Towards a Recommender Engine for Personalized Visualizations

Belgin Mutlu, Eduardo Veas, Christoph Trattner, and Vedran Sabol

Know-Center
Inffeldgasse 16, 8010 Graz, Austria
{bmutlu,eveas,ctrattner,vsabol}@know-center.at

Abstract. Visualizations have a distinctive advantage when dealing with the information overload problem: being grounded in basic visual cognition, many people understand visualizations. However, when it comes to creating them, it requires specific expertise of the domain and underlying data to determine the right representation. Although there are rules that help generate them, the results are too broad as these methods hardly account for varying user preferences. To tackle this issue, we propose a novel recommender system that suggests visualizations based on (i) a set of visual cognition rules and (ii) user preferences collected in Amazon-Mechanical Turk. The main contribution of this paper is the introduction and the evaluation of a novel approach called *VizRec* that is able suggest an optimal list of top-n visualizations for heterogeneous data sources in a personalized manner.

Keywords: Personalized visualizations, Visualization Recommender, Recommender Systems, Collaborative Filtering, Crowd-Sourcing

1 Introduction

Despite recent technical advantages of search engines and content provider services the information overload problem still remains a crucial issue for many application fields. Finding the right piece of information in huge information spaces is a tedious and time consuming task. Recent innovations such as recommender systems help overcome this issue, though with limited success due to limitations in presenting information items of a top-n list, typically in textual form. On the other hand, visualizations have shown to be an effective way to deal with the overload issue providing the possibility to display and explore a huge set of data points at the same time. However, creating useful visual representations of data typically requires expert knowledge. Up to now, only a few approaches attempted to automatically generate useful visual representations of a given set of data [14] [9], albeit with certain limitations. Despite their usefulness these approaches exhibit weaknesses in dealing with highly heterogeneous data and ignore the fact that visual representation of data is a matter of the users taste or preferences. To fill this gap, we present in this paper a novel approach – called *VizRec* – which tackles these issues by: (i) automatically generating a set of visualization in the context of heterogeneous data and (ii) recommending the most useful visualization in a personalized manner, helping the user to explore large amounts of data efficiently.

Problem Statement. The problem we are dealing with in this work is the generation of an optimal list of top-n visualizations for the user given a set of heterogeneous data sources as input. Considering just visual encoding rules as proposed in the literature [14], leads to a huge set of possibilities, "valid" in terms of the visually representing the data, but without consideration about which type best serves user's needs.

VizRec deals with the issue by (1) automatically identifying the set of appropriate visualizations using a rule-based algorithm to analyze compatibility between visuals and input data, and (2) filtering a subset based on user's preferences to be recommended as the list of top-n visualizations that best reflect the user's information needs.

Contributions. The contributions of this work can be summarized as follows:

- A novel visual recommender approach to generate and recommend personalized visualizations.
- An extensive evaluation of visualization types in the context of three data repositories conducted in Amazon Mechanical Turk, providing insights on the usefulness of the approach.

Paper structure. Overall, the paper is structured as follows: Section 2 presents related work in the area. Section 3 introduces *VizRec*. Section 4 presents the methodology we choose to evaluate our approach. Section 5 highlights the results of our evaluation and Section 6 concludes the paper and provides insights in how the current work will be extended.

2 Related Work

Recommending visualizations is a relatively new strand of research and only little effort has been put in so far to tackle this challenge. The closest approach to our intention is a system described by Voigt et al. [4], which uses a knowledge base of numerous ontologies to recommend visualizations. It is essentially a rule based system that pre-selects visualizations based on their support of device, data properties and task. In a second stage, the system ranks visualizations following rules about visualization facts, domain assignments, and user context. One disadvantage of Voigt et al.'s approach is that both visualizations as well as data inputs need to be annotated semantically. Furthermore, both the pre-selection and the ranking stages are rule-based. But more importantly, a large theoretic part of the work lacks empirical support. While the user preferences such as graphical representations and visualization literacy are outlined, the actual collection and validation of user preferences is left for future work.

In contrast, we present a complete Collaborative Filtering (CF) approach, collecting use preferences for personalization form a large study involving the general public, validating them in an offline experiment and drawing insights from empirical evidence. Our approach starts by strictly describing the visual encoding process, hence we represent visualizations in terms of their visual components (see [3] for thorough description of visual components). Instead of pursuing a through specification encompassing all known expert knowledge about visual perception, we concentrate on pragmatic, simple facts that will aid the sensible mapping of data onto visual components (e.g., [6]), extending the description to many different types of visualizations. Next, in contrast to focusing only on specific format and domain, we obtain and visualize our data from heterogeneous data sources.

