Journal of Web Engineering, Vol. 16, No. 1&2 (2017) 063–074

ON THE VALUE OF PURPOSE-ORIENTATION AND FOCUS ON LOCALS IN RECOMMENDING LEISURE ACTIVITIES

BEATRICE VALERI and FABIO CASATI
DISI, University of Trento
via Sommarive 9, 38123 Povo, Trento, Italy
valeri@unitn.it, casati@unitn.it

FLORIAN DANIEL
DEIB, Politecnico di Milano
via Ponzio 34/5, 20133 Milano, Italy
florian.daniel@polimi.it

Received June 21, 2016
Revised October 13, 2016

Recommender systems are omnipresent today, especially on the Web, and the quality of their recommendations is crucial for user satisfaction. Unlike most works on the topic, in this article we do not focus on the algorithmic side of the problem (i.e., searching for the algorithm that better learns from the collected user feedback) and instead study the importance of the data in input to the algorithms, identifying the information that should be collected from users to build better recommendations. We study restaurant recommendations for locals and show that fine-tuned data and state-of-the-art algorithms can outperform the leading recommendation service, TripAdvisor. The findings make a case for better-thought and purpose-tailored data collection techniques.

Keywords: Recommender systems, data collection, mobile recommendations, restaurants, TripAdvisor
Communicated by: M. Gaedke & L. Olsina

1. Introduction

Recommender systems are software systems that, given a set of items and a user, aim to predict the user's interest in the items and to suggest the user which items to inspect, use or buy in a given context. Two ingredients are at the core of each recommender system: first, the *algorithms* that select candidate items; second, the *data* that provide the base for the recommendations [1].

Algorithms are the traditional focus of research: starting from a set of numeric ratings that users assign to items, researchers look for prediction models that can provide good recommendations. They can be split into two main classes: collaborative filtering [2] and content-based recommendation [3]. The former is based on how users interact with items (e.g., if they read, rate, comment, like, buy items) and look for *users that behave similarly* to the target user; the latter is based on the descriptions of both items and users (their profile and preferences) and look for *items* that have similar features to the ones the target user already liked in the past. Hybrid techniques bring both approaches together.

Input data for algorithms is often in the form of ratings on items, typically in terms of 5-stars ratings (such as those in the datasets available from *MovieLens* [4] or *Netflix* [4, 5]). Other common rating scales are unary (like), binary (thumbs-up/thumbs-down) and 3-values scales (thumbs-up/neutral/thumbs-down). Most algorithms compute recommendations based on ratings without considering extra information about the experience or context of the user that lead to the judgments, thereby leaving the interpretation of the ratings to the recommender algorithms. The need of more information other than simple user-item ratings is also highlighted by the interest in feature-based [6] and context-aware [7] recommender algorithms.

In this article, we focus on recommendations for leisure activities, and specifically of restaurants. In our previous studies on this topic [8], we noticed that the quality of recommendations on items (in our cases, restaurants and bars) strongly depends on the specific purpose of the activity: one place may be good for dining but not for drinking; another one may be good for a romantic dinner but not for one with friends. This kind of nuances is usually not captured by state-of-the-art recommender systems.

Another important aspect of this domain is that persons (both tourists and locals) looking for places to eat frequently rely on recommendations from *locals* [9]: They see locals as knowledgeable and trustworthy, since they know that locals know most of the available options. This precious knowledge of locals is often lost in online tourist portals, which are mostly oriented toward and, hence visited by, tourists. Good recommendation services that specifically leverage knowledge of locals are still understudied.

These considerations raise the question of whether data collected with i) a focus on locals and ii) information on specific usage purposes in mind can recommend restaurants to locals with higher *precision* (that is, higher probability of recommending a place the user will actually like) than generic recommender systems, including the leading commercial restaurant recommenders many of us use regularly. The objective of this article is to answer this question. Our goal is not so much that of finding the "best" algorithm, but rather to explore the potential of rating datasets considering the purpose and the local origin of the recommendations with respect to general tourist-oriented datasets, and to understand if and how much we can improve on the quality of recommendations. Given the importance of the topic (millions of people use restaurant recommendations systems every day), finding a positive result would be an indication for commercial systems of the importance of focusing on this kind of ratings, and for researchers that researching optimal algorithms would be an activity with a high potential impact.

