

Preference Thresholds Optimization by Interactive Variation

Mohamed Mouine and Guy Lapalme

RALI-DIRO Université de Montréal
CP 6128, Succ Centre ville
Montréal (Québec) H3C 3J7
{mouinemo,lapalme}@iro.umontreal.ca

Abstract. In order to customize the display of meteorological data for different users, we use clustering to group similar users. We compute a rate of similarity between the current user and all others in the same cluster. We use this rate for weighting users' preferences and then compute an average to be compared with a threshold to decide to display this parameter or not. The optimization of this threshold is also discussed.

Keywords: information visualization, artificial intelligence, clustering, personalization, graph.

1 Introduction

We want to generate customized meteorological reports for users of the Environment Canada website and adapt the reports to the needs and preferences of users. Users can change the layout according to their taste, preferences and needs. We want improve the user experience with the site by learning to *guess* users' preferences. To make this presentation closer to each user preferences, we aggregate *user types* and give each user a visualization close to what s/he wants. The principle that we consider most important is the customization of the display for each user taking into account his profile to be automatically detected without interference with his privacy and personal information. Law in Canada regarding the use of cookies in the government website is strict. We quote here an extract from the Office of the Information and Privacy Commissioner of Ontario¹: “*The collection of personal information by government organizations must be in accordance with section 38(2) of the Freedom of Information and Protection of Privacy Act (FIPPA). This section of the Act requires government organizations to collect only the personal information that is necessary for the administration of a government program*”. For this reason, we will not use cookies.

The information we have about a user (Location, preferred language, date, local time, season) does not allow us to predict user preferences. The first step of the method [9] is to archive (anonymously) interactions of the users with the

¹ <http://www.ipc.on.ca/English/Privacy-Policy/>

visualization and especially the final settings chosen. We take for granted that the final display as chosen by the user reflects her preferences. Users are then grouped according to their information similarity. We then determine the cluster corresponding to the user. Visualizations corresponding to users belonging to the same group as the user will be used to identify similar users' preferences. We weight their similarity rates based on the distance between the feature vectors. The preferences of users are weighted according to the rate of similarity. To decide what preferences be taken into account, we compare the mean of each parameter with a threshold to determine if a parameter is taken into consideration or not. In the next section we describe, for the problem of customizing visualization, the need to know the user profiles. In Section 3, we detail the problem. The solution to this problem is described in Section 4. And finally, a summary is presented in the last section.

2 Need for User Profiles

Graphics are useful to summarize and communicate numerical information found in weather reports. The task of generating graphics cannot be reduced to the encoding of a mass of information, it must take into account the decoding to be done by the user. If the user cannot decode the information, graphics generation has failed. It should take human perception into account. Robbins [11] explains how to create better visualizations by taking into account all the parameters (choice of chart type, the amount of information, choice of style attributes ...) and perception towards graphs.

The problem of information visualization is not necessarily raised by technical obstacles. Chen [2] studied 10 unsolved problems in this area, the first three ones being problems from a user-centric perspective:

usability the information contained in the visualization should meet the needs of users; this explains the increasing number of usability studies and evaluations of visualization [3,10,4,16];

perception the principles of perception were incorporated into rendering algorithms in order to optimize rendering computation and produce an *ideal* visualization human point of view and not from a machine standpoint [7];

prior knowledge of the user can be considered a parameter of the second problem. In general, users need two types of knowledge before understanding the message conveyed in the information displayed: how to use a system for displaying information and how to interpret the content.

There are two types of user profiles: determined by a user who sets preferences and needs or built automatically using several techniques such as history, behavior, rules of associations, classification techniques and algorithms of clustering.

We want the system to learn the preferences of the users. Reports generated for each one are interactive. It is very important that users can modify visualization according to their tastes and needs[14]. A study was conducted to describe why users should interact[15]. They are based on different intentions of users