Mackinlay et al. describe an influential, albeit conceptually different approach, in the ShowMe [8] system. It integrates a set of user interface commands and functions aimed to automatically generate visualizations for (*Tableau*[1]). ShowMe attempts to support the user by searching for graphical presentations that may address their task. Appropriate visualizations are selected based on data properties, such as datatype (text, date, time, numeric, boolean), data role (measure or dimension) and data interpretation (discrete or continuous). The ranking for visualizations is based on static ratings (scores) globally defined for every supported chart type. We follow a similar approach to select visualizations based on encoding rules. In contrast to having global ratings, our methods allows to personalize the resulting visualizations to interests of the individual user using a CF approach.

Nazemi et al.'s system suggests visualizations based on user preferences [9], incrementally gathered during interaction with the visualization system in form of usage profiles for particular charts. Nazemi et al. take a bottom-up approach, analyzing user interaction with visualization to describe user behavior. Instead, we describe a top-down method to elicit user preferences by collecting ratings. These methods are complementary and can be deployed together with user behaviour analytics. As in our case, Nazemi et al. utilize a personalized approach to suggest visualizations, though they only target the content from digital libraries (i.e., bibliographical notes, publications).

Ahn et al.'s work on adaptive visualization attempts to provide user adapted visual representation of their search results [11]. The user context is a collection of user actions accumulated over time, such as the issued search queries, selected documents from the search results and traversed links. The collection serves to capture user interests beyond the query and to define in turn a user model, which is in fact applied to visually highlight the relevance of a particular result set. In contrast, *VizRec* augments user queries with preferences in order to find best representation of the information behind the queried content, instead of only displaying relevant results as clusters.

Despite these notable efforts, the problem of recommending visualizations is still sparsely explored. Especially, in the context of generating and suggesting useful visualizations for heterogeneous multidimensional data sources not much research has been conducted. Also there seems to be a gap in the literature on doing this in a personalized manner, since previous work on recommender systems has shown, that the one-size-fits-it-all principle typically does not hold. To contribute to this sparse strand of research we have invented and evaluated *VizRec*, a novel visual recommender engine capable of recommending different types of visualizations for heterogeneous datasources in a personalized manner.

3 The VizRec Approach

Figure 1 shows the general workflow of *VizRec* to generate personalized visualizations for heterogeneous data sources (HDS). As highlighted, the system responds to a given

[1] Tableau: http://www.tableausoftware.com/

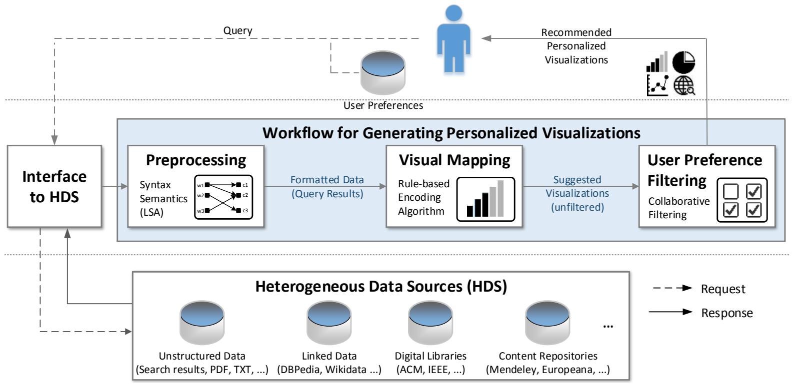


Fig. 1: Schematic representation of the VizRec recommendation pipeline.

