What’s going on in my city? Recommender systems and electronic participatory budgeting

Iván Cantador¹, María E. Cortés-Cediel², Miriam Fernández³, Harith Alani³
¹Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain
ivan.cantador@uam.es
²Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid
Pozuelo de Alarcón, Spain
mcorte04@ucm.es
³Knowledge Media Institute
The Open University
Milton Keynes, United Kingdom
{m.fernandez, h.alani}@open.ac.uk

ABSTRACT
In this paper, we present electronic participatory budgeting (ePB) as a novel application domain for recommender systems. On public data from the ePB platforms of three major US cities – Cambridge, Miami and New York City–, we evaluate various methods that exploit heterogeneous sources and models of user preferences to provide personalized recommendations of citizen proposals. We show that depending on characteristics of the cities and their participatory processes, particular methods are more effective than others for each city. This result, together with open issues identified in the paper, call for further research in the area.

1 INTRODUCTION
Participatory budgeting (PB) is a democratic deliberation and decision-making process in which citizens decide how to spend certain municipal or public budgets. It allows citizens to inform about issues and problems on a wide range of subject areas in a city – e.g., housing, public safety, education, health, transportation and environment–, and propose, debate and support/vote for spending ideas and projects aimed to address such problems.

Since its original invention in Porto Alegre, Brazil, in 1988 [12], PB has gained much popularity. As for 2014, PB had spread to over 1500 cities around the world [4], and in recent years, such number has increased significantly with the adoption of ICTs and online participatory processes [29][32]. In this context, electronic participatory budgeting (ePB) tools constitute an added dimension aimed to support, improve and innovate traditional (offline) participatory methods – e.g., meetings, committees and councils– with virtual (online) services [20], and have been shown to provide benefits to citizens [24][32], such as perceiving quality and transparency, and saving time. Despite these benefits, ePB has not transformed participation as much as expected, and the levels of citizen e-participation are still low. As suggested by [32], part of the problem may arise from website design. In [23], the OECD stated a number of challenges aimed to increase citizen engagement in e-participation. One of these challenges is the development of technology able to support a citizen to actively participate by giving her the electronic means to find others that share similar opinions and points of view. Another key aspect to encourage people to use e-participation rather than traditional forums is saving time. However, in general, ePB platforms of large cities do have hundreds, even thousands, of budgeting proposals and associated comments and debates, and provide very limited search and filtering functionalities. Moreover, when creating a budgeting proposal, a citizen should be aware of similar or related ideas or projects, so she could better define the proposal or find the opportunity to collaborate with others. It is in these scenarios where recommender systems have new, challenging opportunities.

As discussed in [11], in the research literature there are few studies on recommender systems for the e-governance domain, and most of them present very simple, poorly evaluated recommendation approaches, focused on the e-information level [1][2][3]. At the e-consultation and e-participation levels, there are seminal works on recommending political candidates [28] and citizen comments and opinions [16][19][21]. In [7], we preliminary experimented on recommending citizen proposals with data from the ePB platform of Madrid (Spain), showing that a social tag-based recommendation component allowed increasing the coverage and diversity of collaborative filtering. Differently to that work, in this paper, we present ePB as a novel application domain for recommender systems, proposing and evaluating a number of recommendation methods that exploit heterogeneous data from the ePB platforms of three major US cities. Based on our empirical results, we provide valuable insights about which recommendation solutions may be more effective depending on characteristics of the cities and their participatory processes. Moreover, to encourage future work in the area, we make available the generated datasets, and identify several open research issues.

2 DATASETS
The datasets used in this paper contain public data available online in the electronic participatory budgeting platforms¹ of Cambridge (MA), Miami (FL) and New York City (NY). The data can be accessed via web services².

