Fusion strategies to combine topical and temporal information for publication venue recommendation

Luis M. de Campos, Juan M. Fernández-Luna and Juan F. Huete

Departamento de Ciencias de la Computación e Inteligencia Artificial, ETSI Informática y de Telecomunicación, CITIC-UGR, Universidad de Granada, 18071 Granada, Spain

Abstract

We study the publication venue recommendation problem, where a recommender system must help a researcher to decide where to submit a given target article for possible publication. We focus on content-based recommendation approaches, where we explicitly look for a good match between the topics discussed in the article and the recommended venue. But, in addition to this topical dimension, we also want to include a temporal dimension, in such a way that we prefer those venues where the articles which are more related with the target article have been published more recently. We use an information retrieval system to obtain separate topical and temporal recommendations and then combine them by means of different fusion strategies. The experimental results obtained on a collection of biomedical journal articles confirm the effectiveness of our proposals.

Keywords

Publication venue recommendation, topical profiles, temporal information, decay functions, information retrieval, content-based recommendation, fusion strategies

1. Introduction

Every researcher in any discipline has to face the problem of deciding where to publish the scientific article he/she has written. This decision, among other factors, should be based on the suitability of the topics that the article deals with respect to those usually treated by the tentative publication venues (either journals or conferences). This is known as the publication venue recommendation problem [28, 45].

Although there are also approaches to this problem based on collaborative filtering (recommending venues where either similar researchers or coauthors have published in the past [25, 31]), in this paper we are going to adopt a content-based approach, where the decision about which venue to recommend is mainly based on the textual content of the paper to be published [20, 42].

The information that the recommender system possesses about each possible publication venue can be represented by means of a *profile* [19]. Although there are different types of profiles, for example based on semantic networks or concepts [39], the most common and

CIRCLE'22: Joint Conference of the Information Retrieval Communities in Europe, July 4–7, 2022, Samatan, France
lci@decsai.ugr.es (L. M. de Campos); jmfluna@decsai.ugr.es (J. M. Fernández-Luna); jhg@decsai.ugr.es (J. F. Huete)
0000-0001-9125-1195 (L. M. de Campos); 0000-0002-0366-4545 (J. M. Fernández-Luna); 0000-0002-9970-7705 (J. F. Huete)

simple type that can be associated to an item in general, or to a venue in our particular scenario, is based on a set of terms or keywords (perhaps weighted) [16]; for example, a textual description of the item, or a subset of the terms that appear in the articles published in the venue. These term-based profiles can be built automatically.

In [10] we proposed the use of several homogeneous subprofiles to represent each venue, as an alternative to a single and heterogeneous profile comprising all the information about it. The intuition is that if we are able to describe with more precision which are the main topics covered by each venue (these topics being obtained analyzing the articles published in it through a clustering process) and we build separate subprofiles for each topic, probably we could recommend more accurately the most appropriate venues for publishing a new article.

However, we did not consider other information source that could also help to improve the recommendation results, namely the temporal information. The motivation for using this source for publication venue recommendation is that we do not only want that the thematic content of our new (target) article matches with the recommended venues, we would also prefer that the themes this article deals with have been *recently* considered in these venues: If venues A and B both have published articles related with the main topic of the target article but the articles in A are more recent than the articles in B, then probably we would prefer to select venue A instead of B.

In this work we aim to build a publication venue recommendation system considering both the temporal and the topical dimension of the available data (the articles published in the venues). Therefore, on the one hand, we want to study how the temporal information alone can be used in the publication venue recommendation problem and the benefits it may offer; on the other hand, we also want to study different ways of combining the temporal information with the topical information, hoping that these two complementary sources reinforce each other.

The remainder of this paper is organized as follows: Section 2 briefly discusses some related works. Section 3 describes how we build the topical subprofiles by means of Latent Dirichlet Allocation (LDA) [6] applied to the articles published in a set of venues, and how to use them within an information retrieval system (IRS). Section 4 explains how the temporal information of the articles, through the use of a decay function, can be incorporated to modify the output of another IRS for articles. Then, in Section 5 the way of building the publication venue recommender from either the IRS of subprofiles or the IRS of articles is outlined. Section 6 is devoted to study methods to combine the topical and temporal recommenders. Section 7 reports the results of the experiments with our models and the baselines. Finally, Section 8 contains the concluding remarks.

2. Related work

In the literature about content-based publication venue recommendation there are different ways of representing the information within the profiles (terms, n-grams, noun-phrases, topics) but in all the cases each venue has either a single profile [28, 40, 45] or as many subprofiles as published articles [20, 35, 25, 45, 32, 37, 17]. The only work where different subprofiles topically homogeneous are built for each venue is [10].