2. Recommender Systems for Leisure Activities

Collaborative filtering is often successfully applied in the e-commerce sector. However, for the leisure sector, we have to take into consideration specific characteristics such as *context* and *location*: First, only places close to the user can be experienced and people do not usually travel much to find them (within 14 miles from their house [10]). Furthermore, contextual information helps finding more interesting results: Baltrunas et al. [7] show that recommendation systems are able to increment user satisfaction by considering aspects such as weather conditions, companions, time (season, weekday or time of the day) and familiarity with the area. Mobile devices provide support for context-aware recommendation systems.

Thanks to their sensors, they can automatically collect contextual information, such as user position and therefore weather conditions, and also provide for proactive recommendations [11].

The need for support in searching interesting places for leisure activities is widely recognized and many services have been developed with this goal. TripAdvisor and Gogobot are two of the most used recommendation services in the travel sector, while Yelp and Google Local are more focused on locals, providing "yellow pages"-like services. These recommendation systems collect ratings using a 5-star scale or a variation of it, giving users the possibility to express their experience with different nuances of satisfaction/dissatisfaction. They are also very popular: TripAdvisor counts 340 million unique monthly visitors [12], while Yelp counts 142 million unique monthly visitors [13]. The use of mobile devices has been growing steadily: 190 million people downloaded the TripAdvisor mobile app, and 50% of accesses to TripAdvisor and Yelp come from mobile devices (both smartphones and tablets) [12, 13].

Recently, Foursquare added local search functionality to its application [14], using both the check-in information of the location-based social network and user feedbacks, reviews and tastes to provide recommendations. Other services specifically focus on restaurants, such as The Fork (user-provided ratings) and Zagat (expert-provided ratings).

In a recommender system, users express their opinions about items in the form of ratings. *Items* can be anything users can experience and can have an opinion about, while a *user* is any person that has experienced some of the items the system is focused on. In this paper we focus on restaurants, which are physical establishments, so only people able to visit them can also experience them. From this perspective, a restaurant can have two kinds of customers: locals and tourists. *Locals* are people that live in the area, are familiar with the local cuisine and the other restaurants in the area, and can experience them several times. *Tourists* are visitors for business or leisure that generally have fewer chances to sample restaurants in a given area and are less familiar with the local cuisine.

In many systems, when rating items users assign one rating per item, evaluating it according to their overall experience. *Multi-criteria* ratings, instead, ask users to add one rating for each of a set of predefined characteristics of the item. For example, in the case of restaurants, the criteria could be food quality, drink quality, service and popularity. This requires a user to consider the different aspects of her experience and give more ratings.

Orthogonally to this, a restaurant may be perceived differently depending on the *purpose* of the visit: the choice of a restaurant for a dinner with friends may differ from what we would choose for a romantic dinner or for a quick lunch. We already verified the importance of considering purpose in our earlier research [8], where we also identified four main purposes: dinner with tourists, romantic dinner with the partner, dinner with friends and price/quality ratio (e.g., important for a lunch break).

3. Method

In this article we study whether recommendations based on *purpose-specific data* collected from *locals* outperform recommendations computed from the typical data collected from tourists by tourist portals. As representative of tourist portals currently popular we select TripAdvisor: it is very popular worldwide and it is very used in Trento, Italy, the city in which we run our evaluation. TripAdvisor has the advantages of having an almost complete list of restaurants in Trento, being used by many tourists and not having any competitor in the city. The other popular tourist portals have less knowledge of the available options in the area and are not able to collect many information about the preferences of people. On the other hand, TripAdvisor does not disclose any information about its dataset or its algorithm, but this happens also with all the other recommendation services. We can only say that TripAdvisor provides non-personalized recommendations, i.e. a generic rank of restaurants based somehow on an aggregation of all the ratings collected for each restaurant.