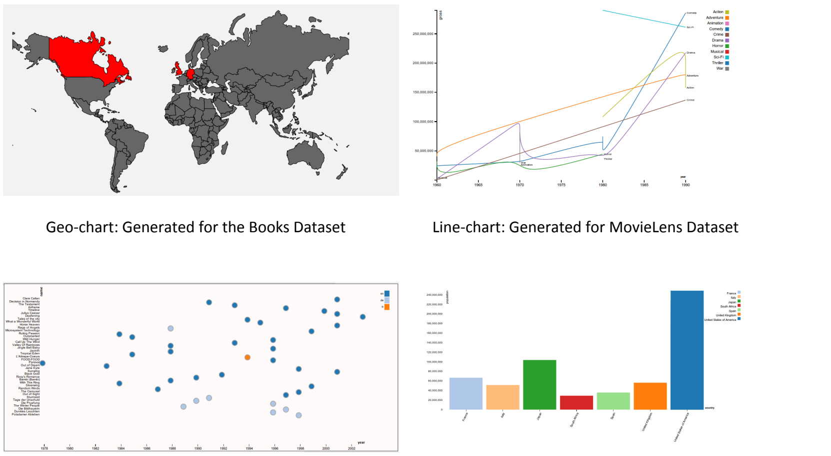
search query and a given data source, with a set of visualizations that reflect the user's personal preferences in a top-n sorted manner. Before deciding on the appropriate visualizations the filter pipeline first annotates retrieved data and then performs data analysis tasks to categorize them into standard and/or specific datatypes. After that, a mapping operation is performed (based on visual perception and visual encoding guidelines [14]) that maps the data to the visual components (encoding some attributes of the data, e.g using axes of a visualization) of the appropriate visualizations.

In the final step, the system includes user preferences in a collaborative filtering [2] approach, which takes into account a set of specific usability preferences that have been collected in the past. In summary, there are three steps to generate personalized visual recommendations: (1) preprocessing, (2) mapping and (3) user preference filtering. In the following subsections we shortly describe each of those units:

Step 1: Preprocessing. The preprocessing unit is responsible for extracting and annotating data attributes appropriate for mapping. Associated data sources, such as Linked Data, ACM digital library, or Mendeley, collect and index various kinds of documents, e.g., conference publications, books, journals, lectures and images. Each data source defines and organizes its repositories according to a (often closed) proprietary data model. Many scientific digital libraries for instance, define the structure of their literature archives in terms of some important attributes, such as title, abstract, author, keywords, etc. following e.g., the Dublin Core metadata format.

Before the mapping algorithm can begin to establish correspondence with visualizations, the data in these various formats have to be first collected in series and then categorized according to datatypes. The data is first categorized into standard datatypes such as categorical, temporal and numerical – represented by primitive data types string, date and number, respectively. This categorization into primitive datatypes is basically done by analysis values of the individual attributes. To do so, the analysis performs a top-down approach, i.e., for a given value, it is first decided to which of the aforementioned standard datatypes a given value belongs. In further step, using gazetteer lists the more specialized datatypes are derived, such as for spatial information.

Step 2: Visual Mapping. A visualization can be broken down in a number k of visual components, each of which encodes a single piece of information visually [3]. If



Geo-chart: Generated for the Books Dataset

Line-chart: Generated for MovieLens Dataset

Timeline: Generated for Books Dataset

Bar-chart: Generated for EU Dataset

Fig. 2: Four example charts generated via *VizRec*. Note that not all charts are equally useful (see e.g., top-right chart).

every visual component could encode any kind of data, the possible number of combinations for a visualization type would be given by $\binom{n}{k}$, where n is the number of data attributes in a dataset (i.e., number of fields). Hence, for an example dataset with one *date*, two *strings* and two *numbers* to be represented in a barchart with two visual components, the total number of combinations would be $\frac{n!}{(n-k)!} = \frac{5!}{(5-2)!} = 20$. But many of these combinations would be perceptually incorrect, because visual components are often suited to represent only some kinds of data attributes, given by the perceptual properties of the channel and the characteristics of the data attribute [3].

Visual mapping identifies which attributes of the data can be related to which visual components of a visualization type [7]. These relationship is established based on datatype similarity between data attributes and visual components. To do so we benefit from an ontology of patterns [14] for a type of visualization. Each pattern describes one possible mapping for a concrete visualization in terms of its visual components and supported datatypes. For instance, two possible patterns for the bar chart could be (1) $\{x-axis : string,\ y-axis : number\}$, and (2) $\{x-axis : date,\ y-axis : number\}$. The patterns specify the types of data required for each visualization to be instantiated. Hence, each pattern i defines for each visual component j, which r_j attributes should be selected from n_j data attributes: $\frac{n_j!}{r!(n_j-r_j)!} = \binom{n_J}{r_j} = C^r_{n_i}$. Note that n_j is a subset of n that complies with datatype compatibility for the j visual component r_j . To obtain the total number of combinations M_i, being generated for a particular pattern i, we multiply every suitable $\binom{n_j}{r_j}$ visual component of a pattern: $M_i = \prod C^{r_j}_{n_j}$. Thus, the final number of patterns M of a visualization is nothing else then the sum of every M_i. In our working example, for bar chart's pattern (1) one attribute with datatype *string* and one with datatype *number* we obtain $M_i = C^2_1 \times C^2_1 = \binom{2}{1} \times \binom{2}{1} = 4$ possible mappings. And for pattern (2) one attribute with datatype *date* and one with datatype