Moreover, there are not much works explicitly using temporal information: [2] uses a personal venue rating which takes into account the years when the articles published in a venue were added to the researcher personal collection. This rating is used to compute a similarity between researchers which is then applied within a collaborative filtering algorithm to recommend venues. In [32], within their system, the authors compute a similarity between venues that gives more weight to the information of the articles associated to more recent years using an age-discounted scheme (inverse log-weighting scheme), and this similarity is then used to build a graph of venues where a random walk with restart algorithm is applied.

The situation is different in the context of general recommender systems, where there are many works which consider organizations of the information different from single profiles. These multi-faceted profiles can be trees [30], graphs of clusters [47] or hierarchies [21, 36]. Other methods generate topical subprofiles using clustering (grouping either terms or documents) [3, 5, 11, 13, 26].

Within general recommender systems there are also many works considering the inclusion of temporal information, although most of them are focused on collaborative filtering techniques [8]. A quite common method is to use decay functions or temporal discounting to penalize old items [15, 46]. Another option is to consider time frames. For example, [33] studies a content-based filter for tweets, learning a specific time frame for each user and only recommending tweets within this personalized frame. Similar ideas are used in [38] for points-of-interest recommendation. Long and short-term profiles are another way to incorporate time, by distinguishing the most recent interests of users from those more stable or permanent [23, 44].

3. Topical subprofiles

Two basic organizational schemes to represent information about the different publication venues are that we call *monolithic* profiles and *atomic* subprofiles, which represent the two extremes in terms of heterogeneity or homogeneity of the information. In monolithic profiles all the articles published in a given venue are grouped together in a single (heterogeneous) macro-document. On the contrary, in atomic subprofiles each individual article published in a venue forms a (very homogeneous) subprofile for this venue, having as many subprofiles as articles it contains. An intermediate way of organizing the information would be to build subprofiles around the different thematic areas covered by each venue.

In [10] we used clustering (namely k-means) to group the published articles in homogeneous groups based on their textual content. Then the text of all the articles in a venue which belong to the same group was used to build the corresponding subprofile for this venue (one subprofile for each cluster). In this paper we are going to build topical profiles in a similar way but using LDA instead of clustering.

We propose to apply a *global* LDA to the collection of all the articles of all the venues[1]. It is known that LDA tries to find latent topics in a document collection. Each of these latent topics is characterized by a different probability distribution of terms (given the topic). Moreover, in this non-supervised method, for each document a probability distribution of topics is obtained, which describes to which extent each document is about each topic. LDA requires the number of topics to be used, k, as an input parameter.

[1]In contraposition to use a separate, *local* LDA applied to all the articles of *each* venue.

Once LDA has been applied to our articles collection and we have computed the probability distributions of topics, we associate each article with its most probable topic. Therefore, we have a partition of the articles in k groups, where all the articles in each group deal mainly about the same global topic.

The next step is to obtain the homogeneous subprofiles for each publication venue. We simply divide each group of articles associated with the same topic into local groups of articles related to this topic *and* each venue. All the articles in each local group associated with a venue are then concatenated to form a single document/subprofile. In this way each venue will have at most k subprofiles, probably much less than k, because it is quite possible that the vast majority of the venues do not deal with all the topics but are especialized in a more reduced number of specific topics (and therefore there will be topics which are not the most probable topic for any article in a given venue).

Finally, the document collection formed by the subprofiles of all the venues is indexed by an IRS that will be used to match the content of the target article (the query) with the subprofiles, providing a score, $sc_s(p)$, for each subprofile p and hence a ranking of subprofiles.

4. Managing temporal information

We assume that a researcher normally will prefer to publish in a venue where the topics related to their article are dealt with *recently*, i.e. they are topics of current interest of the venue. In other words, a venue that has recently published articles which are closer to the target article is preferable. Although reasonable and intuitive, this is only a working hypothesis, which our experiments in Section 7 will confirm[2].

A simple but effective way of incorporating this criterion in the recommendation is through the use of a temporal decay function $f(\Delta t)$ [22]. This function considers the difference, Δt, between the current year[3] and the publication year of an already published article, in such a way that the greater Δt is the less important becomes this article (from a temporal perspective).

Then, the recommender system based on temporal information will work in the following way: we index the collection of articles (the atomic collection) using an IRS. Given the target article, the IRS obtains a score $sc_a(d)$ for each article d in the collection, which represents the degree of similarity between the target and d. Then we compute a modified score $msc_a(d)$ by using also the decay function, $msc_a(d) = sc_a(d) \times f(\Delta t)$, which reduces the similarity of the article as the temporal distance between the target and d increases. Finally, we obtain a ranking of articles according to this modified score.