3.1. *Data Collection*

In May 2014 we collected ratings for 50 restaurants in Trento, Italy. We selected this list considering the most popular restaurants according to TripAdvisor that are located in the city center and easily reachable by everyone. Such restaurants are a convenient choice as they are almost all the options available in Trento's historic city center, where people mostly spend their leisure time. By choosing these restaurants, we are able to assume that every local in Trento is aware of many of them and experienced at least some of them. We invited locals to participate to our data collection through posters and flyers, we enrolled friends and colleagues by email, and we also involved a small group of university students, therefore implementing a convenience sampling of Trento's local population.

As part of the data collection, participants were asked to share with us some personal information about themselves on a voluntary basis. Of the 114 total participants, 91 answered these questions allowing us to understand how the user base is composed. 70% of the respondents were male; most of them were less than 36 years old (58% were aged between 18 and 25, while 37% were aged between 26 and 35). At the time of the study, most of the respondents had been living in Trento for more than 2 years (58%, including 16 people who have been living in Trento for more than 10 years). The convenience sampling of the population was thus biased towards young men.

Ratings were collected using a 3-values thumb-up/thumbs-down scale for each of the purposes identified in [8] (dinner with tourists, dinner with partner, dinner with friends and lunch break): users could specify whether they like or don't like a restaurant, or are neutral about it. In general, the thumbs-up/thumbs-down scale leaves less space for controversy than using 5 stars: the user just has to think about whether the item is good or bad, without having to think about how good (or how bad). The neutral rating prevents forcing the user to like or dislike an item if it is considered borderline.

The participants were asked to rate restaurants for each of the purposes (the 4 identified above) at a time. The process produced a total of 4706 ratings, with 1529 ratings for "dinner with tourists", 1113 ratings for "dinner with the partner", 1112 ratings for "dinner with friends" and 952 ratings for "price/quality ratio". The restaurants received a minimum of 4 ratings and a maximum of 112 ratings per purpose, while users added a minimum of 0 ratings and a maximum of 49 ratings per purpose, with an average of 11 ratings per purpose.

3.1.1. *Recommendation Algorithms*

Computing purpose-specific ratings poses challenges to the recommendation algorithms, as the algorithm has to work with multiple ratings per item per purpose. A first way to approach this multiplicity is to filter ratings to create one dataset for each purpose; in this way, only the information about the purpose the requester is interested in is used to compute recommendations. Another way is to merge all ratings from the different purposes and to compute aggregated ratings valid for all purposes, similarly to how multi-criteria ratings are handled by recommendation algorithms. A third solution is to learn user tastes using all collected data for all purposes and to compute ratings for each purpose individually; in this way, the whole information is used to extract taste features and to compute similarities between users or items, but only the ratings specific to a purpose for a user in a given instant of time are used for the prediction of ratings for unknown restaurants.

We followed this last solution. To handle the presence of 4 ratings per user-restaurant pair (one for each purpose), we split each user's ratings for a restaurant into 4 purpose-specific restaurant-purpose pair, resulting in 200 (50 * 4) items. In this way, all ratings can be considered in the computation of the model used by the algorithm (like building clusters for cluster-based collaborative filtering or computing matrix factorization for SVD), while only the restaurant-purpose pairs for the requested purpose are considered to build the rank when computing recommendations. To adapt the algorithms to this behavior, we only need to extend them with a final filter of items by purpose.

For computing recommendations we select four state-of-the-art, personalized, collaborative filtering algorithms implemented by Apache's Mahout library (`http://mahout.apache.org`):

- *User-based collaborative filtering* identifies a requester's neighbors (the users with similar tastes) and uses their ratings and the level of similarity with the requester to compute a prediction of the requester's ratings for the items she does not know yet.
- *Cluster-based collaborative filtering* pre-groups users into clusters of users with similar tastes and averages the ratings of all users within each cluster to compute a prediction of the requester's ratings for unknown items. We specifically use hierarchical clustering.
- *Slope One* is an item-based algorithm that leverages on the principle of "popularity differential," that is, on how much one item is liked more than another. In order to predict the rating of an item, it considers information both from other items rated by the requester (and their ratings from other users) and from other users who rated the item (and their ratings to other items) [15].
- *SVD* is a matrix factorization algorithm that computes ratings out of features automatically extracted from a known, incomplete user-item matrix. The matrix is decomposed into a user-feature, a feature-item, and a feature-feature matrix. Rating predictions are computed as the product of the requester's row, the feature-feature matrix, and the item's column.