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. © 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-5901-6/18/10...$15.00
https://doi.org/10.1145/3240333.3240391

¹ Electronic participatory budgeting platforms:
Miami, http://www.publicspacechallenge.org
New York City, http://ideas.pbnyc.org

² Web services to access the data of the e-participatory budgeting platforms:
New York City, https://shareaboutsapi.poepublic.com/api/v2/pbnewyork/datasets
In particular, for each of the above cities, we downloaded the data generated in their annual ePB processes from 2014 to 2017. Hence, we built a total of 12 datasets, which are described in Table 1. The datasets have the same structure, with three main entities: proposals, comments and supports. A proposal is a project, initiative or idea commonly submitted by a citizen to an ePB platform, and is composed of a title and description, a timestamp, and a geographic location with longitude and latitude coordinates. A comment is a short text uploaded by a citizen giving her opinion about certain proposal created by other person. Finally, a support is an explicit positive vote given by a citizen to other’s proposal. Both comments and supports have associated a timestamp. From now on, we consider as users those citizens who provided one or more supports, as unary ratings the citizens’ supports, and as items those proposals with one or more supports.

As shown in Table 1, the rating sparsity levels of the datasets are similar and around 99.4, comparable to values of standard recommender systems datasets. The average number of ratings per user is relatively low – being higher in the Cambridge datasets-, whereas the average number of ratings per item is moderate – being higher in the Miami datasets. The table also summarizes location information. It reports the items average and maximum distances (in km) to the ’centroid’ items of the datasets. These values evidence the relatively small area and high item closeness of Cambridge. The table also reports average closeness of the rated items per user, by means of the average and maximum user radius, defined as the average distance between a user’s rated items and her ’centroid’ rated item. As we shall show, these characteristics will allow explaining why the evaluated recommendation methods perform differently among the considered cities/datasets.

In addition to rating and location data, we also investigated the exploitation of textual data for recommendation purposes. In particular, we used the TextRazor tool to perform a semantic annotation of the items titles and descriptions. Table 2 shows an example of a citizen proposal and its generated annotations. Such annotations are of three types –entities and topics (from Wikipedia, Freebase and Wikidata knowledge bases), and categories (from the IPTC taxonomy)–, and have associated confidence scores and their corresponding URLs in the source repositories.

