# Towards a Recommender Engine for Personalized Visualizations

Belgin Mutlu(✉), Eduardo Veas, Christoph Trattner, and Vedran Sabol

Know-Center GmbH, Inffeldgasse 13, 8010 Graz, Austria
{bmutlu,eveas,ctrattner,vsabol}@know-center.at

**Abstract.** Visualizations have a distinctive advantage when dealing with the information overload problem: since they are grounded in basic visual cognition, many people understand them. However, creating them 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 to 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 can 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 advances in search engines and content provider services, the information overload problem still remains a crucial issue in 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 to resolve the issue, though with limited success, due to limitations in the way the recommended items are presented, typically as a list in the textual form. Alternatively, visualizations have shown to be an effective way to deal with the overload issue by opportunity to display and explore a huge set of data points simultaneously. However, creating useful visual representations of data typically requires expert knowledge. To date, only a few approaches attempted to automatically generate visual representations given a set of data [14] [9], albeit with certain limitations. Despite their usefulness, these approaches are ineffective in terms of dealing with highly heterogeneous data and ignore the fact that visual representation of data is a matter of the users' taste or preferences. To address this issue, in this paper we present a novel approach – called *VizRec* – which tackles these challenges by: (i) automatically generating a set of visualizations in the context of heterogeneous data and (ii) recommending the most useful


F. Ricci et al. (Eds.): UMAP 2015, LNCS 9146, pp. 169–182, 2015.
DOI: 10.1007/978-3-319-20267-9_14

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 proposed in the literature [14] leads to a large set of possibilities, valid in terms of representing the data visually, but without considering which type serves the users' needs best.

*VizRec* deals with the issue by (1) automatically identifying the set of appropriate visualizations using a rule-based algorithm to analyze the compatibility between the visuals and the 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.

The paper is structured as follows: Section 2 presents the related work in the area. Section 3 introduces *VizRec*. Section 4 presents our methodology for evaluating our approach. Section 5 highlights the results of our evaluation and Section 6 concludes the paper and provides insights into how the current work will be extended.

## 2 Related Work

Recommending visualizations is a relatively new strand of research and only few efforts have been made in so far to tackle this challenge. The closest approach to our suggestion 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 the device, data properties and task involved. At a second stage, the system ranks visualizations following the rules concerning visualization facts, domain assignments, and user context. One disadvantage of Voigt et al.'s approach is that both visualizations and data inputs have to be annotated semantically beforehand. Furthermore, the pre-selection and the ranking stages are rule-based. More importantly, a large theoretical part of the work lacks the empirical support. While user preferences, such as graphical representations and visualization literacy, are outlined, the actual collection and validation of user preferences are tasks for future work.

In contrast, we present a complete Collaborative Filtering (CF) approach by collecting user preferences for personalization from a large study involving the general public, validating them in an offline experiment and drawing conclusions based on the empirical evidence. Our approach starts by strictly describing the visual encoding process, i.e., we represent visualizations in terms of their visual components (see [3] for thorough description of the 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 types of visualizations. Next, instead of focusing only on specific data format and domain, we obtain and visualize heterogeneous data sources.

Mackinlay et al. propose 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 help the user by searching for graphical presentations that may address their task. Appropriate visualizations are selected based on the data properties, such as datatype (text, date, time, numeric, boolean), data role (measure or dimension) and data interpretation (discrete or continuous). The ranking of visualizations is based on static ratings (scores) globally defined for every supported chart type. We follow a similar approach and select visualizations based on the encoding rules. Rather then using global ratings, our method allows us to personalize the resulting visualizations according to the 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 the form of usage profiles for particular charts. Nazemi et al. follow a bottom-up approach, analyzing user interaction via visualization to describe user behavior. In contrast, we apply a top-down method to elicit user preferences by collecting ratings. These methods are complementary and can be deployed together with user behaviour analytics. Similar to us, Nazemi et al. utilize a personalized approach to suggest visualizations but 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 captures user interests beyond the query and in turn defines a user model, which is applied to visually highlight the relevance of a particular result set. In contrast, *VizRec* augments user queries with preferences in order to find the 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 insufficiently explored, especially little research has been performed on generating and suggesting useful visualizations for heterogeneous multidimensional data.Moreover there seems to be a gap in the literature on doing this in a personalized manner, since previous work on recommender systems has shown