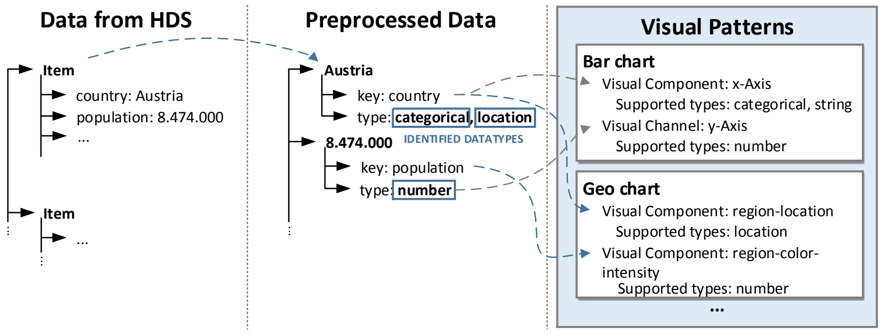


Fig. 3: Visual mapping process: identifying mapping combinations for Bar and Geo charts considering datatype compatibility between their visual components and data form HDS.

number, we obtain $M_i = C_1^1 \times C_1^2 = \binom{1}{1} \times \binom{2}{1} = 2$ possible mapping combinations. Hence, the total number of combinations for this type of chart would be 6 using this particular dataset.

Having obtained all the combinations, the mapping operator finally maps data to the corresponding visual components of a visualization based on the following principles: (i) one data attribute will be instantiated to one visual channel of a visualization, (ii) the datatype of the attributes should be compatible with those of the channels, and (iii) every mandatory visual channel of a visualization should be instantiated. Once the mapping process is completed, *VizRec* presents the mapping combinations as a set of appropriate visualization configurations to the user. This process is illustrated in Fig. 3.

Visual patterns in cooperation with rule based mapping algorithm generate all mapping combinations which are plausible for the data, but not all of them represent what the user needs or prefers. Therefore, we need better mechanisms for selecting the visualization. To achieve that we involve users to validate the mapping results. We benefit from collaborative filtering (CF) [18] which allows us to collect user feedback in form of ratings and to apply them in such a way as to provide reasonable prediction of the active user's preferences. In our context, we provide predictions for the mapping combinations the user might prefer based on the current user's and similar user's preferences.

Step 3: User Preference Filtering. To finally filter the generated mapping combinations according to the user's preferences, we employ a simple user-based CF approach . Hence, for a given dataset, the mapping algorithm provides a set of possible combinations M each serving as a possible *item* to be recommended to the user. The list of recommendations R for the current user is nothing else but a subset of M. Concretely, given a set of active user's ratings U and a set of predictions P, where both should contain ratings for the items from M, we denote $R = U \cup P$. Note that the calculation of P involves calculating the k-nearest neighbors (based on *Pearson correlation*) to the active user, which liked the same mapping combinations as the active user in the past, and have rated mapping combinations $x \in M$ active user has not yet seen.

For the calculation of R, we first take the set U_p, containing all ratings of the active user given for different mappings and the set N_p, containing all ratings given by other users and build the set $M_p = U_p \cup N_p$. From M_p we construct the matrix A, consisting of user-ids, item-ids, and the ratings, which is passed to generate the predictions for the current user. For this purpose we make use from memory based CF approach [2] which generates a list of top-n visual recommendations.

4 Evaluation

This section describes the experimental setup in detail, the data sources, the method and metrics used to validate our approach.

Datasets & Mappings. The study used the following three open-source datasets:

(Movielens[2]*) Dataset (*movies*):* This dataset comprises information about the top ranked movies for the years 1960, 1970, 1980, and 1990. It counts 41 entries which are selected of items from the respective dataset and are characterized by the attributes: (movie) name, budget, gross, creation year, and shooting location. Based on this, the mapping unit produced four types of visualizations (see Fig. 2) with the following mapping frequencies: 32 bar-charts, 9 line-charts, 13 timelines and 1 geo-chart. Hence, a total of 55 mapping combinations were generated.