These algorithms have been selected as they are popular and simple, two properties that allow us to communicate better the effects of the data on recommendation quality. Other algorithms have been shown to perform similarly or even better under certain conditions, but our goal in this article is not that of finding the best algorithm for our specific dataset, and we aim instead at understanding and communicating the potential of adopting of purpose-specific ratings from locals. Since all restaurants in our dataset are easily reachable by foot, user location and time (the usual contextual information) are not needed; we consider instead the purpose the requester is interested in.

3.2. *Quality Metric*

We compare algorithms based on their *precision* (since we don't have full knowledge of the users' interests they may have rated only a subset of the restaurants they actually know we cannot compute meaningful recall values). Given a user u, the *list* of computed recommendations, and the purpose p, we compare the performance of the algorithms using the following *precision* metric (following [16]):

$$Precision(u, p, list) = \frac{||Good(u, p, list)||}{||Good(u, p, list)|| + ||Bad(u, p, list)||} \quad (1)$$

where:

$Good(u, p, list)$ = items in *list* that have been rated positively by user u for purpose p

$Bad(u, p, list)$ = items in *list* that have been rated negatively by user u for purpose p

For the comparison, we split the users' ratings for each purpose p into a training set (the ratings the algorithms can use to build the user profile) and a test set (the ratings used to compute the precision of recommendations) with a 70/30 proportion. We tested only users that had at least 6 ratings per selected purpose, leaving at least 2 ratings for testing (the ceiling of the 30% split), and omitted items the users didn't express any opinion for. Test ratings were randomly collected half from users' positive ratings and half from their neutral or negative ratings (to test good and bad predictions). To make the test independent of the computed split of ratings, each query was repeated with 5 different random splits.

3.3. *Algorithms tuning and configuration*

Given this evaluation strategy, all algorithms underwent a dry run to configure them for best performance: we collected $Np = 5$ recommendations for each purpose from each algorithm and averaged the precision of each list of recommendations (20 per user: 4 purposes by 5 training/test splits). For simplicity, all tests were run on the full dataset available, causing a possible overfit of the parameters on the data. Despite this possible overfitting, we can still see the effect of data on their performance (that is exactly our main focus) and we can justify the possible bias with the availability of better-quality versions of these algorithms that can obtain the same (or better) results without overfitting parameters. An important point, to which we will come back later in the paper, is that *Slope One* has no parameters and does not require any tuning, so it was used as is, providing bias-free results.

User-based collaborative filtering depends on the used similarity metric, neighborhood strategy and neighborhood size. We tested Pearson correlation, log likelihood, Spearman correlation, Tanimoto coefficient, cosine similarity, Euclidean-distance-based similarity and Yule similarity. The best precision was obtained with neighborhood selected by similarity threshold, using Yule similarity and similarity threshold 0.3, with a precision of 76%. For *cluster-based collaborative filtering*, we used the same similarity metrics as for user-based collaborative filtering and identified the best configuration in log likelihood similarity and stopping condition expressed as fixed number of clusters, set to 3, with a precision of 70%. The best precision with *SVD* was obtained with 10 features and 30 iterations, with 65% of precision.

4. Results

4.1. *Aggregate precision*

TripAdvisor provides non-personalized recommendations. For this reason, for the comparison we can consider the top Np restaurants in the order proposed by TripAdvisor as the ones recommended to each user. We vary Np from 2 to 15 to study the effect of the recommendation set size on precision and compute the precision of our recommender algorithms by averaging the results of all the purposes over 570 individual data points (114 users times 5 random splits per run) per purpose. We explicitly consider also low values of Np because the typical use case is that of searching for restaurants on a smartphone while on-the-go, and hence often with limited time and screen real estate.

