Personalized recommendations in e-participation: Offline experiments for the 'Decide Madrid' platform

Iván Cantador[1], Alejandro Bellogín[1], María E. Cortés-Cediel[2], Olga Gil[2]

[1]Escuela Politécnica Superior
Universidad Autónoma de Madrid
Spain
{ivan.cantador, alejandro.bellogin}@uam.es

[2]Facultad de Ciencias Políticas y Sociología
Universidad Complutense de Madrid
Spain
{mcorte04, olgagil}@ucm.es

ABSTRACT

In e-participation platforms, citizens suggest, discuss and vote online for initiatives aimed to address a wide range of issues and problems in a city, such as economic development, public safety, budges, infrastructure, housing, environment, social rights, and health care. For a particular citizen, the number of proposals and debates may be overwhelming, and recommender systems could help filtering and ranking those that are more relevant. Focusing on a particular case, the 'Decide Madrid' platform, in this paper we empirically investigate which sources of user preferences and recommendation approaches could be more effective, in terms of several aspects, namely precision, coverage and diversity.

KEYWORDS

recommender systems, e-participation, citizen participation, urban planning, smart cities, social tagging, folksonomy

1 INTRODUCTION

Plans related to smart cities are drafted to mitigate and remedy urban challenges and problems in a sustainable way through innovation [2]. They commonly entail a strong integration of information and communications technologies (ICTs) into planning, operations and management. In addition, interaction and participation of citizens and residents are common when digital media are set in place.

In this context, e-participation platforms are commonly used in smart cities in order to upgrade the relations among stakeholders in civil society –including citizens, residents, firms and the local government itself–, and perform as a mechanism to put the citizens at the center of the process [1]. This all reflects that at the core of the smart city there is an attempt to develop new forms of collaboration and urban development through ICTs [17].

One of the pillars of smart city plans is governance. Local governance within the smart cities encapsulates collaboration, cooperation, partnerships and participation, which might become success factors in the city [18]. Hence, when we talk about governance in the smart city, we refer to the adoption of inclusive and participatory processes that allow for the deliberation of social actors [12], and their implication in the formulation of public policies. Governance thus involves multiple interactions between the different stakeholders, public institutions, citizens, researcher institutes, and firms.

Adding ICTs to governance might facilitate better exchange of information and sustainability in such interactions [9]. In particular, the modes of participatory interactions involving local governments with citizens and residents might be characterized as three-fold. In each mode, ICTs are applied and used to a different extent. On the first mode, interactions reach the level of information; on a second mode, interactions would involve e-consultation; and on the third mode, interactions would entail e-participation, which requires the greater degree of involvement.

Moreover, according to the scientific literature, citizenship participation in public affairs is related to three relevant aspects. First of all, the improvement of democratic legitimacy in increasingly complex societies. Secondly, the improvement of effectiveness and efficiency of public policies and, lastly, the development of an active citizenship through experiences of participation [4].

Reports produced by the OECD [20][21] have covered the issue of increasing citizen participation in politics through ICTs. The theoretical frameworks in these reports focus on four objectives [15]: (i) promoting participation through a widened audience; (ii) seeking participation from citizens and residents through ICTs to leap forward on technical and communication skills; (iii) facilitating relevant information through a more accessible format for audiences; and (iv) engaging in deliberative process with an ample majority. These objectives have been assumed –to a different extent– by governments using ICTs as a way to increase public participation and the possibility to enhance the benefits for citizens [11].

In so doing, there are convenient strategies to identify needs of the citizenship, and to provide the tools for participation [23]. E-participation platforms are institutionalized mechanisms that allow the citizens to participate in democratic life, and thus be an active part of government plans and decisions [7]. However, these platforms have a number of problems, such as an excess of information, and a requirement for customization.

For a particular citizen, the number of initiatives and discussions in an e-participation platform may be overwhelming. Addressing this situation, recommender systems could filter and rank the initiatives and discussions that are more relevant for the citizens based on previous explicit interests and analyzed implicit

CitRec, August 27, 2017, Como, Italy

ACM ISBN 978-1-4503-5370-0/17/08...$15.00
https://doi.org/10.1145/3127325.3127330

behavior. In this way, they may not only promote the citizens' participation, but also could increase their engagement.