EU Open Linked Data Portal[3] *Dataset (*eu*):* The *eu* dataset collects for 28 EU countries, the percentage of the population looking for educational information online in the years 2009–2011. It counts 91 entries characterized by attributes: (country) name, year, language, population, constitutional form and value in percent of the population looking for educational information. The mapping unit suggested 30 possible chart combinations, concretely 15 bar charts, 6 line charts, 8 timeline and 1 geo chart.

Book-Crossing Dataset[4] *(book):* This dataset contained 41 randomly chosen books, published between 1960 and 2003, characterized by the attributes: name, country, publisher, and year. The mapping unit suggested 3 chart types: bar chart with 2 combinations, geo chart with 1 combination and timeline with 3 combinations, totaling 7 mapping combinations.

Procedure. Our experimental approach was to gather user preferences for visualizations obtained from the rule-based system and to then train a RS to suggest visualizations. A crowdsourced study was designed to obtain personalized scores for each chart suggested by the visual recommender. To give a score, a participant would have to perform some cognitively demanding task with the chart (i.e., a minimal analysis). Based on the experiments conducted by Kittur et al. [13], this preparatory task should bring participants to accurately study the combination and prevent a randomly or rash rating. Hence, we designed a task as follows: 1) a participant was given a one line description of the dataset originating the chart, 2) looking at the chart she had to write tags (at most five) and a title for it, then 3) the participant rated the chart. The score system used a multidimensional scale adapted from a list of usability factors presented in [10] and [12]: (1) cluttered, (2) organized, (3) confusing, (4) easy to understand, (5) boring, (6)

[2] Movielens: https://movielens.org/
[3] Eu: https://open-data.europa.eu/en/linked-data
[4] Book-Crossing Dataset: http://www2.informatik.uni-freiburg.de/ cziegler/BX/

exciting, (7) useful, (8) effective, and (9) satisfying. Note that dimensions 1–6 are duplicated with opposing sentiment (e.g., cluttered vs organized). Opposing dimensions were used to ensure meaningful ratings for scales with complex meaning..Dimensions were rated on a 7-point Likert scale (1=not applicable – 7 very applicable).

As the chart scores were intended for the offline experiment, it was required that a participant rate more than one chart. We experimented with varying sizes of HITs (Human Intelligent Task), collecting ten (10) and five (5) tasks (chart/combinations and their corresponding ratings). In pilot studies these turned out to take overly long (around 15mins), so we settled for collecting three (3) chart/combinations per HIT. Suggested combinations were distributed in 32 HITs, each of which contained 3 randomly chosen mapping combinations. Pilot studies also helped streamline dataset descriptions, task descriptions and instructions across the experiment. After accepting a HIT, the participant (worker or turker) received a tour to complete a task. The tour showed a chart and corresponding tags, title and ratings in the exact same format as the subsequent experiment. When ready the worker started the first task in the HIT by pressing a button. Workers were allowed to write *not applicable* or NA for tags, but were alerted if they failed to write any tags. The rating dimensions were not assigned a score until a worker did it. Workers could only proceed if they had rated all dimensions. A HIT with three chart/combinations was compensated with $1.00. A worker rated a minimum of three charts, but to ensure a more realistic training set for the CF-RS, workers were allowed to perform more than one HIT. Only expert workers how achieved consistently a high degree of accuracy by completing HITs were allowed to take part in the study.

Evaluation Protocol. A set of studies was carried out to analyze variability in preference scores. To compute the overall score for a chart for each worker, the scores in opposing dimensions (clutter, confusing, boring) were inverted and then all dimensions were averaged together according to the following formula $SC = \left(\sum_{i=1}^{k} \rho_k D_k\right)/k$. Where $k = 9$ is the number of dimensions, ρ_k is the coefficient 1 and D_k is k dimension score. The chart score was obtained by averaging the worker scores.