For a first assessment of the difference between the dataset underlying TripAdvisor and our own dataset, we compare the recommendations of TripAdvisor with a similar non-personalized, average-based recommendation algorithm using our dataset. This baseline algorithm computes the predictions of user ratings by computing the "lower bound of Wilson's score confidence interval" (`http://www.evanmiller.org/how-not-to-sort-by-average-rating.html`). This formula computes a confidence interval for the average rating we would obtain if we had all ratings by the full population, starting from a sample of ratings. The lower bound tells "the item is liked at least that much." As TripAdvisor algorithm is not publicly available, our baseline is the best simulation of TripAdvisor-like algorithm on our dataset.

Our own dataset differs from TripAdvisor's one in four key aspects: (i) 3-value vs. 5-value rating scales, (ii) purpose-based vs. generic ratings, (iii) locals vs. tourists, and (iv) small amount vs. large amount of ratings. Since we don't have access to the actual dataset and algorithm used by TripAdvisor, we cannot distinguish the effects of each of these aspects, but we can still see in Figure 1 how TripAdvisor generally produces better recommendations than the baseline (except for $Np = 5$), i.e., the TripAdvisor-like algorithm run on our data. The key to this better performance most likely lies in the bigger amount of ratings TripAdvisor can rely on.

Interestingly, if we now look at the precision of the personalized algorithms, we see that they *all* perform better than both TripAdvisor and the baseline. Slope One and User-based have the best precision and are very close to each other. Cluster-based is not far from the top recommenders, with only a distance of 2 percentage points in precision for higher Np, while SVD performs worse. TripAdvisor's precision is highest for $Np = 15$, where it reaches the same precision of SVD, while it still is 10 percentage points lower than the best performance. This shows that as the size of the recommendation set grows, TripAdvisor has higher probability to contain good recommendations. In order to assess the expressive power of the charts in Figure 1, we took the precision values for $Np = 15$ and performed pair-wise t-tests. The tests confirm also statistically what is communicated by the chart visually: except for User-based/Slope one and TripAdvisor/SVD, all precision values are significantly different (p-value < 0.0001, α-level $= 0.05$, considering the precision of 1280 recommendation lists for each algorithm).

Many of these results are not surprising: algorithms were trained on this dataset, so the comparison with TripAdvisor is not entirely fair. What is however interesting is that Slope One is not trained on our dataset, being the only one without parameters, and the comparison here is indeed fair. Despite this, Slope One produces recommendations of quality very close,

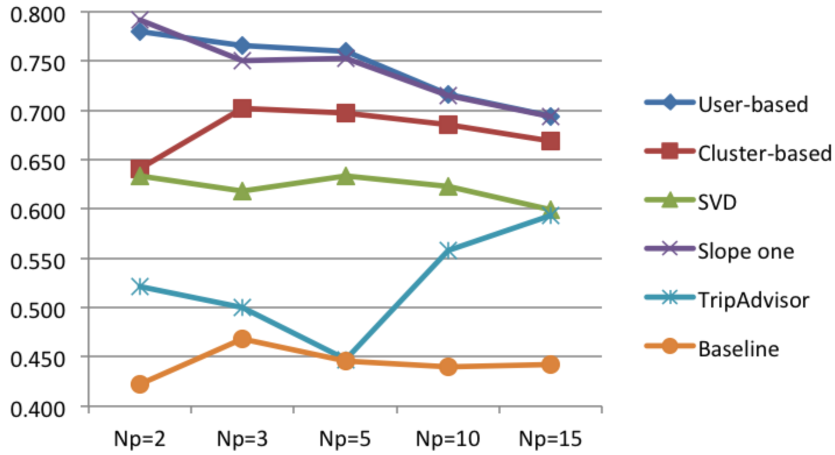

Fig. 1. Precision of the recommendation algorithms for varying result set sizes Np.

and in some cases better, than the best recommender algorithm considered, i.e., user-based collaborative filtering, even if TripAdvisor has a better knowledge of people tastes, building recommendations using a dataset much bigger than the one we collected.

Overall, Figure 1 shows that the precision of the best algorithm between the chosen personalized algorithms (user-based collaborative filtering) is from 10 ($Np = 15$) to 31 ($Np = 5$) percentage points higher than that of TripAdvisor (from 17% to 68% in relative terms).