5. Building the publication venue recommenders

In the previous models, the ranking obtained by the IRS, given the target article as the query, is either a ranking of individual articles (for the temporal and the atomic models) or a ranking of

[2]However, in these experiments, we only have access to the venues where the articles were in fact published but we do not know which venues the authors considered in the first place.

[3]In our context this is the year of publication of the test article.

subprofiles (for the topical model)[4]. However, a ranking of publication venues is required, as we are going to recommend venues to publish the target article. To combine the scores of the different documents associated with each venue v (i.e. for those articles $d \in v$ published in v and for those subprofiles $p \in v$ associated to v) and generate a final venue ranking, we use the *CombLgDCS* method [9], which sums up the scores of all of the subprofiles/articles associated with each venue but taking into account their positions in the ranking (using a logarithmic devaluation):

$$sc_{top}(v) = \sum_{p \in v} \frac{sc_s(p)}{\log_2(rank(p)+1)}, \tag{1}$$

$$sc_{temp}(v) = \sum_{d \in v} \frac{msc_a(d)}{\log_2(rank(d)+1)}, \tag{2}$$

where $sc_s(p)$ and $msc_a(d)$ denote the original score values obtained by the IR systems based on subprofiles and articles, respectively, and $rank(p)$ (respect. $rank(d)$) is defined as 1 in the case that p (respect. d) being the first occurrence of a subprofile (respect. an article) of v in the original ranking and, otherwise, it is the raw value of the position of the subprofile (respect. the article) in this ranking.

6. Combining topical and temporal information

Once we have developed publication venue recommenders based on both topical and temporal information, the next logical step is to combine them to try to improve performance. As the previous methods are based on IR systems and each generates a ranking of venues given the target article, it seems reasonable to use some fusion/aggregation methods which are usually employed to combine the rankings of different IR systems [43]. There are two basic types of fusion methods: score-based and rank-based methods [34].

Score-based methods combine the scores generated by each IRS for a document, thus obtaining a combined score that is used to rerank the documents. However, as we are combining the scores of different IRSs with different characteristics (in our case scores obtained from different types of documents, subprofiles in one case and articles in other case), previous to the combination a score normalization is necessary to make the scores comparable to each other. We normalize the scores of each IRS by dividing by the corresponding maximum score. Rank-based methods do not use the scores (in case these scores exist) but only the information provided by the rankings.

Among the score-based methods, we are going to use the so-called CombSUM and CombMNZ, proposed originally in [18], and LC, a linear combination of scores.

CombSUM simply computes the sum of the scores of the different IRSs, in our case the combined score of venue v is

$$sc(v) = nsc_{top}(v) + nsc_{temp}(v),$$

[4]For the monolithic model we directly obtain a ranking of venues.

where nsc_{top} and nsc_{temp} are the normalized scores generated by the topical and the temporal recommenders, respectively (and assuming that if a venue v does not appear in some ranking, its corresponding score is equal to 0).

CombMNZ is similar to CombSUM but tries to promote those venues appearing more frequently in the rankings, by multiplying the sum of scores by the number of rankings where the document appears. In our case the score of venue v is

$$sc(v) = \begin{cases} 2 * (nsc_{top}(v) + nsc_{temp}(v)) & \text{if } v \text{ appears in the two rankings} \\ nsc_{top}(v) & \text{if } v \text{ appears only in the topical ranking} \\ nsc_{temp}(v) & \text{if } v \text{ appears only in the temp ranking} \end{cases}$$

The linear combination LC is

$$sc(v) = a * nsc_{top}(v) + (1 - a) * nsc_{temp}(v),$$

where a is a parameter controlling the relative importance given to the topical recommender.

Among the rank-based fusion methods, we shall use Borda Count [4, 24] and Condorcet [29], which originally were election methods.

Borda methods transform the rankings into scores which are later combined using different functions. Each document is associated with its position in the ranking: the first document gets score 1, the second document obtains score 2, and so on until the last document in the ranking, which gets score Y (if there are Y documents in the ranking). If a document does not appear in the ranking (but we are going to consider it because it appears in another ranking), its score is $Y + 1$. Then the scores obtained from each ranking are combined with some functions and a new ranking is generated (in this case sorting from least to greatest). In our case, having only two rankings and hence two scores, sc_{btop} and sc_{btemp}, the different functions used give rise to the following expressions:

BordaSUM: $sc_b(v) = sc_{btop}(v) + sc_{btemp}(v)$
BordaPROD: $sc_b(v) = sc_{btop}(v) * sc_{btemp}(v)$
BordaL2: $sc_b(v) = sc_{btop}(v)^2 + sc_{btemp}(v)^2$

Another common option is to associate with each document some transformation of the ranking instead of the ranking itself. For example, we can link with a document in position i the score $1/i$, thus obtaining (using the same functions to combine the scores as before) the Reciprocal Borda methods, BordaRSUM, BordaRPROD and BordaRL2.