In the second part of our evaluation, we performed an offline experiment to estimate the performance of personal preferences for visualization recommendations. To this end, we used the preferences collected from turkers as a training data for our recommender. Following the method described in [15], we split the preference model into the two distinct sets, one for training the recommender (training-set), and another one for testing (test-set). The test-set acts here as a reference value that, in an ideal case, has to be fully predicted for the given training-set. From each of the datasets in the preference model, we randomly selected 20% of user-rated mapping combinations (visualizations) and put them into the test-set. The recommendations produced out of the training-set are further used to evaluate the performance of *VizRec*. The performance of *VizRec* depends generally on how good it predicts the test-set. We compared the generated recommendations (prediction-set) and the test-set by applying a variety of well-known evaluation metrics in information retrieval[16]: Recall (R), Precision (P), F-Measure (F), Mean Average Precision (MAP) and the Normalized Discounted Cumulative Gain ($nDCG$). The first three metrics basically express the quantity of relevant recommended results,

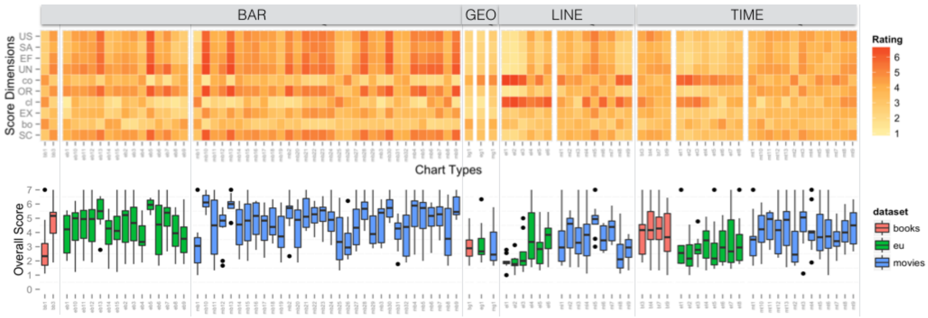


Fig. 4: Mean and Variability in Scores (1=completely disagree, 7=totally agree). The heatmap illustrates the contribution of the 9 dimensions (US=useful, SA=satisfying, EF=efficient, UN=understandable, co=confusing, OR=organized, cl=cluttered, EX=exciting, bo=boring) to the overall score (SC). The boxplot below illustrates the high variability in personal ratings.

whereas MAP, and $nDCG$ quantify the concrete ordering of results (i.e., giving penalties if the results are not on the top but are relevant for the user).

5 Results

Participants. Each HIT was completed by ten different workers. For the 92 visualizations, 8280 scores across 9 dimensions were collected by 70 participants. Participants completed in average 4.7 HITs. The experiment started on 26th of November and ended on 3th of December 2014. The allotted working time per HIT was 900 sec. and the average working time of workers was 570 sec. per HIT.

Visual Quality. The heatmap in Fig. 4 shows the mean rating for every dimension for each chart. Firstly, the results confirm a clear understanding of the opposing dimensions, Negative dimensions in lower case received opposite scores to corresponding positive ones (UN-co, OR-cl, EX-bo, in Fig. 4 top). The aggregated score for each chart in the bottom row of the heat map (SC) shows that only a handful of charts achieved clearly high scores whereas in category there were charts above the midline. More importantly, boxplot at the bottom explains these scores: there is a broad variability in scores for most chart instances. This confirms our assumption that user preferences matter when choosing the right representation. With regards to H1, results confirm that only a very small number of charts achieved high scores and the rest present wide variability.

From the heat map, it is already possible to identify individual top-scoring charts. To establish differences in chart categories and datasets we performed a factorial ANOVA with chart type and dataset as factors (chart-type: *bar, line, time, geo* and dataset: *Movies, Books, Eu*). Homogeneity of variance was confirmed with a Levene test. The factorial ANOVA revealed a significant effect of dataset $F(2, 908) = 21.19, p < 0.0001$, a significant effect of chart type $F(3, 908) = 38.98, p < 0.001$ and significant interaction effect dataset chart type $F(5, 908) = 3.81, p < 0.01$. TukeyHSD multiple comparisons revealed a significant difference in scores between *movies* ($M = 4.86$) and *books* ($M = 3.82$) $p < 0.05$, as well as between *movies* and *Eu* data ($M = 3.68$),

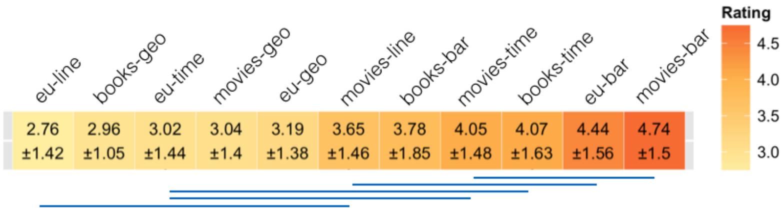


Fig. 5: Significant Interactions Chart Type / Dataset. The heat-map illustrates the mean score and standard deviation for each combination of *dataset-chart type* (1=completely disagree, 7=totally agree). The lines below show where differences start to be significant. Note that due to its high variability, *books-bar* is not significantly better than *eu-line*, whereas *movies-line* is.

