributes include movie title, actor, director and genre. Furthermore, the user can restrict the mining over a specific time interval, so that the evolution of rating behaviors over a period can be observed. The user can enter additional search settings such as the maximum number of groups to be returned and its rating coverage. Suppose in Figure 1, a user wants to interpret how the reviewer ratings have evolved over the years for the movie *Toy Story*. For this, the user enters the search query "Toy Story" and sets the type of the query to Movie Name. Once the additional search settings have been entered, the user clicks on Explain Ratings and fetches the results. Moving the time slider over the range of values allows the user to observe reviewer groups that provide best interpretations for the movie and how they change over time.

The result of such a query is shown in Figure 2. Groups from different sub-problems (Similarity Mining and Diversity Mining) are visualized in two different tabs. For this demo, each of the groups always specify the state as their geo condition in order to allow rendering of the explanation in the map. The average rating of the group is used for highlighting the state. We use a red (rating 1.0) to green (rating 5.0) Likert Scale for depicting the average rating. The other reviewer attributes associated with the group are highlighted through icons as a visual aid to the user. The color of the pin holding the icons depicts the age group of the sub-population.

For example, Figure 2 shows the best three groups for Similarity Mining : *male reviewers from California*, *male reviewers from Massachusetts* and *female teen student reviewers from New York*. In this particular instance, the displayed groups neatly correspond to the major market segments of animation movie box office - male and young movie goers. However, it must be noted that MapRat strives to highlight representative groups that the user can self-identify with. Explaining why the chosen groups exhibited such rating behavior is significantly more complex and is not the primary aim of MapRat. All three groups have rated the movie positively as indicated by the color used for highlighting the respective states; the average rating by female teen student reviewers from New York is however lower than those by the remaining groups. These groups consists of reviewers whose individual rating is very closer to the group average and also cover a reasonable fraction of rating tuples. Clicking on any of the groups displays additional statistics about the group's rating. Figure 3 shows the statistical details that are shown to the user when she clicks on the result *Male reviewers from California* for further exploration. This provides a convenient way to compare the rating patterns of related groups. It is also possible to drill down and view the city level aggregate movie rating statistics for each of the groups during such interactive exploration. Finally, MapRat can exploit any user demographic information (gender, age, location or occupation) available to constrain the groups that are highlighted. This ensures that the resulting groups are the ones that user most self-identifies with and hence most relevant for decision making.