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

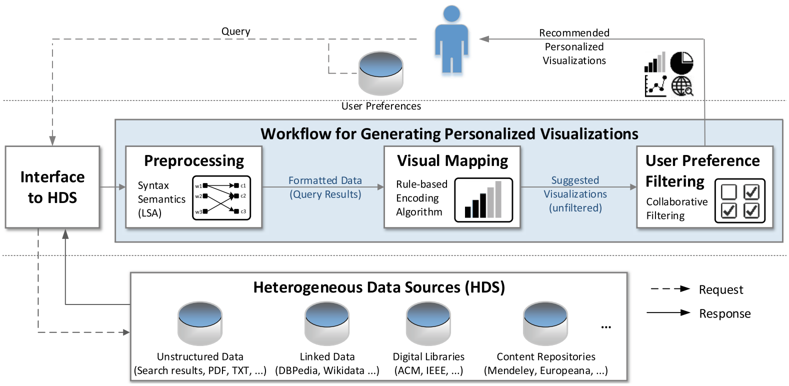


**Fig. 1.** Schematic representation of the VizRec recommendation pipeline

that the one-size-fits-it-all principle typically does not hold. To contribute to this small body of research we developed and evaluated *VizRec*, a novel visual recommender engine capable of recommending various 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 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 the 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.

As the final step, the system includes user preferences via a collaborative filtering [2] approach, which takes into account a set of specific usability preferences that have been collected in the past. In summary, the three steps to generating personalized visual recommendations are: (1) preprocessing, (2) visual mapping and (3) user preference filtering. In the following subsections, we briefly 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 and 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 an (often closed) proprietary data model. Many scientific digital

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).

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 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 performed by analyzing values of the individual attributes. To do so, the analysis employs a top-down approach, i.e., for a given value it is first decided to which of the aforementioned standard datatypes it belongs. Next, by using gazetteer lists more specialized datatypes are derived, e.g., 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 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). For example a 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 $\frac{n!}{(n-k)!} = \frac{5!}{(5-2)!} = 20$. However many of these combinations would be perceptually incorrect, since 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].

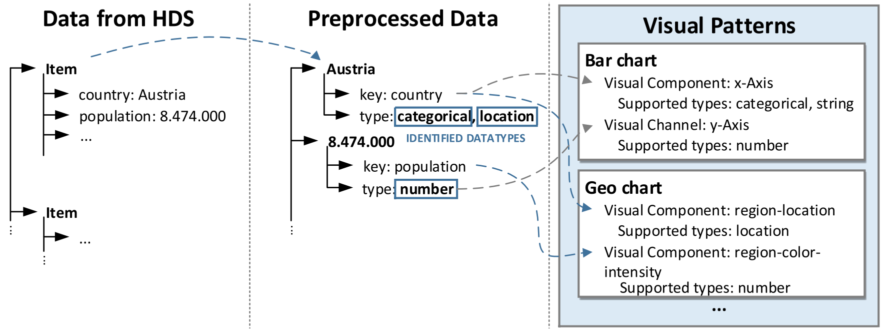


**Fig. 3.** Visual mapping process: identifying mapping combinations for Bar and Geo charts considering datatype compatibility between their visual components and data form HDS

Visual mapping identifies which attributes of the data can be related to which visual components of a visualization type [7]. The relationship is established based on the datatype similarity between the data attributes and the 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, 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 that are 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 $M_i$, 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 *number*, we obtain $M_i = C^1_1 \times C^2_1 = \binom{1}{1} \times \binom{2}{1} = 2$ possible mapping combinations. Hence, using this particular dataset the total number of combinations for this type of chart would be 6.

Having obtained all the combinations, the mapping operator 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 the datatypes 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 together with rule-based mapping algorithm generate all mapping combinations which are plausible for the data. Since not all of them represent what the user needs or prefers, better mechanisms for selecting the visualization are required. To that end, we ask 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 a way that provides reasonable prediction of the active user's preferences. In our context, we make predictions for the mapping combinations that the user might prefer based on her and similar users' 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. 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 $P$, both of which 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*) of the active user, who liked the same mapping combinations as the active user in the past and rated mapping combinations $x \in M$ active user has not seen yet.

For the calculation of R, we first take the set $U_p$ containing all ratings of the active user given for various mappings and the set $N_p$ containing all ratings given by other users and develop the set $M_p = U_p \cup N_p$. Based on $M_p$, we construct the matrix $A$ consisting of user-IDs, item-IDs and the ratings, that generates the predictions for the current user. For this purpose, we applied the memory based CF approach [2] that generates a list of top-n visual recommendations.

## 4 Evaluation

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

**Datasets and 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 has 41 entries, which are selected from items of 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.