The Condorcet method is based on pairwise comparisons between the documents in the ranking: given a ranking, if document d_i is previous to (is prefered to) document d_j in the ranking then we add 1 to the element m_{ij} in a matrix. This process is repeated for all the rankings, thus obtaining at the end in m_{ij} the number of times that d_i was prefered to d_j. If this number is strictly greater than half the number of existing rankings (one in our case), we count a victory of d_i on d_j. In this way we are accumulating the number of victories of d_i over the other documents. This number is the score associated to d_i and based on these scores a new ranking is generated[5].

[5]Ties are resolved in favor of the document having greater $\sum_k m_{ik}$.

7. Experiments

In this section we are going to experimentally tests our proposals for publication venue recommendation. We explain first the experimental setting and next the obtained results.

7.1. Experimental settings

We used in the experimentation a set of 309,551 biomedical journal papers extracted from the PMSC-UGR test collection [1], namely those articles published between 2007 and 2016 in one of the 1002 journals having more than 100 papers in this period. The information used from each article was title, abstract, keywords and year of publication. The 276,679 articles published between 2007 and 2015 were used as training set and the test set is formed by the 33,872 articles published in 2016.

As mentioned previously, firstly we used LDA to find the latent topics in the training set and secondly to build the subprofiles associated to each journal. More precisely, we employed the implementation of LDA in the Gensim Python library, with default parameters. Before applying LDA, in order to reduce dimensionality, we performed stopwords removal and stemming; also, the terms appearing in more than 90% of the articles and those appearing in fewer than 750 articles were removed[6]. Concerning the LDA parameter fixing the number of topics, k, we proposed a value equal to the number of descriptors or categories in the second level of the MeSH thesaurus, which is 110.

Concerning the decay function used by the temporal model, we have considered two options: a linear decay, $f(\Delta t) = \frac{1}{1+\Delta t}$, and a power decay, $f(\Delta t) = \frac{1}{(1+\Delta t)^{1/4}}$. In the first case, the linear decay imposes a penalization more severe to older articles, whereas the power decay penalizes them more smoothly.

All the recommendation models being considered are supported by an IRS. As the original individual articles and also the different subprofiles are text documents (in the case of subprofiles these documents are formed by the concatenation of the articles within each subprofile), they form a training document collection which can be indexed and searched for. We have used the Lucene library[7] for these purposes, removing stopwords and performing stemming before indexing and using the Language Model (with Jelinek-Mercer smoothing) as retrieval model to compute the ranking of documents.

Each article from the test set is considered as the article for which we are seeking a recommendation (i.e. the target article), and its textual content is used to form a query to the IRS.

In order to evaluate the quality of the recommendations offered by the different models, we adopt a conservative but objective approach: only one journal is relevant for each query (test article) and this is the journal where the paper has actually been published. As evaluation measures we use accuracy@X (with X=1,5) and mean reciprocal rank, MRR. Accuracy@X computes the ratio between the number of recommendations where the true journal at which a test article was published is among the first X recommended journals and the number of all the

[6]All these reductions of terms are only used to apply LDA and determine the topics; the subprofiles obtained from these topics will contain all the terms appearing in the original documents associated with the subprofile.
[7]https://lucene.apache.org/

recommendations. MRR tries to reflect how high in the ranking the only relevant journal is recommended: it is computed as the average of the inverse of the positions in the ranking at which the actual journal where each test paper was published is found (0 if the actual venue does not appear in the top 40 positions in the ranking).

7.2. Results

We have experimented with all the fusion methods explained in Section 6, as well as the original Topical and Temporal models and also the baselines Monolithic and Atomic models. The results of our experiments with the different models are displayed in Table 1 and in Figures 1, 2 and 3. The decay function used for the temporal model (and hence for all the fusion models) in these results is always the power decay. We do not show the results obtained when using the linear decay because they are quite poor, even worse than the baselines. This implies that a smooth penalization of older articles is positive but it becomes self-defeating when the penalization is abrupt.

Table 1
Results of the experiments (best results in bold).