This work intends to proof the usefulness of personalized recommendations in e-participation. It does so through offline experiments carried out over the 'Decide Madrid' platform[2], which is the online digital medium set in place by the local government in the city of Madrid. Aiming to investigate which sources of user preferences (e.g., comments and social tags) and recommendation approaches (e.g., content-based and collaborative filtering) could be more effective, we perform offline experiments on a dataset obtained from the 'Decide Madrid' platform, and with a variety of evaluation metrics, such as precision, coverage and diversity.

2 RELATED WORK

Within the different forms of e-governance, government-to-citizens (G2C) governance aims to provide citizens with a variety of online information and e-services in an efficient and cost-effective manner, and to strengthen the relationship between government and citizens with ICTs. This is addressed at different levels of interaction between such actors, distinguishing among information, consultation and participation levels [20].

At the *e-information level*, the government offers websites with information on policies and programs, laws and regulations, budgets, and other issues of public interest, as well as software tools –such as email subscription lists, online newsgroups, and web forums– for the dissemination, and timely access and use of public information and services. In this context, recommender systems have been mainly proposed to provide the citizens with personalized government e-notifications and e-services; see e.g. [3], [5] and [6].

At the *e-consultation level*, the government offers online consultation (a.k.a. e-voting) mechanisms and tools, which present citizens with choices about public policy topics, allowing for the deliberation in real time, as well as the access to archived audios and videos of public meetings. Citizens are thus encouraged to contribute to the government consultations. In this context, recommender systems could assist voters in making decisions by providing recommendations about candidates close to the voters' preferences and tendencies. Terán and Meier [22] proposed a recommendation framework aimed to assist voters in making decisions by providing information about candidates close to the voters' preferences and tendencies. Its recommendations are based on similarities between voters and candidates –whose profiles are created by filling a questionnaire about values, attitudes and political issues on a number of topic categories. The system performs a fuzzy based clustering algorithm, and generates a graphical representation of political parties distributed in generated clusters, helping citizens to analyze politicians.

Finally, at the *e-participation level*, the government provides online participation platforms where citizens can propose, discuss, give feedback, and vote for initiatives aimed to solve or improve a wide range of situations and problems in different aspects of a city, such as health and social care, culture and education, energy and environment, and urban mobility and transport. In these platforms, recommendation approaches can assist the citizens in finding relevant proposals, discussions, individuals and associations, according to personal interests explicitly declared though votes, or implicitly expressed by means of online comments and social links. Nelimarkka et al. [19] present CRC, an online civic engagement platform that, differently to other analogous platforms, facilitates the participants' consideration of diverse viewpoints, an issue that is desirable in democratic processes and increases civic engagement, as shown by the authors. For such purpose, the platform recommends comments from individuals who hold similar and dissimilar opinions, by means of uncertainty minimizing sampling and PCA techniques.

Kavanaugh et al. [14] present Virtual Town Square, a location-based information aggregator system aimed to support and facilitate citizens' discussion and interaction. The system captures such information from several local news providers and user-generated media. Then, it filters and recommends relevant information items according to several aspects, such as topic, social media popularity, citizens' comments, and collaborative filtering similarities with like-minded people within trusted groups. Finally, instead of providing recommendation to citizens, Marsal-Llacuna and De la Rosa-Esteva [16] propose an agent-based model that, mining citizens' opinions expressed on the web, makes recommendations to planners on the design of an urban plan. A particular innovative feature of the model is that public participation occurs before and during the design and development of the plan. The recommendations are generated by a demographic collaborative filtering agent that exploits citizens' satisfaction surveys concerning a variety of issues about the city, and a content-based filtering agent that mines opinions of citizens from others cities about projects related to the target plan.

3 THE 'DECIDE MADRID' PLATFORM

In September 2015, Madrid city council launched the 'Decide Madrid' e-participation platform, a web system designed to allow Madrid residents to make, debate and vote proposals for the city on a variety of topics, such as transport, natural environment, urbanism, social rights, health care, education, and culture. Through this system, citizens contribute and decide how to spend part of the assigned participatory budget, which has been set to 100 million Euros for 2017.