[2] Movielens: https://movielens.org/

*EU Open Linked Data Portal*[3] *Dataset (*eu*):* The *eu* dataset collects the percentage of the population looking for educational information online in the years 2009–2011 for 28 EU countries. It has 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] *(books):* This dataset contained 41 randomly chosen books published between 1960 and 2003 and 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, the total of 7 mapping combinations.

**Procedure.** Our experimental approach was to gather user preferences for visualizations obtained from the rule-based system and train a RS to suggest visualizations. A crowdsourced study was designed to obtain personalized scores for each chart suggested by the visual recommender. Before giving a score, a participant had 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 random or rash rating. We designed the task as follows: 1) a participant was given a one line description of a dataset originating the chart, 2) looking at the chart she had to write tags (at most five) and a title for it and 3) rate 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) 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).

Since the chart scores were intended for the offline experiment, the participant had to 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). Since in pilot studies these turned out to take overly long (around 15mins), 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 to 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, which 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

[3] Eu: https://open-data.europa.eu/en/linked-data

[4] Book-Crossing Dataset: http://www2.informatik.uni-freiburg.de/~cziegler/BX/

but were alerted if they failed to write any tags. The rating dimensions were not assigned a score until the 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 who consistently achieved 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 the 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, $\rho_k$ 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 workers as 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 is a reference value that, ideally,can 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 entered 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* generally depends on how well 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, whereas $MAP$ and $nDCG$ quantify the concrete ordering of the 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 workers. For 92 visualizations, 8280 scores across 9 dimensions were collected from 70 participants. The participants completed on average 4.7 HITs. The experiment started on November 26,2014 and ended on December 3, 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. The results confirm a clear understanding of the opposing

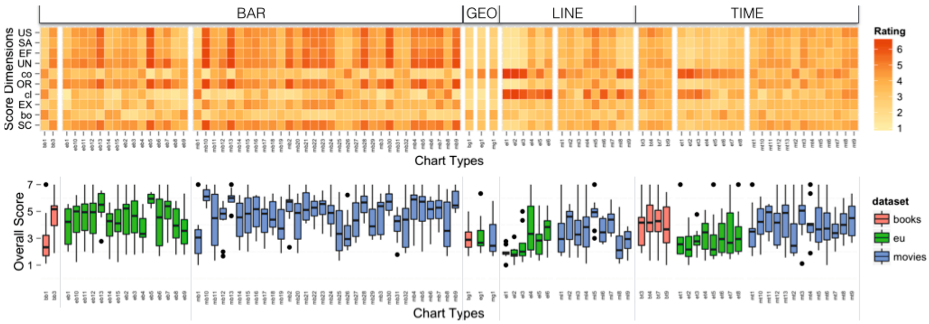


**Fig. 4.** Mean and variability in scores (1=completely disagree, 7=totally agree). The heatmap illustrates the contribution of 9 dimensions (US=useful, SA=satisfying, EF=efficient, UN=understandable, co=confusing, OR=organized, cl=cluttered, EX=exciting, bo=boring) to the overall score (SC). The boxplot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

dimensions. Negative dimensions in the 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 the 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. The results confirm that only a very small number of charts achieved high scores and the rest were variable.

From the heat map individual top-scoring charts can be identified. To establish differences in the chart categories and datasets, we performed a factorial ANOVA with the chart type and dataset as factors (chart-type: *bar, line, time, geo* and dataset: *Movies, Books, Eu*). Homogeneity of variance was confirmed by 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 ($M = 3.68$), $p < 0.001$. For the 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 among them. The interaction effects illustrate several differences amongst chart type. For instance, the majority of the users preferred bar chart, probably since it is familiar to most people. Another reason may be that it is easier to compare the values of several numbers at once using bar chart. Yet

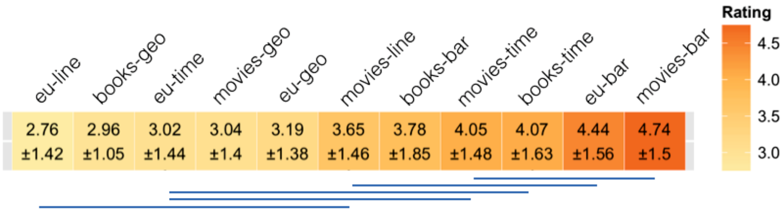


**Fig. 5.** Significant Interactions Chart Type / Dataset. The heat-map illustrates the mean score and standard deviation for each combination of *dataset-chart type* (1=completely disagree, 7=totally agree). The lines below show where differences begin to be significant. Note that due to its high variability,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

these results merely indicate 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 accommodate those user preferences better.