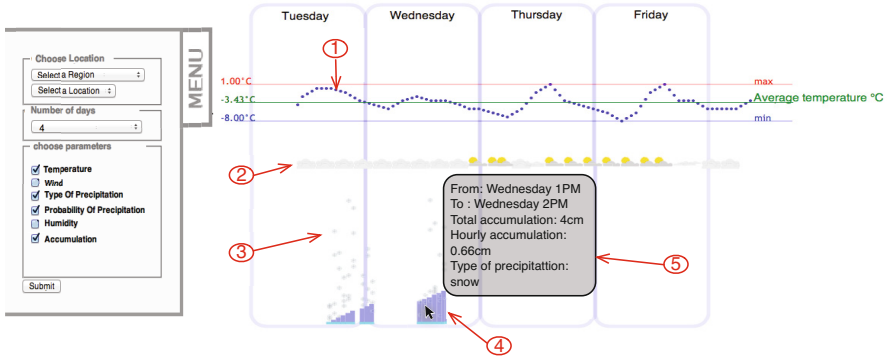


Fig. 1. Generated visualization showing some types of information that can be selected by the user. Circled numbers are added here for reference purposes. ① Temperature: how temperature is displayed gives an idea of the trend. Maximum and minimum line show limits of the temperature for the chosen period. ② Cloud cover. ③ TOP (Type Of Precipitation): rain or snow. Quantity displayed is proportional to the quantity provided. ④ Accumulation: shows the accumulation zone and total rainfall (mm) or snow (cm). ⑤ The user may have more details about aspect by putting the cursor over it (tooltip). Using the preference menu (in left), the user can select a region (Province) and location (city), modify number of the days to display and which parameter to display.

and introduced a list of categories (Select, Explore, Reconfigure, Encode, Abstract/Elaborate, Filter and Connect). The majority of these principles are used in our visualization.

3 Problem Statement

Environment Canada (EC) produces an enormous amount of weather information on a continuous (26 Mb twice a day). This information is used to provide Canadians with up to date information on weather conditions. Problem is that this amount of information must be summarized to be displayed. Weather reports prepared in advance may not contain all the information that all users hope to find. Already, more than 1,000 weather reports presenting the weather in Canada are issued twice per day. First solution is to prepare more weather reports in advance. But most of these reports will not likely be used. We thus propose to build a weather reports generator to answer users on demand in either English or French. Each report must meet the specific needs of the user for which it was generated. In order to summarize and analyze large amounts of information, we have presented a method that automatically generates a visual report (graph, image, text...).

To allow users to select the information to display, our system should be interactive. We have already proposed [9] generating reports based on the similarity of user profiles. Clustering was used to group similar users to produce a visualization corresponding to the needs of most of them. But visualization generation was based on an arbitrary threshold to decide the parameters to be taken into account. In this paper, we study the robustness of this threshold and present a sensitivity analysis.

4 Approach

To generate the visualization (see Fig. 1) based on the user's profile to learn more about their preferences, our system saves each final configuration chosen by a user. The system uses this data, for future customization of the weather reports according to user profiles. When a new report is generated, it is based on several parameters including: the user's profile and choice of former users similar to the current user. Clustering is used for determining the settings of similar users. The distance between the current user and others in a cluster will be used to compute the similarity. A user can modify the visualization settings using a menu (left in Fig. 1).

4.1 Clustering

K-means [5] is a simple algorithm for clustering. It classifies a set of data among a fixed number of clusters. The main idea is to define K centroids, one for each cluster. K being between 10 and 20 corresponding roughly to the number of provinces in Canada. The next stage is to take each user in our database for which a visualization has already been generated and to associate her with the nearest centroid. Variables of user profiles belonging to the same category represents the features for the clustering process.

4.2 Similarity Computing

Profile data and preferences for each user are saved (anonymously). We assign a weight, depending on the significance, to each criterion for users' profiles (see Table 1). The rate of similarity is given by:

$$R_j = \sum_{i=1}^n C_i * W_i \text{ with } j \in [1..m]$$

where: R is the rate of similarity, m number of user, n number of criterion, C is the criterion taken into account by the user, and W is its weight.

An example of the calculated similarity is shown in table 2. similarity rate is used to weight the preferences.

Table 1. Criterion weight

critereon	weight
Lang	0.2
Province	0.3
City	0.3
Season	0.1
Period	0.1

Table 2. Similarity rates

	Lang	Province	city	season	period	% of similarity
current user	fr	Qc	Montreal	winter	morning	-
user1	fr	Qc	Quebec	autumn	Afternoon	50%
user2	fr	Qc	Longueuil	winter	morning	70%
user3	en	Qc	Gaspé	autumn	morning	40%
user4	fr	Qc	Montreal	spring	morning	90%
user5	fr	Qc	Laval	autumn	evening	50%
user6	fr	Qc	Sutton	summer	evening	50%
user7	en	Qc	Delson	summer	morning	40%
user8	en	Qc	Montreal	spring	Afternoon	60%
user9	fr	Qc	Gatineau	winter	morning	70%
user10	fr	Qc	Montreal	winter	morning	100%
average						62%

4.3 Computing the Preferences

To predict the preferences of the current user, we rely on preferences data recorded in our database of similar user. In the database, we register number of day fixed in visualisation by user and for each parameter 1 if used, 0 if not. Preferences of each user is weighted by the level of similarity with the current user.

$$P_{wj} = P_j * R_j$$

4.4 Threshold Analysis

Threshold analysis is often used in multicriteria decision methods [1,6,12,13] in which a decision maker sets thresholds arbitrarily and then evaluates their robustness using sensitivity analysis [6,8] to modify their values. If a small modification results in a large change in the results, the threshold is considered sensible and the decision is not robust.