This process consists of three main phases, namely *submit*, *support* and *vote* phases. In the *submit phase*, any person can create a proposal by signing up on the platform, and filling a simple questionnaire specifying a title, description, and some optional tags for the proposal (Figure 1 left). Then, the *support phase* is aimed to prioritize the most interesting and relevant proposals. For such purpose, city residents who are over 16 years old are allowed to explicitly express their support to existing proposals. The proposals that get support from 27064 people (i.e., 1% of the allowed residents) in a period of 30 days are approved. Before passing to the next phase, during a period of 45 days, approved proposals are commented and discussed by the citizens in the platform (Figure 1 right). The proposals without enough support are discarded and archived. Finally, in the *vote* phase, during a period of one week from its approval date, each approved proposal can be voted by allowed residents. In case there are more people in favor than against, a proposal is accepted as a 'collective proposal' of Madrid citizens, and the city council government assumes it as its own and carries it out. To achieve this, within a maximum period of one month, the corresponding technical reports on the legality, feasibility and economic cost of the proposal are published on the web. Then, citizens can access the plan to accomplish the proposal, and track its progress.

[2] 'Decide Madrid' e-participation platform, https://decide.madrid.es

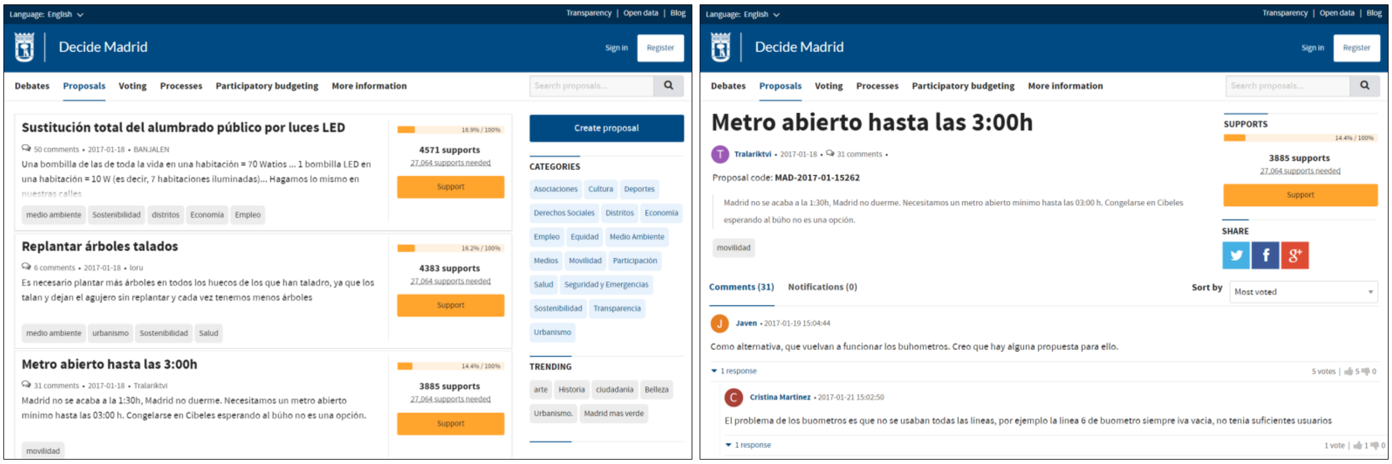

Figure 1: Screenshots of the 'Decide Madrid' platform, showing proposals metadata (title, author, date, description, tags, number of supports) and citizen comments.

3.1 The 'Decide Madrid' folksonomy

When a citizen creates a proposal in the 'Decide Madrid' e-participation platform, she has the opportunity to annotate it with a number of freely chosen words, which are called as *categories* in the platform, and are commonly known as *social tags* in the literature.

In a particular system, the whole set of tags constitutes an unstructured collaborative classification scheme that is referred as *folksonomy*. This implicit classification is then used to search, discover and recommend (tagged) resources of interest.

In general, within the 'Decide Madrid' platform, the tags assigned by a user to a particular proposal correspond to the places and topics related with the proposal. In both cases, a tag, which is a plain-text word, does not follow any categorization schema, and has not been assigned any metadata, which would allow establishing the meaning, type and properties of the concept underlying the tag. Hence, a 'place tag' may refer to a district (e.g., *centro*), a neighborhood (e.g., *sol*), a street (e.g., *gran via*), a square (e.g., *plaza mayor*), and a museum (e.g., *el prado*), to name a few; and a 'topic tag' may refer to a city asset (e.g., *transporte*), a particular issue (e.g., *precio metro*), and a social group (e.g., *jovenes*), among others.