This means that even though our dataset is significantly smaller than that of TripAdvisor, the focus on locals and personalization yield recommendations that are of significantly higher quality compared to recommendations computed with a generic algorithm from a much larger dataset. TripAdvisor's restaurant rank is in fact built using a huge amount of reviews mostly by tourists and specifically focuses on recommending restaurants to tourists. Our experiment aims to understand how to recommend restaurants to locals and shows that locals are a special class of users that are simply more demanding than generic tourists.

We have to keep in mind that these results have been obtained by averaging the precision of purpose-based recommendations. TripAdvisor starts with a disadvantage since it is built for tourists, and its recommendations could be worse for other purposes (as we will see next).

4.2. *Purpose-specific Precision*

We now analyze the importance of purpose-specific ratings in recommending restaurants. In [8] we found that the restaurants perceived as good for bringing a tourist are similar to those for a romantic dinner with the partner, while the ones for going out with friends are very different and more related to the price/quality ratio. Next, we analyze concretely how the different recommenders behave depending on the purpose a user has in mind. The test setting of the experiments is the same as above, with the only difference that now we no longer aggregate results and instead keep purposes separated.

Figure 2 reports the precision graphs for each purpose. If we concentrate on TripAdvisor,

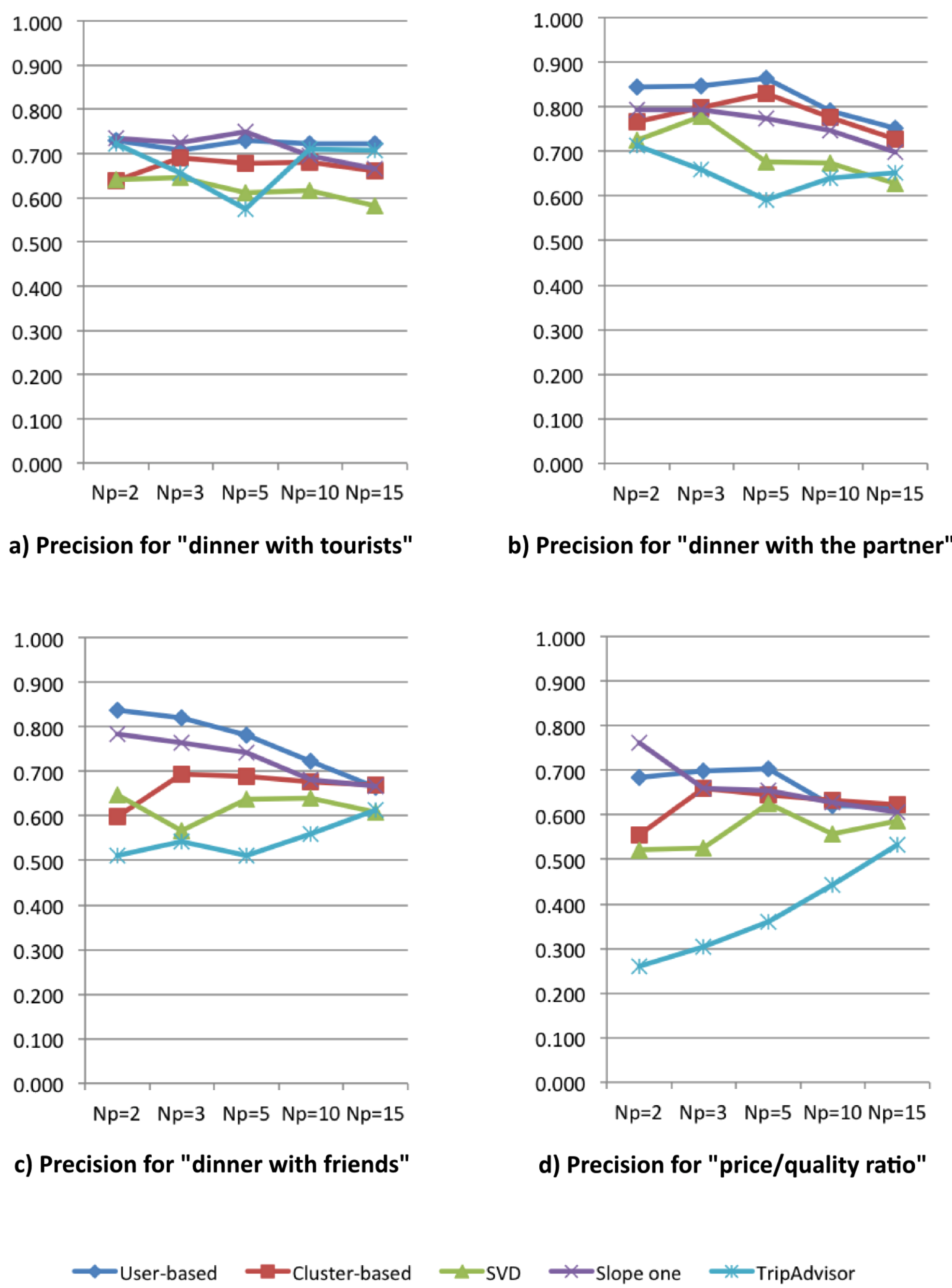

Fig. 2. Purpose-based precision for the five recommendation algorithms.

we see that it provides good predictions for a dinner with tourists, while its precision decreases if the meal is to be consumed with the partner or friends, and it reaches its lowest value if a good price/quality ratio is the target (only 26% of precision for $Np = 2$). The personalized algorithms seem less affected by the purpose, with slightly higher precision for a dinner with the partner and slightly lower precision for the price/quality ratio. Slope One, User-based and Cluster-based collaborative filtering always outperform SVD.

These results clearly indicate that each purpose is different from the others, and algorithms that take care of these differences are able to build better recommendations than generic algorithms. TripAdvisor shows the best precision for $Np = 2$ and “dinner with tourists”, while the worst precision is obtained for $Np = 2$ and “price/quality ratio”, with a difference of 46 absolute percentage points. Purpose-based recommender algorithms