Model	acc@1	acc@5	MRR
CombSUM	**0.2408**	0.5533	**0.3831**
CombMNZ	**0.2408**	**0.5534**	0.3830
LC(0.75,0.25)	0.2385	0.5517	0.3814
BordaPROD	0.2385	0.5516	0.3812
BordaSUM	0.2384	0.5503	0.3805
BordaL2	0.2381	0.5490	0.3797
LC(0.25,0.75)	0.2379	0.5503	0.3798
BordaRSUM	0.2354	0.5527	0.3799
BordaRPROD	0.2354	0.5519	0.3798
BordaRL2	0.2354	0.5516	0.3798
Condorcet	0.2349	0.5498	0.3782
Topical	0.2346	0.5448	0.3763
Temporal	0.2331	0.5403	0.3731
Atomic	0.2282	0.5370	0.3696
Monolithic	0.2236	0.5278	0.3653

Although the differences between the different models are not quantitatively very big, the tendencies are clear and the three metrics being considered essentially move in the same direction (the rankings of the systems for all of them are very similar). All the fusion strategies improve the results of the two individual components, Topical and Temporal, which in turn both improve the results of the two baselines, Atomic and Monolithic, which do use neither topical nor temporal information. This confirms two facts: first, the use of topics to build more homogeneous subprofiles, as well as the use of temporal information (in the form of a decay function) is positive to get better venue recommendations; second, the topical and

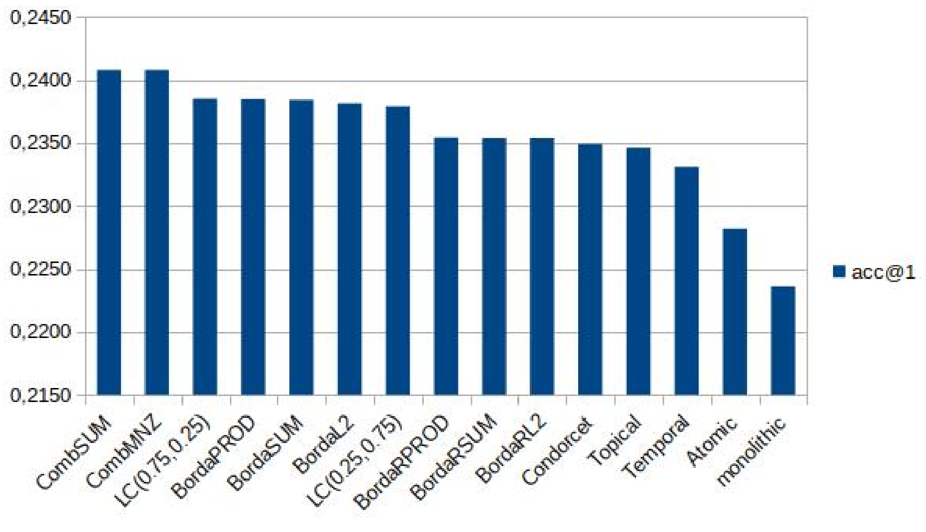

Figure 1: Accuracy@1 of the different models.

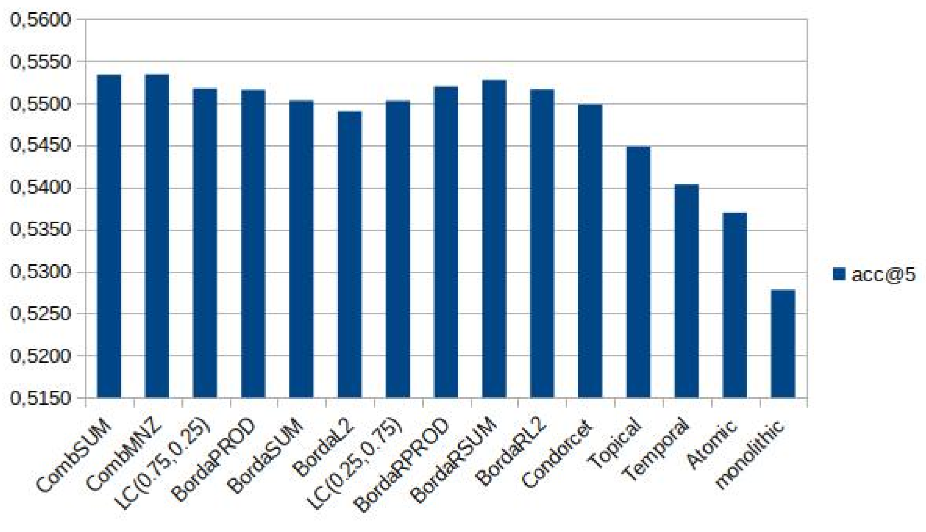

Figure 2: Accuracy@5 of the different models.

temporal information are complementary and their combination through a fusion method further improves the results. Moreover, the individual results for topical and temporal models are quite similar, although topics performs slightly better than temporal.

It can also be observed that the fusion methods based on scores (CombSUM, CombMNZ and LinearComb) perform better than those ones based only on the ranking information (Borda and Condorcet). This confirms findings obtained in other contexts [4].