$p < 0.001$. For chart type, there was a significant difference in scores between *bar* ($M = 4.60$) and *geo* ($M = 3.06$) $p < 0.001$, *bar* and *line* ($M = 3.29$) $p < 0.001$, *bar* and *time* ($M = 3.72$) $p < 0.001$, as well as between *time* and *line*, $p < 0.02$. The significant effects of multiple comparisons for interaction are shown in Fig. 5.

The main outcomes are the information about user preferences and the clear differences amongst them. The interaction effects illustrate several differences amongst chart type. But these results are merely a hint that there are varied preferences. Looking at each dataset, chart and chart type in the heat map of Fig 4, it is clear that while a small number of charts are generally preferred, in most cases the ratings vary widely and a personalized approach would better accommodate those user preferences.

Recommendation Quality. At a glance, the results of our offline evaluation reveal significant improvements in recommendation quality achieved through the use of individual user preferences. To measure the improvements in quality, we compared the *VizRec* CF with the baseline filtering algorithms Most Popular (MP) [17], and Random (RD). The RD simulates the recommender behavior providing an arbitrary order of visualizations – i.e., it can be compared with having only the first two units in the *VizRec* pipeline from Fig. 1. The MP, in contrast, generates the results sorted according to global ratings, accumulated in our case from ratings of individual user. Considering RD and MP baseline algorithms should unveil whether the recommender systems in general help in providing useful visualizations, and whether the personalized approach improves the quality of results respectively.

For the comparison, we analyzed the top 3 recommendations, as our datasets are of relatively small size (compared to some commonly used datasets, such as BibSonomy and CiteULike [15]). The results of the evaluation are summarized in Tab. 1.

The results show that *VizRec* CF outperforms both baseline algorithms in all three datasets. Concretely for the RD, the first three quality metrics clearly show that results are more accurate with *VizRec* CF compared to simply generating arbitrary visualizations (cf., $F@3(CF) = .1257$ and $F@3(RD) = .0055$ for Movies). Additionally, *MAP@3* and *nDCG@3* reveal that *VizRec* CF sorts individual visualizations according to their relevance for a user significantly better. Note that the difference between individual metrics amongst datasets is to a large extent influenced by the considerable difference in size of the three datasets (e.g., Books has only 7 different visualizations – $F@3(CF) = .4778$, whereas Movies has 55 – $F@3(CF) = .1257$, see Fig. 4).

Dataset	Alg.	Metric				
		R@3	P@3	F@3	MAP@3	nDCG@3
Movies	**CF**	.1152	.2111	.1257	.0793	.1271
	MP	.0488	.0926	.0591	.0163	.0419
	RD	.0039	.0093	.0055	.0020	.0048
EU	**CF**	.1526	.2632	.1877	.1263	.1721
	MP	.0263	.0175	.0211	.0088	.0161
	RD	.0132	.0175	.0150	.0044	.0103
Books	**CF**	.5333	.4555	.4778	.4889	.5000
	MP	.1333	.0444	.0667	.0444	.0667
	RD	.0667	.0222	.0333	.0333	.0420

Table 1: Quality metrics values P@3, R@3, F@3 MAP@3, NDCG@3 estimated for the three different datasets using baseline algorithms MP, RD, and *VizRec* CF.

Another interesting finding is that the recommender strategy based on global ratings, MP, generated less accurate results than *VizRec* CF for collected user preferences, both in providing relevant visualizations and in their ranking order. This supports our main assumption that in the face of the wide variability in user preference ratings, the personalized approach performs better recommendations.

6 Discussion and Outlook

This work builds on the premise that the preference of a visual representation for a dataset is a matter of personal preference. Empirical evidence collected through a crowd sourced experiment supports the assumption that preferences widely vary for visual representations generated automatically. The second motivation driving our work is that a CF approach to recommend visualizations can account for such variability in personal preferences and significantly improve the recommendation. Our offline experiment supports our assumptions showing that *VizRec* CF outperformed both the random approach (RD) and the global best approach (MP). A major contribution is that the presented work is based on empirical evidence collected with a methodical study involving the general public. The approach to generate and suggest visualizations, the process of elicitation of user's preferences, and the insights obtained therewith are, to the best of our knowledge novel contributions.