**Recommendation Quality.** At a glance, the results of our offline evaluation show significant improvements in the 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, in our case accumulated from ratings of individual users. Considering RD and MP, baseline algorithms should unveil whether the recommender systems can in general help with providing useful visualizations and whether the personalized approach improves the quality of the results, respectively.

For the comparison, we analyzed the top 3 recommendations, since our datasets relatively smaller than some commonly used datasets, such as BibSonomy and CiteULike [15]. The results of the evaluation are summarized in Table 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 indicate that the results are more accurate using *VizRec* CF than 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 can sort individual visualizations according to their relevance to 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).

**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

| | | Metric | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Alg.** | **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 |

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 with regard to providing relevant visualizations and their ranking order. This supports our main assumption that in terms of the wide variability in user preference ratings, the personalized approach performs better recommendations.

## 6 Discussion and Outlook

This work is based on the premise that the preference of a visual representation for a dataset is a 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 recommending visualizations can account for such variability in personal preferences and significantly improve the recommendations. 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 of our work is that it is based on the empirical evidence collected via a methodical study involving the general public. Our approach to generating and suggesting visualizations, the process of elicitation of users' preferences and the insights described in this paper are to the best of our knowledge, novel.

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 will investigate applying the measuring semantic similarity of the data attribute array to establish if a similar structure has been observed before and suggest from global ranking of other users. Furthermore, the investigation associated with our crowdsourced experiment is still ongoing. Although exploration of the relationship between quality of content features (such as textual description) and 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 attempting to determine the tasks that a user associates with the preferred visualizations. Furthermore, we will conduct an online evaluation to ascertain whether our recommender performs as expected, compared with the results of the offline experiment.

**Acknowledgments.** This work is funded by the EC 7th Framework project EEXCESS (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 with regard to the offline evaluation.

## References

1. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.A.: Indexing by latent semantic analysis. Journal of the American Society for Information Science **41**, 391–407 (1990)
2. Xiaoyuan, S., Khoshgoftaar, T.M.: A Survey of Collaborative Filtering Techniques. Adv. in Artif. Intell. (2009)
3. Bertin, J.: 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 2012 (2012)
5. Schelter, S., Owen, S.: Collaborative filtering with apache mahout. In: RecSysChallenge 2012 (2012)
6. Mackinlay, J.: Automating the design of graphical presentations of relational information. ACM Trans. Graph., 110–141 (1986)
7. Rahm, E., Bernstein, P.A.: A Survey of Approaches to Automatic Schema Matching. The VLDB Journal (2001)
8. Mackinlay, J., Hanrahan, P., Stolte, Ch.: Show Me: Automatic Presentation for Visual Analysis. IEEE Transactions on Visualization and Computer Graphics, 1137–1144 (2007)
9. Nazemi, K., Retz, R., Bernard, J., Kohlhammer, J., Fellner, D.: Adaptive semantic visualization for bibliographic entries. In: Bebis, G. (ed.) ISVC 2013, Part II. LNCS, vol. 8034, pp. 13–24. Springer, Heidelberg (2013)
10. Seffah, A., Donyaee, M., Kline, R., Padda, H.: Usability measurement and metrics: A consolidated model. Software Quality Journal, 159–178 (2006)
11. Ahn, J., Brusilovsky, P.: Adaptive Visualization of Search Results: Bringing User Models to Visual Analytics. Information Visualization, 167–179 (2009)
12. Zheng, X.S., Lin, J.J.W., Zapf, S., Knapheide, C.: Visualizing user experience through "Perceptual Maps": concurrent assessment of perceived usability and subjective appearance in car infotainment systems. In: Duffy, V.G. (ed.) HCII 2007 and DHM 2007. LNCS, vol. 4561, pp. 536–545. Springer, Heidelberg (2007)
13. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with mechanical turk. In: Proc. of CHI 2008 (2008)
14. Mutlu, B., Hoefler, P., Tschinkel, G., Sabol, V., Granitzer, M., Stegmaier, F.: Suggesting visualizations for published data. In: Proc. of VISIGRAPP 2014 (2014)

15. Trattner, C., Kowald, D., Lacic, E.: TagRec: Towards a Toolkit for Reproducible Evaluation and Development of Tag-Based Recommender Algorithms. ACM SIG-WEB Newsletter (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. Jaeschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in social bookmarking systems. Ai Communications **21**(4), 231–247 (2008)
18. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)