### Table 1: Descriptions of the generated datasets. Standard deviations within parentheses; distances in kilometers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>First item date</th>
<th>Last item date</th>
<th>#users #items #ratings</th>
<th>Rating sparsity</th>
<th>Avg. ratings/user</th>
<th>Avg. ratings/item</th>
<th>Avg. item distance</th>
<th>Max. item distance</th>
<th>Avg. user radius</th>
<th>Avg. max. user radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cambridge-2014</td>
<td>2014-11-26</td>
<td>2015-01-02</td>
<td>874 348 3518</td>
<td>98.843</td>
<td>4.0 (10.1)</td>
<td>10.1 (15.3)</td>
<td>1.6 (6.9)</td>
<td>4.2</td>
<td>0.4 (0.6)</td>
<td>4.5 (1.2)</td>
</tr>
<tr>
<td>Cambridge-2015</td>
<td>2015-07-31</td>
<td>2015-09-02</td>
<td>528 252 1761</td>
<td>98.676</td>
<td>3.3 (5.2)</td>
<td>7.0 (8.9)</td>
<td>1.5 (1.0)</td>
<td>3.9</td>
<td>0.4 (0.7)</td>
<td>4.3 (1.2)</td>
</tr>
<tr>
<td>Cambridge-2016</td>
<td>2016-06-01</td>
<td>2016-08-01</td>
<td>856 378 2546</td>
<td>99.213</td>
<td>3.0 (7.6)</td>
<td>6.7 (9.4)</td>
<td>1.5 (1.0)</td>
<td>3.9</td>
<td>0.3 (0.5)</td>
<td>4.2 (1.0)</td>
</tr>
<tr>
<td>Cambridge-2017</td>
<td>2017-06-01</td>
<td>2017-08-01</td>
<td>1575 550 6412</td>
<td>99.260</td>
<td>4.1 (9.5)</td>
<td>11.7 (16.0)</td>
<td>1.7 (1.0)</td>
<td>4.4</td>
<td>0.4 (0.6)</td>
<td>4.7 (1.2)</td>
</tr>
<tr>
<td>Miami-2014</td>
<td>2013-08-14</td>
<td>2014-04-09</td>
<td>4387 294 6294</td>
<td>99.512</td>
<td>1.4 (2.3)</td>
<td>21.4 (40.6)</td>
<td>7.2 (6.5)</td>
<td>29.1</td>
<td>0.4 (1.5)</td>
<td>27.8 (2.6)</td>
</tr>
<tr>
<td>Miami-2015</td>
<td>2015-03-04</td>
<td>2015-04-02</td>
<td>6947 251 7921</td>
<td>99.546</td>
<td>1.1 (0.9)</td>
<td>31.6 (79.9)</td>
<td>6.8 (6.5)</td>
<td>30.6</td>
<td>0.2 (1.0)</td>
<td>27.8 (1.7)</td>
</tr>
<tr>
<td>Miami-2016</td>
<td>2016-03-15</td>
<td>2016-04-22</td>
<td>6907 328 8640</td>
<td>99.619</td>
<td>1.3 (1.7)</td>
<td>26.3 (77.5)</td>
<td>9.5 (8.3)</td>
<td>28.7</td>
<td>0.5 (2.3)</td>
<td>27.3 (2.8)</td>
</tr>
<tr>
<td>Miami-2017</td>
<td>2017-03-07</td>
<td>2017-04-07</td>
<td>6341 327 7886</td>
<td>99.620</td>
<td>1.2 (1.6)</td>
<td>24.1 (62.5)</td>
<td>8.4 (7.6)</td>
<td>39.9</td>
<td>0.5 (1.9)</td>
<td>35.2 (2.8)</td>
</tr>
<tr>
<td>NewYorkCity-2014</td>
<td>2014-09-12</td>
<td>2014-12-03</td>
<td>1861 403 2686</td>
<td>99.642</td>
<td>1.4 (2.0)</td>
<td>6.7 (25.2)</td>
<td>9.7 (5.7)</td>
<td>24.7</td>
<td>0.2 (0.9)</td>
<td>18.1 (1.6)</td>
</tr>
<tr>
<td>NewYorkCity-2015</td>
<td>2015-09-29</td>
<td>2015-11-16</td>
<td>2365 542 5497</td>
<td>99.727</td>
<td>1.5 (4.0)</td>
<td>6.5 (25.0)</td>
<td>9.6 (4.8)</td>
<td>23.5</td>
<td>0.2 (0.8)</td>
<td>21.5 (0.5)</td>
</tr>
<tr>
<td>NewYorkCity-2016</td>
<td>2016-08-23</td>
<td>2016-10-01</td>
<td>2253 643 3569</td>
<td>99.754</td>
<td>1.6 (2.0)</td>
<td>5.6 (14.7)</td>
<td>10.0 (5.7)</td>
<td>26.3</td>
<td>0.3 (1.1)</td>
<td>22.1 (1.8)</td>
</tr>
<tr>
<td>NewYorkCity-2017</td>
<td>2017-08-08</td>
<td>2017-10-14</td>
<td>3045 1065 5238</td>
<td>99.829</td>
<td>1.7 (3.2)</td>
<td>5.2 (22.2)</td>
<td>8.3 (4.6)</td>
<td>24.4</td>
<td>0.2 (0.9)</td>
<td>23.6 (1.8)</td>
</tr>
</tbody>
</table>

To allow reproducibility, we make available all the datasets with their users anonymized. The datasets contain both the original proposals and citizens’ comments and supports, and the metadata we generated, namely the item semantic annotations, and semantic- and location-based user/item profiles (Section 3.2).