There is almost no difference between the different methods based on scores, although the linear combination performs somewhat worse if the weights are not uniform (considering the

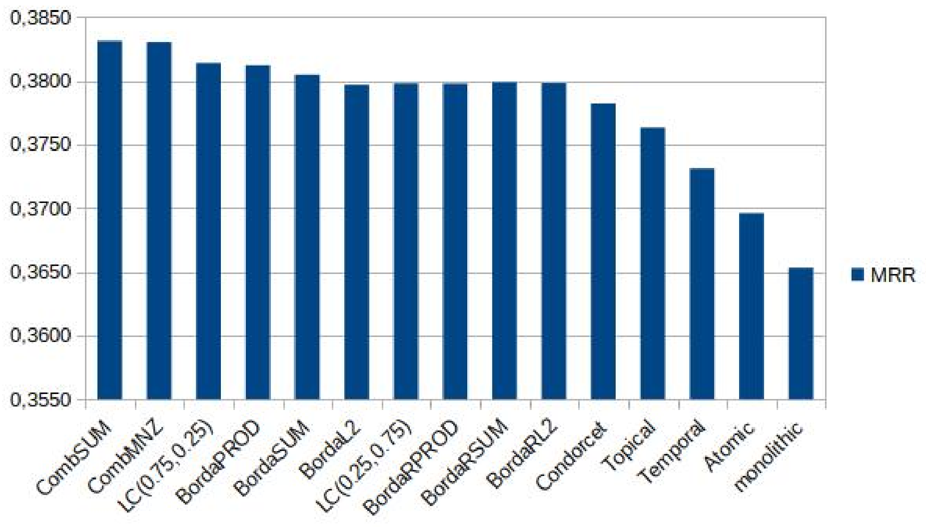

Figure 3: Mean reciprocal rank of the different models.

three weight combinations being used, taking into account that CombSUM is equivalent to LinearComb with both weights equal to 0.5). In other words, we should attribute the same importance to the topical and the temporal information (although perhaps it would be preferable to give somewhat more importance to topics than to time).

Within the methods based on ranking, the differences are also quite small, although the best method is always a variant of Borda (depending on the metric, either direct Borda or Reciprocal Borda is better) and Condorcet produces the worst results.

To assess the robustness of our main findings, we have used a statististical significance test, namely the McNemar test [27, 14]. This is a non-parametric test for paired data and we have applied it for the accuracy@1 metric. Essentially it compares, for each article in the test set, the result obtained by two different models, which in our case may be success or failure (i.e. the top ranked journal is the true journal where the test article was published or not), builds a 2×2 contingency table and computes a statistics related to the number of cases where the two models get different results. We selected a confidence level of 90%. For the pairwise comparisons we have used the best score-based and ranking-based fusion models (CombSUM and BordaPROD) as well as the Topical and Temporal components and the baselines Atomic and Monolithic.

The results of these pairwise tests (the p-values) are displayed in Table 2.

These results confirm that CombSUM is the best method, showing statistically significant differences with all the other methods. BordaPROD is the second best, although its differences with Topical and Temporal are not significant (but they are with all the other methods). Topical is somewhat better than Temporal (but without significant differences), which in turn is better than Monolithic and Atomic (having significant differences). Finally, Atomic is significantly better than Monolithic.

Table 2
p-values of the pairwise McNemar tests.

	BordaPROD	Topical	Temporal	Atomic	Monolithic
CombSUM	0.0232	0.0652	0.0200	0.0001	1.81e-07
BordaPROD	-	0.2507	0.1027	0.0017	6.15e-06
Topical	-	-	0.3716	0.0001	4.32e-07
Temporal	-	-	-	5.98e-06	0.0001
Atomic	-	-	-	-	0.0636

8. Concluding remarks

We have tackled the publication venue recommendation problem by building two content-based recommender systems using different dimensions: One system is based on a process that analyzes the articles published in the different venues through latent Dirichlet allocation and generates homogeneous subprofiles for each venue associated with the different discovered topics, which are then indexed by an IRS. The other system uses the temporal information of each published article, by means of a decay function, to modify the output of another IRS. The lists of recommended venues for publishing a given target article generated by each method are then combined by using several score-based and ranking-based fusion strategies.

We have carried out experiments with a collection of biomedical journal articles. The results obtained indicate that both the topical and the temporal recommenders improve the performance of the baseline models which do not use these types of information. Moreover, all the fusion strategies further improve the performance, thus showing that the two dimensions, temporal and topical, are complementary and their combined use is always beneficial.