[5] `http://www.grouplens.org/node/73`

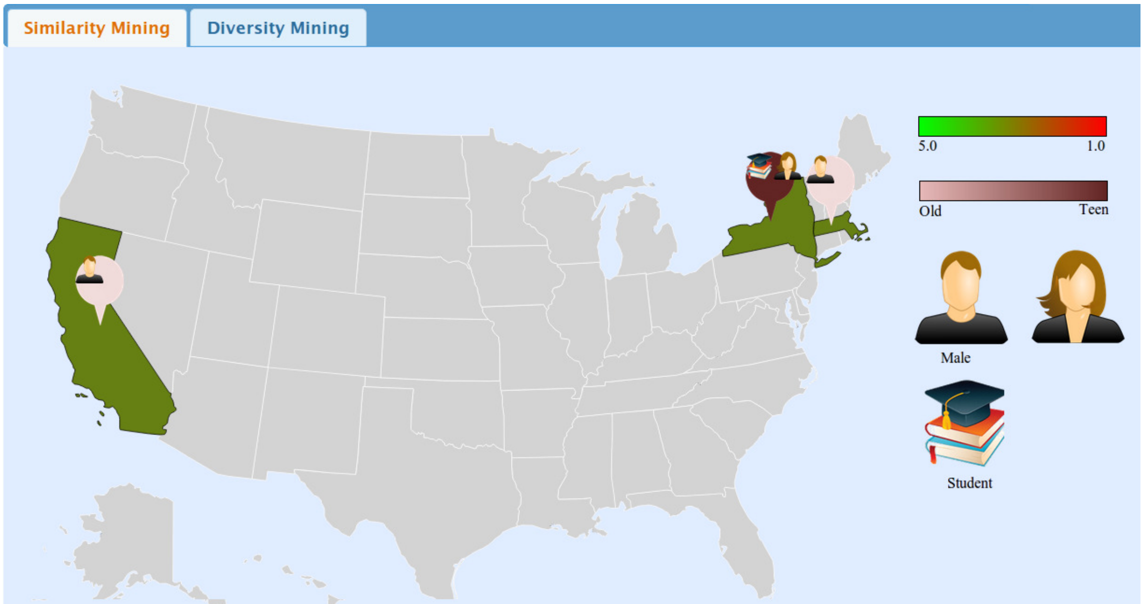

Figure 2: MapRat Explanation Result for Query in Figure 1.

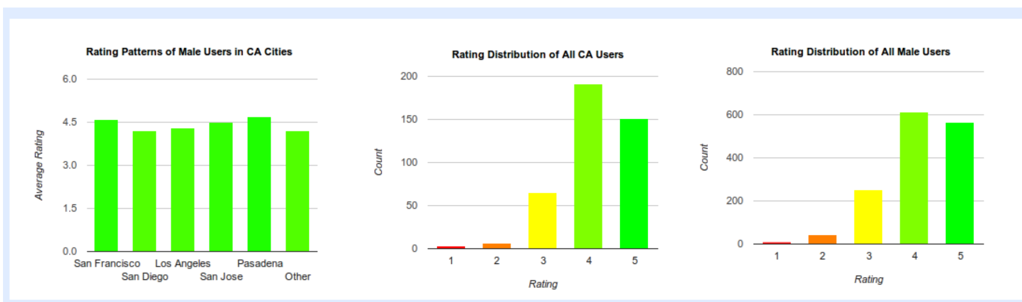

Figure 3: MapRat Exploration Result for Explanation *Male reviewers from California*.

3.2 Demonstration Plan

Our demo allows the audience to use a web interface (as shown in Figure 1) and specify arbitrary search query involving one or more movie attributes. Example queries include *The Social Network*, *Tom Hanks*, *The Lord of the Rings film trilogy*, *thriller movies directed by Steven Spielberg* and so on. The audience can specify other search settings and the time interval to restrict the mining.

Based on the query, our system will display the visualizations for the two meaningful rating interpretation problems. The audience can explore the results to have a better understanding of the reviewer rating patterns for the query. They can observe how the rating patterns fluctuate over a period of time or drill down deeper to view the rating statistics at city level. Such exploration will give the audience a deeper appreciation of our system's utility to aid users make informed judgments about movies quickly. It will also clearly show the superiority of our system in describing rating explanations in terms of meaningfully labeled user groups, over existing collaborative rating sites.

4. CONCLUSION

Given an item (or set of items), MapRat generates meaningful explanations of rating behaviors that help users make informed decisions about items. The system visualizes the interpretations on a map and facilitates exploration over different dimensions. The interactive and dynamic aggregate analytics in our system goes beyond the static and often generic aggregate statistics provided by popular collaborative rating sites. Our demo allows users to generate meaningful rating interpretations for popular movies in MovieLens dataset and interactively explore and geo-visualize each of the explanations.

5. REFERENCES

[1] M. Bilgic and R. J. Mooney. Explaining Recommendations: Satisfaction vs. Promotion. *Intelligent User Interfaces*, pages 13–18, 2004.

[2] M. Das, S. Amer-Yahia, G. Das, and C. Yu. MRI: Meaningful Interpretations of Collaborative Ratings. *PVLDB*, 4(11):1063–1074, 2011.

[3] J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh. Data Cube: A Relational Aggregation Operator Generalizing Group-by, Cross-Tab, and Sub Totals. *Data Mining and Knowledge Discovery*, pages 152–159, 1997.

[4] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard. *Thematic Cartography and Geovisualization*. Pearson Education, 3rd edition, 2009.