As we shall explain in the next section, we have developed methods to determine whether a tag refers to a place or to a topic, and specify its corresponding concept and type. In our experiments, we evaluate content-based and hybrid recommendation approaches that exploit tags with and without the above processing.

3.2 The 'Decide Madrid' proposal forums

As mentioned before, and it is shown in Figure 1 (right), the 'Decide Madrid' platform provides an online forum for each proposal where citizens discuss it by means of comments and replies threads. In the forum, every comment and reply may receive positive and negative votes from the system users.

The platform website shows the number of positive/negative votes given to all proposals, comments and replies. In a collaborative filtering context, the votes would be the citizens' ratings for the proposals.

However, the website does not show the users who gave such votes, and thus does not make the explicit [user, item, rating] tuples publicly available. To address this situation, in our experiments, we consider the users' comments as a signal of their interest for the corresponding proposals, and thus we treat them as ratings, for collaborative filtering offline evaluation.

4 EXPERIMENTS

4.1 Dataset

We conducted the experiments on a dataset generated with data publicly available in the 'Decide Madrid' website. Specifically, we crawled the website gathering information about the proposals recorded in the system from 15th September 2015 to 31st May 2017. To obtain the users' comments, we only accessed the first page of each proposal forum, according to the decreasing popularity of its comments.

Table 1 shows statistics about the dataset. We obtained a total of 54357 ratings (i.e., unique [user, proposal] comment pairs) given by 17991 users to 16880 proposals, which leads to a 99.98% of rating sparsity. All the proposals were tagged. The users provided 58294 tag assignments (i.e., unique [user, tag, proposal] tuples), using 2967 distinct tags, from which the 7.65% were mapped to places, and 52.28% were mapped to topics. In the next two subsections, we explain how we performed such tag mappings.

Number of users	17991
Number of users who created proposals	11489
Number of users who commented proposals	10481
Number of proposals	16880
Number of proposals with place tags	16880 (100%)
Number of proposals with topic tags	11724 (69.45%)
Number of tags	2967
Number of place tags	227 (7.65%)
Number of topic tags	1551 (52.28%)
Number of tag assignments	58294
Number of place tag assignments	24179 (41.48%)
Number of topic tag assignments	31691 (54.36%)
Number of ratings	54357
Number of proposals with ratings	12055 (71.42%)
Rating sparsity	99.98%

Table 1: Dataset statistics.

4.1.1 Place tags

To determine whether each plain-text tag corresponds to a certain place in Madrid, we first created a repository of places in the city. Madrid has 21 districts (Figure 2), each of them with several neighborhoods. In total, it has 129 neighborhoods. We also considered an artificial district we called 'whole city' since there are proposals that are applicable to all the districts.

From a public database available at the 'madrid.org' Open Data portal[3], we downloaded lists of streets (and their neighborhoods) and representative places (e.g., hospitals and museums) in Madrid. Then, we processed the database transforming the names of the places to words in lowercase without numbers, punctuation symbols and accented vowels (as we also did with the social tags from the platform), e.g., *Gran Vía, 1* was transformed to *gran via*. Table 2 shows statistics about the generated places database.

Type	Places	Type	Places
Districts	22	Police stations	50
Neighborhoods	129	Shopping centers	48
Streets, squares, bridges	7979	Cinemas	36
Universities, colleges, schools	543	Parks	31
Religious buildings	317	Cemeteries	22
Hospitals, community health centers	170	Fire stations	12
Theaters	113	Amusement parks	4
Museums	63	Bullrings	1

Table 2: Places database statistics.

Afterwards, we mapped tags (and consequently proposals) to places by exact matching with the places names. As can be observed in Table 3, there was a quite uniform distribution of the citizens' proposals along the city districts.

Figure 2: Districts of Madrid. The numbers are district ids.