have a more constant quality, with less difference between the best and the worst precision: for example, user-based collaborative filtering has the highest precision for $Np = 5$ and “dinner with the partner”, while the lowest one is for $Np = 15$ and “price/quality ratio”, with a difference of 25 absolute percentage points. This minor difference demonstrates a higher quality of purpose-based, personalized recommendations under all circumstances. This let us conclude that *recommendations computed from purpose-specific data outperform TripAdvisor* for the purposes dinner with partner, dinner with friends and price/quality ratio, and may represent a strategic value for competitors of TripAdvisor that want to target locals instead of generic tourists.

TripAdvisor recommendations have instead a high precision for a dinner with tourists, and for this purpose their quality is in line with the ones computed with personalized recommendation algorithms. Given that these latter algorithms use data that stem from locals, this means that locals essentially agree with TripAdvisor on where to bring a tourist and where not. This, in turn, is a quality certificate for TripAdvisor for this specific purpose.

5. Limitations and Conclusion

Our experiments show that providing locals with restaurant recommendations is a tricky endeavor, because providing them with added value – compared to generic tourist portals – asks for advanced personalization, not only based on identity but also on purpose. Purpose is not generically available in recommender systems’ datasets and cannot be extracted by the usual ratings collected. The importance of purpose identified in [8] and in this paper shows how important data collection is: there is a need for better understanding of which information should be collected as user feedback to better learn which experience the users had in a restaurant and better predict how much the other users could enjoy the same restaurant. The experiments further show that if data are collected from/for locals, even basic algorithms outperform generic recommendations. The improvement in recommendation quality thanks to tailored data is not only significant, but has a big effect size. These results are somewhat surprising, given that also more advanced and precise algorithms are available in the literature. What we therefore take home from these experiments is the high potential for considering purpose and origin of ratings in a hugely important area that potentially impacts all of us - outperforming what is by far the leading commercial solution in this space. The results however also show that TripAdvisor is still competitive in its own domain, i.e., recommendations for tourists.