Our approach is inspired by this type of sensitivity analysis. A threshold determines whether to use the parameter or not. We analyzed its robustness to the user feedback with the visualization. We consider that:

- if the user changes the display and adds a new parameter that has not been used, that means that the threshold was very strict and led to the exclusion of the parameter in question; its value should be decreased;

Table 3. Result of visualization interaction

	temperature	wind	TOP	POP	humidity	accumulation
added	-	-	-	-	20%	40%
removed	0%	20%	10%	10%	-	-
kept	100%	80%	90%	90%	80%	60%
average	0.57	0.41	0.38	0.34	0.32	0.23
old threshold	0.33					
new threshold	0.33	0.35	0.34	0.34	0.31	0.29
retained (new)	yes	yes	yes	yes	yes	no

- if the user modifies the visualization and regenerates it by removing one or more parameters, it means that the value of the threshold for a removed parameter is very low and its value should be increased;
- if the user does not interact, the threshold is robust and corresponds to the wish of the user.

To assess the robustness of visualization, we analyze user interactions and compute the new thresholds according to the following formula:

1. Vary the threshold ± 0.01 for each 10 % of dissatisfied users.
2. Redo the experiment (Table 3).
3. Repeat step 1 and 2 until we haven't more dissatisfaction superior to 10 %.

Table 4 shows the final result of our example. We stop the threshold variation after 3 iterations because we did not have any dissatisfaction superior to 10 %.

Table 4. Final result of visualization interaction and threshold variation (after 3 iterations)

	temperature	wind	TOP	POP	humidity	accumulation
average	0.57	0.41	0.38	0.34	0.32	0.23
Final threshold	0.33	0.36	0.34	0.35	0.30	0.30
retained (new)	yes	yes	yes	yes	yes	no

It can be seen that the thresholds have been revised downward or upward according to users' dissatisfaction. Although the threshold of the parameter *temperature* that satisfies users has not been changed. The thresholds for *wind*, *TOP* and *POP* were increased because users were not satisfied with the visualization generated for them. The threshold parameters *humidity* and *accumulation* were lowered because multiple users wanted to have it in their visualizations that the former threshold does not allow.

5 Conclusion

Our job is to customize a visualization according to the profile of the user. We have little information about the users using our system. We rely on the history

of user preferences similar to our current user. We set a minimum threshold for deciding which parameters will be used in the visualization. We use the same threshold to evaluate the level of satisfaction of the user. The interaction of the user with the visualization used for evaluation improves our system to better match preferences of users according to their profile.

References

1. Brans, J.P., Vincke, P., Mareschal, B.: How to select and how to rank projects: The promethee method. *European Journal of Operational Research* 24(2), 228–238 (1986)
2. Chen, C.: Top 10 unsolved information visualization problems. *IEEE Computer Graphics and Applications* 25(4), 12–16 (2005)
3. Elmqvist, N., Yi, J.S.: Patterns for visualization evaluation (2012)
4. Goldberg, J., Helfman, J.: Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization* 10(3), 182–195 (2011)
5. Hartigan, J.A., Wong, M.A.: Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, 100–108 (1979)
6. Maystre, L.Y., Pictet, J., Simos, J.: Méthodes multicritères ELECTRE: Description, conseils pratiques et cas d'application à la gestion environnementale, vol. 8. PPUR (1994)
7. McNamara, A., Mania, K., Gutierrez, D.: Perception in graphics, visualization, virtual environments and animation. In: *SIGGRAPH Asia 2011 Courses*, p. 17. ACM (2011)
8. Mena, S.B.: Une solution informatisée à l'analyse de sensibilité d'electre iii. *Biotechnol. Agron. Soc. Environ.* 5(1), 31–35 (2001)
9. Mouine, M., Lapalme, G.: Using clustering to personalize visualization. In: *2012 16th International Conference on Information Visualisation (IV)*, pp. 258–263. IEEE (2012)
10. Plaisant, C.: The challenge of information visualization evaluation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp. 109–116. ACM (2004)
11. Robbins, N.B.: *Creating more effective graphs*. Wiley-Interscience (2012)
12. Roy, B.: Classement et choix en présence de points de vue multiples (la méthode electre). *RIRO* 2(8), 57–75 (1968)
13. Roy, B.: Electre iii: Un algorithme de classement fondé sur une représentation floue des préférences en présence de critères multiples. *Cahiers du CERO* 20(1), 3–24 (1978)
14. Ward, M., Grinstein, G., Keim, D.: *Interactive data visualization: Foundations, techniques, and applications*. AK Peters, Ltd. (2010)
15. Ware, C.: *Information visualization: Perception for design*. Morgan Kaufmann (2012)
16. Weaver, C.: Look before you link: Eye tracking in multiple coordinated view visualization. In: *BELIV 2010: BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, p. 2 (2010)