### Table 2: Example of a proposal and its semantic annotations

<table>
<thead>
<tr>
<th>Proposal id: 1</th>
<th>Dataset: Cambridge-2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Created time: 2014-11-26, 20:19:01</td>
<td></td>
</tr>
<tr>
<td>Location: [longitude = -71.094882, latitude = 42.360129]</td>
<td></td>
</tr>
<tr>
<td>Title: Diagonal crosswalks</td>
<td></td>
</tr>
<tr>
<td>Description: Lots of pedestrians here, rather than waiting twice you could cross diagonally.</td>
<td></td>
</tr>
<tr>
<td>Entities: pedestrian crossing</td>
<td></td>
</tr>
<tr>
<td>Topics: pedestrian infrastructure, streets and roads, urban planning</td>
<td></td>
</tr>
<tr>
<td>Categories: economy, business and finance &gt; economic sector &gt; transport &gt; traffic</td>
<td></td>
</tr>
</tbody>
</table>

3 RECOMMENDATION METHODS

In this section, we present the recommendation methods we evaluated on the built participatory budgeting datasets. The methods are of different types –namely collaborative, content-based and hybrid–, and exploit the variety of available data: the citizens’ supports, and the proposals text descriptions and geographic locations. We implemented the methods on top of the RankSys framework [31] for reproducibility. Since citizen supports represent unary ratings, we focused on the top-N recommendation task, limiting the size of the methods recommendation lists to 50.

3.1 Collaborative methods

The collaborative methods recommend a user (citizen) items (proposals) rated (supported) by like-minded users. The methods compute similarities between users and items based on rating patterns. Specifically, we considered the well-known k-nearest neighbor (k-NN) heuristics and matrix factorization models.

3.1.1 k-nearest neighbors methods

We evaluated both the user- and item-based k-NN heuristics [22]. The user-based approach exploits rating-based similarities between users to create neighborhoods, which are used to estimate scores for a (user, item) pair. The item-based approach works in a similar way, but computing similarities between items. Since the datasets have unary ratings, we used the cosine similarity. We tested several neighborhood sizes: k = 5, ..., 50, in

---

1. Some of the downloaded locations did not belong to the corresponding cities; to identify and filter out the wrong cases, we used the Google Maps API services.
2. TextRazor semantic annotation tool, [https://www.textrazor.com](https://www.textrazor.com)
steps of 5. We will refer to these methods as \textbf{ub} and \textbf{ib}. For a particular instance of the methods, the used \(k\) value will be included in the method name, as \textbf{ub} \(k\) and \textbf{ib} \(k\), e.g., \textbf{ub}5 for referring to user-based K-NN with neighborhood size \(k = 5\).

### 3.1.2 Matrix factorization methods

We also evaluated a matrix factorization model for collaborative filtering, which addresses the rating sparsity by compressing the user-item matrix into a low dimensional representation in terms of latent factors. More specifically, we used the variation proposed in [14], since it is well suited for implicit feedback datasets; we recall that in our datasets the user-item matrices contain unary ratings. From now on, we will refer to this method as \textbf{mf}. Similarly to the \(k\)-NN heuristics, in the experiments, we tested several numbers of latent factors \(K\), ranging from 5 to 100 in steps of 5, and using the notation \textbf{mf} \(K\), e.g., \textbf{mf}15 to refer to the matrix factorization method with \(K = 15\) factors.

### 3.2 Content-based methods

The content-based methods recommend a user (citizen) items (proposals) “similar” to those she liked (positively supported). The similarity between users and items is computed on profiles built from either textual or location item information. Mining the (positive/negative) opinions of the users’ text comments may be investigated in the future.

#### 3.2.1 Text feature-based methods

In these methods, user preferences and item attributes correspond to text features \(f_i\) (e.g., keywords and categories) extracted from the items titles and descriptions, and recommendations are generated by means of user and item similarities in the text feature space. Formally, an item \(i_u\)’s profile consists of a vector \(i_u = \{w_{n,1}, w_{n,2}, ..., w_{n,l}\} \in \mathbb{R}^l\), where \(w_{n,l}\) denotes the relative relevance (weight) of feature \(f_l\) for \(i_u\), and \(l\) is the number of existing features. As done in [8], we considered several techniques to compute the weights \(w_{n,l}\), namely binary, TF-IDF and BM25. Empirically we observed that recommendation methods using TF-IDF and BM25 weights achieved very similar results, and were clearly superior to those using binary weights. For this reason, the performance values of the text feature-based methods reported in this paper correspond to the TF-IDF weighting scheme. Similarly, a user \(u_m\)’s profile is represented as a vector \(u_m = \{w_{m,1}, w_{m,2}, ..., w_{m,l}\} \in \mathbb{R}^l\), where \(w_{m,l}\) denotes the relative relevance (weight) of feature \(f_l\) for \(u_m\), computed by aggregating the weights of \(f_l\) in the profiles of (training) items rated by \(u_m\), as \(w_{m,l} = \frac{1}{|R(u_m)|} \sum_{i \in R(u_m)} eR(u_m) w_{n,l,i}\), where \(R(u_m)\) is the set of training relevant items (i.e., supported proposals) of user \(u_m\).