We have considered the topical and temporal recommenders essentially in a separate way and then we have combined them by means of fusion strategies. However, for future work it would be interesting to consider other ways of merging these two dimensions, for example through the use of time-based topic models [7, 41]; or further subdividing the topical subprofiles into periods of time, thus obtaining topical-temporal subprofiles, as done in [12] in the context of expert finding; or creating temporal subprofiles for different periods of time and then further subdividing them topically by learning separate topical subprofiles within each temporal subprofile, thus obtaining temporal-topical subprofiles.

Acknowledgments

This work was supported by the Spanish “Agencia Estatal de Investigación” [grant number PID2019-106758GB-C31]; the Spanish “Programa operativo FEDER Andalucía 2014-2020 de la Junta de Andalucía y la Universidad de Granada” [grant number A-TIC-146-UGR20]; and the European Regional Development Fund (ERDF-FEDER).

References

[1] C. Albusac, L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, PMSC-UGR: A test collection for expert recommendation based on PubMed and Scopus, Lecture Notes in Artificial Intelligence 11160, Advances in Artificial Intelligence, CAEPIA 2018, 34-43, 2018.

[2] H. Alhoori, R. Furuta, Recommendation of scholarly venues based on dynamic user interests, Journal of Informetrics 11:553-563, 2017.

[3] B. Amini, R. Ibrahim, M.S. Othman, A. Selamat, Capturing scholar's knowledge from heterogeneous resources for profiling in recommender systems, Expert Systems with Applications 41:7945-7957, 2014.

[4] J.A. Aslam, M. Montague. Models for metasearch. In Proceedings of the 24th Annual International ACM SIGIR Conference, pp. 276-284, 2001.

[5] C. Au Yeung, N. Gibbins, N. Shadbolt, Multiple interests of users in collaborative tagging systems, In King, I. and Baeza-Yates, R., editors, Weaving Services and People on the World Wide Web, pp. 255-274, Springer, 2009.

[6] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, Journal of Machine Learning Research 3:993-1022, 2003.

[7] D.M. Blei, J.D. Lafferty, Dynamic topic models, Proceedings of the 23rd international conference on Machine learning, pp. 113-120, 2006.

[8] P.G. Campos, F. Díez, I. Cantador, Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols, User Modeling and User-Adapted Interaction 24:67-119, 2014.

[9] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, Committee-based profiles for politician finding, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 25(Suppl. 2):21-36, 2017.

[10] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, Publication venue recommendation using profiles based on clustering, IEEE Access, submitted.

[11] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, L. Redondo-Expósito. Automatic construction of multi-faceted user profiles using text clustering and its application to expert recommendation and filtering problems, Knowledge-Based Systems 190, 105337, 2020.

[12] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, L. Redondo-Expósito, Temporal and topical profiles for expert finding, Joint Conference of the Information Retrieval Communities in Europe, CIRCLE 2020, Eds. Iván Cantador, Max Chevalier, Massimo Melucci, Josiane Mothe, Volume 2621 of CEUR workshop proceedings, 2020.

[13] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, L. Redondo-Expósito, LDA-based term profiles for expert finding in a political setting, Journal of Intelligent Information Systems 56:529-559, 2021.

[14] T.G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Computation 10:1895-1923, 1998.

[15] Y. Ding, X. Li, Time weight collaborative filtering, In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pp. 485-492, 2005.

[16] C.I. Eke, A.A. Norman, L. Shuib, H.F. Nweke, A survey of user profiling: state-of-the-art, challenges, and solutions, IEEE Access 7:144907-144924, 2019.

[17] M. Errami, J.D. Wren, J.M. Hicks, H.R. Garner, eTBLAST: a web server to identify expert

reviewers, appropriate journals and similar publications, Nucleic Acids Research 35 W12-W15, 2007.

[18] E.A. Fox and J.A. Shaw. Combination of multiple searches. In Proccedings of the Second Text REtrieval Conference (TREC-2), pp. 243-252, 1994.

[19] S. Gauch, M. Speretta, A. Chandramouli, A. Micarelli, User profiles for personalized information access. In: The Adaptive Web, Lecture Notes in Computer Science 4321:54-89 2007.

[20] N. Kang, M. Doornenbal, B. Schijvenaars, Elsevier journal finder: recommending journals for your paper, RecSys'15, 261-264, 2015

[21] H.J. Kook, Profiling multiple domains of user interests and using them for personalized web support, Proceedings of the International Conference on Intelligent Computing, 512-520, 2005.

[22] S. Larrain, C. Trattner, D. Parra, E. Graells-Garrido, K. Nørvåg, Good times bad times: A study on recency effects in collaborative filtering for social ragging, Proceedings of the 9th ACM Conference on Recommender Systems, 269-272, 2015.

[23] L. Li, L. Zheng, T. Li, LOGO: a long-short user interest integration in personalized news recommendation. Proceedings of the 5th ACM conference on Recommender systems, 317–320, 2011.