Id	Name	Proposals	Id	Name	Proposals
0	Whole city	13118 (77.71%)	13	Puente de Vallecas	184 (1.09%)
1	Centro	377 (2.23%)	11	Carabanchel	181 (1.07%)
17	Villaverde	353 (2.09%)	19	Vicálvaro	166 (0.98%)
8	Fuencarral-El Pardo	351 (2.08%)	3	Retiro	147 (0.87%)
18	Villa de Vallecas	311 (1.84%)	15	Ciudad Lineal	141 (0.84%)
16	Hortaleza	286 (1.69%)	6	Tetuán	125 (0.74%)
10	Latina	246 (1.46%)	4	Salamanca	113 (0.67%)
12	Usera	230 (1.36%)	14	Moratalaz	101 (0.60%)
2	Arganzuela	219 (1.30%)	7	Chamberí	98 (0.58%)
9	Moncloa-Aravaca	205 (1.21%)	5	Chamartín	88 (0.52%)
20	San Blas-Canillejas	202 (1.20%)	21	Barajas	85 (0.50%)

Table 3: Number of proposals per district.

[3] Madrid places, http://www.madrid.org/nomecalles/DescargaBDTCorte.icm

4.1.2 Topic tags

The 'Decide Madrid' platform has 16 tags as main categories for the proposals. However, many other categories could be considered as topics of interest for the citizens, and several tags may refer to the same category. For these reasons, we extended the number of topic categories, and performed a semi-automatic method for assigning existing tags to each category.

Specifically, we manually inspected the 150 most popular tags in our dataset, and grouped them in 30 tag sets, each of them representing a topic of interest. The selection of the 30 topics was done during the above inspection.

Then, iteratively and following a decreasing popularity ordering, for every remaining tag t, we computed its Levenshtein distance with the tags assigned to each category. Next, the tag t was assigned to certain category if the category contained (i) a tag with distance lower than 3 to t –e.g., *alcalde* and *alcaldesa* for the 'City hall & Public Administration' category–, or (ii) a tag that contains t as substring, e.g., the tag *accesibilidad metro* was assigned to 'Accessibility' and 'Transport' categories, since *accesibilidad* and *metro* already belonged to such categories, respectively.

For each category, Table 4 shows the number of tags assigned to the category, and the number of proposals tagged with at least one tag of the category. Issues in transport and urban mobility (e.g., traffic jams, and public parking), natural environment and sustainability (e.g., pollution and waste), and health care (e.g., hospital resources) are well known, major problems in Madrid.

Topic category	Proposals	Tags	Topic category	Proposals	Tags
Transport	4372	237	City hall, Public Administration	444	70
Natural environment	4092	210	Education	268	94
Urbanism	2932	139	Animals	168	98
Health care	2246	125	Family and childhood	133	48
Sustainability	2187	39	Civic virtue	131	50
Social rights	1974	99	Justice	112	46
Citizen participation	1825	38	Leisure, entertainment	105	46
Culture	1638	53	Accessibility	83	23
Economy	1324	66	Politics	66	43
Sports	1210	73	Housing	60	31
Security, emergencies	1077	57	Adolescence, youth	56	21
Equity and integration	1025	68	Tourism	62	8
Government transparency	932	18	Delinquency	49	34
Job	886	30	Old age	33	22
Associations	603	21	Religion	11	6

Table 4: Number of tags per topic.

4.2 Citizen and proposal profiles

As usually done in content-based recommender systems, we defined user (citizen) and item (proposal) profiles as vectors in the same space. In particular, we built the profiles according to the tags used by the users to annotate the items.

We experimented with both *binary* and *weighted* tag profiles for the items, and the binary versions were consistently worse, in agreement with previous work [10]. As we will show in Section 4.5, we tested TF-IDF and BM25 weighting techniques, as done in [10]. Furthermore, since there are tags that are intrinsically related to the proposals by indicating specific topics and affected places, we also experimented with item profiles containing one specific tag type, i.e., *place tags* or *topic tags*. Note that this filtering shrinks the vector space, favoring less sparse representations of the items.

Once the item profiles were built, we generated the user profiles by considering the proposals 'rated' by each citizen, since they describe her interests and tastes. Specifically, for each user, we aggregated her rated item profiles, and accumulated the weights computed for each dimension. By performing this transformation, we obtain a vector representation of users in the same space as items, allowing comparisons between them.