The recommendation score of an item \(i\) for a target user \(u\) is then computed as the cosine similarity \(score(u, i) = \cos(u, i)\). Other similarities could be considered in the future.

In the experiments, we used as textual features the three types of semantic annotations introduced in Section 2: entities, topics and categories. Exploiting entities achieved the best performing results. We will refer to this method as \textbf{cb-ent}.

#### 3.2.2 Location-based method

In this method, users and items are represented with geographic location coordinates. An item profile \(i\) consists of a 2-dimensional vector with the item longitude and latitude values. A user’s profile \(u\) has the same vector representation, where the longitude (latitude) value is the average of the longitude (latitude) values of the user’s training items. Then, the recommendation score of an item \(i\) for a target user \(u\) is computed as \(score(u, i) = 1 - \frac{\text{dist}(\mathbf{u}, \mathbf{i})}{\text{dist}(\mathbf{u}, \mathbf{0})}\), where \(\text{dist}(\mathbf{u}, \mathbf{i})\) is the Haversine distance between the location vectors of \(u\) and \(i\), and \(C\) is a constant (greater than the dataset maximum distance; see Table 1) to normalize scores to the [0,1] range. We will refer to this method as \textbf{cb-loc}.

### 3.3 Hybrid methods

The hybrid methods jointly exploit rating, text and location data, by combining the collaborative and content-based methods presented in Sections 3.1 and 3.2. According to Burke’s taxonomy of hybrid recommender systems [6], the evaluated methods belong to the feature combination and weighted hybridization categories.

#### 3.3.1 Feature combination hybrid methods

These methods apply the collaborative filtering \textbf{ub} (or \textbf{ib}) method using a content-based user (or item) similarity of some of the \textbf{cb} methods. They will be referred as \textbf{cbub} (or \textbf{cbib}). We leave for future work evaluating other methods, e.g., factorization machines [25], to jointly exploit ratings and content-based features.

Without considering the content-based similarities based on topics and categories due to their relatively low recommendation performance with respect to the entity-based similarity, in the experiments section, we will only report results for the entity- and location-based methods, having \textbf{cbub-ent}, \textbf{cbib-ent}, \textbf{cbub-loc} and \textbf{cbib-loc} hybrid methods. As in previous cases, the name of each of them will indicate the neighborhood size used in \textbf{ub} or \textbf{ib}, e.g., \textbf{cbub10-ent}.

#### 3.3.2 Weighted hybrid methods

Each of these methods consists of an ensemble of 2 recommenders. Ensembles with larger number of recommenders remain open for future investigation. To integrate the recommenders of an ensemble, as recently proposed in [30], we tested search fusion strategies from the information retrieval area. In particular, we evaluated the CombMAX (for each target user-item pair, it retrieves the maximum of the two recommenders scores), CombMIN, CombSUM and CombMNZ (CombSUM × number of non-zero recommender scores) score-based fusion techniques [17], and the Reciprocal Rank Fusion (RRF) [10] rank-based fusion technique. We also evaluated the weighted version of CombSUM, called wCombSUM, which computes a linear combination of the ensemble recommenders, using weights \(\alpha = 0.1, 0.2, ..., 0.9\).