Several open questions remain that we plan to address in our continuing research. First, our solution suffers the cold start problem of CF-RS: a user who has not rated any chart cannot be recommended anything. To tackle this issue, we investigate applying measuring semantic similarity of the data attribute array, to establish if a similar structure has been seen before and suggest from global ranking of other users. Furthermore, the investigation on our crowdsourced experiment is still ongoing. A thorough exploration of the relations between quality of content features (such as textual description) with the valued quality of a visualization is beyond the scope of this paper. But we are currently investigating the application of content features to the cold start problem and also to determine the tasks a user associates with preferred visualizations.

Acknowledgements. This work is funded by the EC 7th Framework project EEX-CESS (grant 600601). The Know-Center GmbH is funded within the Austrian COMET Program - managed by the Austrian Research Promotion Agency (FFG). The study was carried out during the tenure of an ERCIM "Alain Bensoussan" fellowship program by the third author. The authors would like to thank Dominik Kowald for providing very constructive and helpful suggestions to conduct the offline evaluation.

References

1. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.A.: Indexing by latent semantic analysis. In: Journal of the American Society for Information Science 41. 391-407 (1990)
2. Xiaoyuan, S., Khoshgoftaar, T.M.: A Survey of Collaborative Filtering Techniques. In: Adv. in Artif. Intell. (2009)
3. Bertin, Jacques : Semiology of graphics. University of Wisconsin Press (1983)
4. Voigt, M., Franke. M., Meissner, K.: Using and Empirical Knowledge for Context-aware Recommendation of Visualization Components. In: Proc. of eKnow'12 (2012)
5. Schelter, S., Owen, S.: Collaborative Filtering with Apache Mahout. In: RecSysChallenge'12 (2012)
6. Mackinlay, J.,: Automating the design of graphical presentations of relational information. In: ACM Trans. Graph. 110-141. (1986)
7. Rahm, E., Bernstein, Philip A.: A Survey of Approaches to Automatic Schema Matching. In: The VLDB Journal. Springer (2001)
8. Mackinlay, J., Hanrahan, P., and Stolte, Ch.: Show Me: Automatic Presentation for Visual Analysis. In: IEEE Transactions on Visualization and Computer Graphics. 1137-1144 (2007)
9. Nazemi, K., Retz, B., Bernard, J., Kohlhammer, J.,Fellner, D.: Adaptive Semantic Visualization for Bibliographic Entries. Springer (2013)
10. Seffah, A., Donyaee, M., Kline, R., Padda, H.: Usability measurement and metrics: A consolidated model. In: Software Quality Journal. 159-178 (2006)
11. Ahn, J., Brusilovsky, P.,: Adaptive Visualization of Search Results: Bringing User Models to Visual Analytics. In: Information Visualization. 167-179 (2009)
12. Zheng, Xianjun S., Lin, James j W., Zapf, S., Knapheide, C. : Visualizing User Experience Through "Perceptual Maps": Concurrent Assessment of Perceived Usability and Subjective Appearance in Car Infotainment Systems. Springer (2007)
13. Kittur, A., Chi, Ed H.,Suh , B.: Crowdsourcing User Studies with Mechanical Turk. In: Proc. of CHI'08 (2008)
14. Mutlu, B., Hoefler, P., Tschinkel, G., Sabol, V., Granitzer, M., Stegmaier. F.: Suggesting Visualizations for Publisched Data. In: Proc. of VISIGRAPP'14 (2014)
15. Trattner, C., Kowald, D. and Lacic, E.: TagRec: Towards a Toolkit for Reproducible Evaluation and Development of Tag-Based Recommender Algorithms, ACM SIGWEB Newsletter, Spring 2015, ACM, New York, NY, USA, 2015.
16. Herlocker, J.L., Konstan, J.A., Terveen, L., G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst., 22(1):5-53, (2004)
17. R. Jaeschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme.: Tag recommendations in social bookmarking systems. Ai Communications, 21(4):231247, 2008
18. J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen.: Collaborative filtering recommender systems. In The adaptive web, pages 291324. Springer, 2007