4.3 Recommendation algorithms

We evaluated the following recommendation algorithms, implemented on top of the RankSys framework[4]:

- **cb**: a content-based (CB) recommendation approach that exploits social tags to build the user and item profiles (as explained in Section 4.2). The score produced by this method is the Cosine similarity between the user's profile and the profile of every proposal (not previously seen by the user) in the system. As described before, in addition to exploiting all tags, we have performed different tag mappings, which are transparently evaluated by generating different user and item profiles, where, for instance, *place* vs. *topic* tags are compared.
- **ub:** a user-based nearest neighbor approach (UB k-NN) that exploits the rating-based similarity between users to create neighborhoods, which are used to compute a score for each (user, item) pair. In the experiments, we used the Cosine similarity between users and several neighborhood sizes, namely k = 5, …, 100, in steps of 5.
- **mf**: a matrix factorization collaborative filtering algorithm. We decided to use the variation proposed in [13] (the HKV factorizer implemented in RankSys,), since it is well suited for implicit feedback datasets; recall that there are only 1's in the user-item rating matrix. Several numbers of latent factors were tested: from 5 to 100, in steps of 5.
- **ib**: an item-based nearest neighbor approach (IB k-NN). This CF method works in a similar way as the *ub* approach, but the similarities are computed between items. In our experiments we used the Cosine similarity without any constraint on the neighborhood size; hence, the neighborhood is limited to the items rated by the user.
- **cbcf**: a hybrid recommendation approach where a user-based CF strategy is computed using CB user similarities. More specifically, we compute a Cosine similarity in a similar way as in the *cb* algorithm, but between two user profiles instead of a user and an item profiles. Similarly to the *cb* algorithm, *cbcf* is evaluated by using place, topic and all tags separately.
- **ipop**: a popularity-based recommender. The items (proposals) with more ratings (comments) are recommended to the users, without considering any personal information.

4.4 Evaluation methodology and metrics

We focused our evaluation on ranking-based metrics, hence no rating prediction metrics (such as the Mean Absolute Error and the Root Mean Square Error) will be reported. The rankings were generated following the *TrainingItems* methodology described in [8], where every item in the training split, except the ones already seen by the user in training, is considered as a possible candidate to be part of a user's final ranking. More specifically, we followed a 5-fold cross validation strategy to split the dataset into training and test: 80% of the interactions were randomly selected to build the training split, and the remainders were used for the test split.

The reported metrics are the following:

- **Precision** and **recall**: these metrics measure the amount of relevant returned items, either normalized by the amount of items returned (precision) or the amount of relevant items known for each user (recall).
- **MAP** and **nDCG**: these two metrics (Mean Average Precision and normalized Discounted Cumulative Gain) allow considering differences in the positions of the relevant returned items.
- **USC (User Space Coverage)**: this metric measures the amount of users who can receive a recommendation (user coverage). It is important to consider the tradeoff between USC and recommendation quality (as measured by the previous metrics).
- **ISC (Item Space Coverage)**: this metric measures the number of different items a recommender is able to recommend. It is thus related to the diversity of the recommendations, since the larger its value, the more diverse the recommendations presented to the users.

For these metrics, we tested several cutoffs, but decided to report the performance at ranking size of 50 because it was more stable.

4.5 Results

Table 5 shows the results obtained in the evaluation of the recommendation algorithms using the e-participation dataset described in Section 4.1.

In general, we observe that using all the tags led to the best results in *cb* and *cbcf* approaches. Although not reported here, we tested different weighting schemes –binary, TFIDF, and BM25 [10]–, and for every metric, TFIDF and BM25 achieved very similar results, and were clearly superior to the binary weights. Because of this, the values shown in the table correspond to the TFIDF weighting scheme. Regarding the type of tags used to create the user and item profiles, we observe that topic tags outperformed place tags in terms of Precision, MAP, nDCG and ISC. Place tags, however, provided better coverage than topic tags, in terms of USC for the *cb* and *cbcf* approaches, and Recall for the *cb* approach.