[24] S. Lin, Rank aggregation methods, Wiley Interdisciplinary Reviews: Computational Statistics 2(5) 555-570, 2010.

[25] H. Luong, T. Huynh, S. Gauch, L. Do, K. Hoang, Publication venue recommendation using author networks publication history, in Intelligent Information and Database Systems, LNCS 7198, pp. 426-435, 2012.

[26] J.P. McGowan, N. Kushmerick, B. Smyth, Who do you want to be today? Web personae for personalised information access, Web-Based Systems: Second International Conference, 514–517, 2002.

[27] Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages, Psychometrika 12:153–157, 1947.

[28] E. Medvet, A. Bartoli, G. Piccinin, Publication venue recommendation based on paper abstract, Proceedings of the 26th IEEE International Conference on Tools with Artificial Intelligence, 1004-1010, 2014.

[29] M. Montague, J.A. Aslam, Condorcet fusion for improved retrieval, In Proceedings of the ACM CIKM Conference, pp. 538–548, 2002.

[30] M. Pavan, E.W.D. Luca, Semantic-based expert search in textbook research archives, Proceedings of the 5th International Workshop on Semantic Digital Archives, CEUR Workshop Proceedings, 1529:18-29, 2015.

[31] M.C. Pham, Y. Cao, R. Klamma, M. Jarke, A clustering approach for collaborative filtering recommendation using social network analysis, Journal of Universal Computer Science 17:583-604, 2011.

[32] T. Pradhan, S. Pal, CNAVER: A Content and Network-based Academic VEnue Recommender system, Knowledge-Based Systems 189, 105092, 2020.

[33] C. Ramos Casimiro, I. Paraboni, Temporal aspects of content recommendation on a microblog corpus. In: Baptista J., Mamede N., Candeias S., Paraboni I., Pardo T.A.S., Volpe Nunes M..G. (eds) Computational Processing of the Portuguese Language. PROPOR 2014.

Lecture Notes in Computer Science 8775:189-194, 2014.
[34] M.E. Renda, U. Straccia, Web metasearch: rank vs. score based rank aggregation methods, In Proceedings of the 2003 ACM symposium on Applied computing, pp. 841–846, 2003.
[35] J. Rollins, M. McCusker, J. Carlson, J. Stroll, Manuscript Matcher: a content and bibliometric-based scholarly journal recommendation system, 5th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2017), CEUR Workshop Proceedings vol. 1823, 2017.
[36] J. Rybak, K. Balog, K. Nørvåg, Temporal expertise profiling, In de Rijke, M., Kenter, T., de Vries, A. P., Zhai, C., de Jong, F., Radinsky, K., and Hofmann, K., editors, Proceedings of the 36th European Conference on IR Research, pp. 540-546, 2014.
[37] M.J. Schuemie, J.A. Kors, Jane: suggesting journals, finding experts, Bioinformatics 24:727-728, 2008.
[38] Y. Si, F. Zhang, W. Liu, CTF-ARA: An adaptive method for POI recommendation based on check-in and temporal features, Knowledge-Based Systems 128:59-70, 2017.
[39] A. Sieg, B. Mobasher, R. Burke, Web search personalization with ontological user profiles. Proceedings of the 16th ACM International Conference on Information and Knowledge Management, pp. 525-534, 2007.
[40] T. Silva, J. Ma, C. Yang, H. Liang, A profile-boosted research analytics framework to recommend journals for manuscripts, Journal of the Association for Information Science and Technology 66:180-200, 2015.
[41] I. Vayansky, S.A.P. Kumar, A review of topic modeling methods, Information Systems 94, 101582, 2020.
[42] D. Wang, Y. Liang, D. Xu, X. Feng, R. Guan, A content-based recommender system for computer science publications. Knowledge-Based Systems 157:1-9, 2018.
[43] S. Wu, Data Fusion in Information Retrieval, Adaptation, Learning, and Optimization, vol. 13, Springer, 2012.
[44] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, J. Sun, Temporal recommendation on graphs via long- and short-term preference fusion, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 723-732, 2010.
[45] Z. Yang, B.D. Davison, Venue recommendation: submitting your paper with style, Proceedings of the 11th International Conference on Machine Learning and Applications, vol. 1, pp. 681-686, 2012.
[46] R. Yeniterzi, J. Callan, Moving from static to dynamic modelling of expertise for question routing in CQA sites, Proceedings of the 9th International AAAI Conference on Web and Social Media, 702-705, 2015.
[47] H.-J. Zeng, Z. Chen, W.-Y. Ma, A unified framework for clustering heterogeneous web objects, Proceedings of the 3rd International Conference on Web Information Systems Engineering, 161-172, 2002.