### 4 EXPERIMENTS

In the experiments, we evaluated the collaborative, content-based and hybrid recommendation methods presented in Section 3, each of them with several parameter settings. We also included \textbf{pop}, a popularity-based recommender, as baseline. Due to lack of space, we only present the best performing configurations, and others of interest for comparison purposes. For the same reason, we do not present the results on each of the 12 datasets, but the average results from the 4 datasets of each city. We did not observe significant differences between results at city and dataset levels.

On a particular dataset, the recommendation performance results were averaged by 5-fold cross validation, keeping 80% of the ratings for training and the remaining 20% for test in each run.
The rating sets were split following the TestItems methodology [5]. Moreover, to analyze the methods performance, we considered a variety of ranking-based metrics, as well as coverage and diversity metrics. The metrics, implemented in the RiVal framework [26], were precision, recall, $F_1$, $nDCG$ (normalized Discounted Cumulative Gain), USC (User Space Coverage) and ISC (Item Space Coverage).

Table 3 shows a summary of the achieved performance values. An interesting result is the difference in the methods that best performed in each city, which could be explained analyzing the dataset characteristics reported in Table 1. For Cambridge, the ub method was the best performing in terms of $F_1$ and $nDCG$. This could be due to the fact that in its ePB processes, the average number of supports by citizen was relatively high. Besides, the city is relatively small, and its citizen proposals are located quite close, as shown by its average item and user radius. In fact, the location-based methods achieved the highest ISC values.

For Miami, in contrast, the hybrid cbub-ent method, which exploits semantic entity-based user similarities in a collaborative filtering fashion, was the best performing in terms of $P_1$ and $nDCG$. In the city ePB processes, there was a large number of participants. However, the number of supports by citizen was relatively low. Since, on average, each proposal had a relatively high number of supports, content-based similarities were more effective. Location-based similarities, in contrast, were not effective, since the city has relatively large geographic dispersion among proposals. Nonetheless, again, the location-based metrics achieved the highest ISC. Finally, it has to be noted that, probably due to the larger number of users, matrix factorization methods showed good performance in comparison to the other cities.

Finally, for New York City, the hybrid cbib-loc methods, which apply item-based collaborative filtering with location-based item similarities, were the best performing in terms of $F_1$ and $nDCG$. We believe that this result may be related to the fact that the PB processes in New York City were done separately in each district of the city. Hence, the relevant proposals for a citizen tend to be those located in her neighborhood.

In the table, we do not show the performance values of the weighted hybrid methods, which performed worse than the feature combination methods. In general, the score-based CombMIN method using cbub was the best performing in terms of $F_1$ (0.082, 0.211 and 0.183 for the three cities respectively), whereas CombMAX and CombMNZ using cbib achieved the highest ISC. CombSUM and wCombSUM showed the best recall and high ISC values, but very low precision values; and the rank-based ComBRRF did not obtain significant competitive results.

5 CONCLUSIONS AND FUTURE WORK
In this paper, we have introduced e-participatory budgeting as a novel application domain for recommender systems. By evaluating a variety of collaborative, content-based and hybrid recommendation methods on real data from ePB processes of three major US cities, we have shown that the selection of a recommendation solution should take into account characteristics of the cities and their participatory scenarios. Specifically, we have observed that the average degree of participation by citizen, the relative geographic closeness of the citizens’ proposals, and particular participatory restrictions, such as proposal voting at district level, were key elements. We believe that demographic, socio-cultural, political and economic aspects could also be considered, and for such purpose, government (Linked) Open Data [13][15][27] and citizen-generated contents in social media [9][18] may be exploited.

In the ePB context, we could go beyond recommending citizen proposals. Among others, we envision recommendation of user communities (e.g., citizens who share the same interests, needs or opinions) and recommendation of user comments (e.g., to find out arguments in favor or against certain initiative) as very promising open tasks. Moreover, as discussed in [11] for e-governance applications, recommendations could be targeted not only to citizens, but also to other stakeholders, such as government managers, politicians, public organizations, and businesses. In each case, defining the relevance of generated recommendations will have peculiarities to investigate.
ACKNOWLEDGMENTS
This work was supported by the Spanish Ministry of Economy, Industry and Competitiveness (TIN2016-80630-P).

REFERENCES