The content-based algorithms outperformed the popularity-based recommender and two of the CF approaches, namely item-based kNN (*ib*) and matrix factorization (*mf*). This could be attributed to a dataset highly skewed towards content features. However, the best performing algorithm for most of the metrics was user-based kNN (*ub*). This evidences the importance of the user-item patterns contained in the data, which were not properly exploited by the other CF algorithms, maybe due to the very high rating sparsity or the lack of non-unary ratings between users and items. In contrast, it should be noted that *ub* recommendations had a low coverage, for both users (USC) and items (ISC), and thus despite being precise, they are not diverse and are generated for a limited number of users. Differently, the hybrid *cbcf* approach showed a good tradeoff between recommendation precision and user coverage, as well as the best item coverage, i.e., the highest recommendation diversity.

[4] RankSys recommender systems evaluation framework, http://ranksys.org

		Precision	Recall	MAP	nDCG	USC	ISC
cb	*place tags*	0.001	0.040	0.005	0.013	0.579	0.087
	topic tags	0.001	0.026	0.006	0.010	0.579	0.183
	all tags	0.002	0.056	0.013	0.023	0.579	0.241
cbcf5	*place tags*	0.002	0.011	0.003	0.006	0.577	0.214
	topic tags	0.003	0.013	0.005	0.008	0.504	0.230
	all tags	0.005	0.019	0.008	0.012	0.563	0.238
cbcf10	*place tags*	0.002	0.018	0.005	0.008	0.579	0.233
	topic tags	0.002	0.021	0.006	0.010	0.513	0.264
	all tags	0.003	0.029	0.010	0.015	0.577	0.274
ub5		0.008	0.044	0.017	0.024	0.389	0.243
ub10		0.006	0.059	0.019	0.029	0.467	0.260
ub15		0.006	0.067	0.020	0.031	0.491	0.265
mf5		0.001	0.011	0.001	0.003	1.000	0.044
mf10		0.001	0.016	0.002	0.005	1.000	0.081
mf15		0.001	0.018	0.003	0.006	1.000	0.102
ib		0.002	0.035	0.009	0.015	0.526	0.237
ipop		0.001	0.027	0.006	0.011	1.000	0.004

Table 5: Experimental results.

5 CONCLUSIONS AND FUTURE WORK

Motivated by the need of incorporating personalized information retrieval and filtering functionalities into e-participation systems, in this paper we have empirically compared a number of recommendation approaches for the 'Decide Madrid' platform, where Madrid citizens create, debate and vote proposals for the city since September 2015.

Assuming that a citizen's comment on a proposal is a signal of her interest for the proposal, we have shown that user-based collaborative filtering heuristics seem to be the best performing approaches in terms of precision-recall and ranking-based metrics, such as MAP and nDCG. In contrast, exploiting the tags assigned by the system users to the proposals, a content-based approach has achieved the highest coverage values. Finally, a simple content-based collaborative filtering approach that jointly uses rating and tag-based information has obtained very good coverage and diversity values. The real impact of these results has to be evaluated in a user study.

In this context, the design and evaluation of alternative hybrid recommendation approaches are left for future work. For instance, we plan to evaluate the exploitation of other place- and topic-based user/item profiles inferred from tagging information. In fact, we have already mapped 100% (11.2%) of the proposals to their corresponding districts (neighborhoods). Using distance metrics between districts/neighborhoods may be valuable.

In addition to tags, the citizens' comments can be further exploited. Applying NLP and Opinion Mining techniques on the comments may allow us to determine whether each comment is in favor or against a particular proposal. With this information, we would be able to consider binary (like/dislike) ratings, instead of the unary ratings used in our experiments. Moreover, evaluating matrix factorization models that exploit content-based information, conducting experiments on data from the 2016 and 2017 participatory budgeting editions separately, and considering the 'whole city' artificial district differently, are issues we want to address in the future.

ACKNOWLEDGMENTS This work was supported by the Spanish Ministry of Economy, Industry and Competitiveness (TIN2016-80630-P).

REFERENCES

[1] Ahlers, D., Driscoll, P., Löfström, E., Krogstie, J., Wyckmans, A. 2016. Understanding smart cities as social machines. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 759–764.

[2] Alawadhi, S., Aldama-Nalda, A., Chourabi, H., ... & Walker, S. 2012. Building understanding of smart city initiatives. In *Proceedings of the 11th International Conference on Electronic Government*, pp. 40–53.

[3] Álvarez-Sabucedo, L., Soto-Barreiros, R., Santos-Gago, J. M., Fernández-Iglesias, M. 2012. A hybrid semantic driven recommender for services in the eGovernment domain. In *Proc. of the 2nd Intl. Conference on Digital Information and Communication Technology and its Applications*, pp. 409–414.