The results of our experiments also reveal another, slightly hidden message that is of particular importance to the world of mobile recommender systems: Mobile devices have typically small screens and are often used in situations in which the user cannot pay full attention to the device. This means that the user can see only few recommended items at a time and may not be willing or able to go through a long list of recommendations [17]. A mobile recommender system is thus particularly challenged even more than a desktop one to compute precise recommendations. The data in Figure 2 show that TripAdvisor performs particularly weakly for small result sets. The lesson is that simply porting a desktop version of a recommendation algorithm to a mobile recommender system may be dangerous, and personalization and data quality become even more important.

Despite the potential, the studies presented here have several limiations. First of all, we have not separated the impact of the various independent variables (such as amount of ratings, rating scale, rating origin and purpose) on the effect we measure. However, we now know that this is an area worth exploring so that further studies are justified. Furthermore, as we mentioned earlier in the paper, the comparison results are fair only for Slope One. For the other algorithms they are also promising but further tests on different datasets (and possibly other algorithms) to determine what works best are needed. As discussed, we tested the algorithms on locals. Therefore the results are applicable to recommendations from locals to locals. While we expect that recommendations from locals may also benefit tourists, the results of this study cannot be generalized in this direction. Finally, another limitation of the study is that our comparison of algorithms is based on the externally visible behavior of TripAdvisor. Its actual, internal algorithm and dataset are not made publicly available and we were only able to run a TripAdvisor-like algorithm directly on our data, i.e., our baseline that is an average-based recommender algorithm. Despite these limitations, we believe this article uncovers an untapped potential to improve our choices of where we spend our leisure time, also laying the directions for future research in this area.

Acknowledgements

This work was supported by TrentoRISE and by the project "Evaluation and enhancement of social, economic and emotional wellbeing of older adults" under the agreement no. 14.Z50.310029, Tomsk Polytechnic University.

References

1. G. Adomavicius, A. Tuzhilin (2005), *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,* TKDD
2. P. Lops, M. De Gemmis and G. Semeraro (2011), *Content-based recommender systems: State of the art and trends,* Recommender Systems Handbook
3. Y. Koren and R. Bell (2011), *Advances in collaborative filtering,* Recommender Systems Handbook
4. F. Cacheda, V. Carneiro, D. Fernndez, and V. Formoso (2011), *Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems,* TWEB
5. J. Lee, M. Sun, and G. Lebanon (2012), *A comparative study of collaborative filtering algorithms,* arXiv preprint arXiv:1205.3193.
6. E. H. Han and G. Karypis (2005), *Feature-based recommendation system,* CIKM
7. L. Baltrunas, B. Ludwig, S. Peer, and F. Ricci (2011), *Context-aware places of interest recommendations for mobile users,* Design, User Experience, and Usability. Theory, Methods, Tools and Practice

8. B. Valeri, M. Baez and F. Casati (2013), *Come Along: understanding and motivating participation to social leisure activities,* CG
9. P. Rompf, R. B. Dipietro, and P. Ricci (2005), *Locals' involvement in travelers' informational search and venue decision strategies while at destination,* J. Travel Tour. Mark.
10. T. Horozov, N. Narasimhan and V. Vasudevan (2006), *Using location for personalized POI recommendations in mobile environments,* SAINT
11. D. G. Vico, W. Woerndl, and R. Bader (2011), *A study on proactive delivery of restaurant recommendations for android smartphones*, RecSys Workshop on Personalization in Mobile Applications
12. *TripAdvisor Fact Sheet,* http://www.tripadvisor.com/PressCenter-c4-Fact_Sheet.html
13. *Yelp Fact Sheet,* http://www.yelp.com/factsheet
14. J. LinkLater (2014), *Foursquare: The New And Improved Yelp,* blog.sweetiq.com
15. D. Lemire and A. Maclachlan (2005), *Slope One Predictors for Online Rating-Based Collaborative Filtering,* SDM
16. G. Groh and C. Ehmig (2007), *Recommendations in taste related domains: collaborative filtering vs. social filtering,* GROUP
17. F. Ricci (2010), *Mobile recommender systems,* Inf. Technol. Tour.