[4] Andrés, J., Bonivento, H. 2009. Problemas institucionales de la participación ciudadana: Análisis conceptual y aplicación al caso colombiano. *XXII Concurso del CLAD sobre la reforma del Estado y modernización de la Administración Pública*, pp. 1–23.

[5] Ayachi, R., Boukhris, I., Mellouli, S., Ben Amor, N., Elouedi, Z. 2016. Proactive and reactive e-government services recommendation. *Universal Access in the Information Society* 15(4), 681–697.

[6] Baldassarre, C., Cremaschi, M., Palmonari, M. 2013. Bridging the gap between citizens and local administrations with knowledge-based service bundle recommendations. In *Proceedings of the 24th International Conference on Database and Expert Systems Applications*, pp. 157–161.

[7] Basañes, F., Bonnefoy, J. C., Yitani, J. E., ... Schmuck, E. 2015. *Liderando el desarrollo sostenible de las ciudades.* Banco Iberoamericano de Desarrollo.

[8] Bellogín, A., Castells, P., Cantador, I. (2011). Precision-oriented evaluation of recommender systems: An algorithmic comparison. In *Proceedings of the 5th ACM Conference on Recommender Systems*, pp. 333–336.

[9] Buccoliero, L., Bellio, E. 2016. Marketing and citizen web empowerment: An index for benchmarking Italian municipalities' web strategies. In *Proc. of the 9th Int. Conference on Theory and Practice of Electronic Governance*, pp. 17–26.

[10] Cantador, I., Bellogín, A., Vallet, D. 2010. Content-based recommendation in social tagging systems. In *Proceedings of the 4th ACM Conference on Recommender Systems*, 237–240.

[11] Coleman, S. 2008. Doing IT for themselves: Management versus autonomy in youth e-citizenship. In *W. Lance Bennett (Ed.), Civic life online: Learning how digital media can engage youth*, pp. 189–206.

[12] Hernández-Bonivento, J. A. 2011. *La gobernanza local y el fortalecimiento institucional.* GIGAPP-IUIOG Estudios Working Papers, num. 2011-08.

[13] Hu, Y., Koren, Y., Volinsky, C. 2008. Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 8th IEEE International Conference on Data Mining*, pp. 263–272.

[14] Kavanaugh, A., Krishnan, S., Pérez-Quiñones, M., Tedesco, J., Madondo, K., Ahuja, A. 2014. Encouraging civic participation through local news aggregation. *Information Polity* 19(1, 2), 35–56.

[15] Macintosh, A. (2004). Characterizing e-participation in policy-making. In *Proc. of the 37th Hawaii Intl. Conference on System Sciences*, pp. 10–pp.

[16] Marsal-Llacuna, M. L., De la Rosa-Esteva, J. L. 2013. The representation for all model: An agent-based collaborative method for more meaningful citizen participation in urban planning. In *Proc. of the 13th International Conference on Computational Science and its Applications*, pp. 324–339.

[17] Meijer, A., & Bolivar, M. P. R. 2015. Governing the smart city: A review of the literature on smart urban governance. *International Review of Administrative Sciences* 82(2), 392–408.

[18] Nam, T., Pardo, T. A. 2011. Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference*, pp. 282–291.

[19] Nelimarkka, M., Nonnecke, B., Krishnan, S., ... Patel, J. 2014. Comparing three online civic engagement platforms using the "spectrum of public participation" framework. In *Proceedings of the 3rd Internet, Policy, and Politics Conference.*

[20] Organisation for Economic Co-operation and Development (OECD). 2001. *Citizens as partners: Information, consultation and public participation in policy-making.* OECD Publishing.

[21] Organisation for Economic Co-operation and Development (OECD). 2003. *Promises and problems of e-democracy; Challenges of citizen online engagement.* OECD Publishing.

[22] Terán, L. F., Meier, A. 2010. A fuzzy recommender system for eElections. In *Proceedings of the 1st International Conference on Electronic Government and the Information Systems Perspective*, pp. 62–76.

[23] Salim, F., & Haque, U. 2015. Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things. *International Journal of Human-Computer Studies* 81, 31–48.