Exploiting Citation Knowledge in Personalised Recommendation of Recent Scientific Publications

Conference or Workshop Item

How to cite:

For guidance on citations see FAQs.

© [not recorded]

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher’s website:

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
Exploiting Citation Knowledge in Personalised Recommendation of Recent Scientific Publications

Anita Khadka¹, Iván Cantador², Miriam Fernandez¹
¹Knowledge Media Institute, The Open University, UK
²Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain
anita.khadka@open.ac.uk, ivan.cantador@uam.es, miriam.fernandez@open.ac.uk

Abstract
In this paper we address the problem of providing personalised recommendations of recent scientific publications to a particular user, and explore the use of citation knowledge to do so. For this purpose, we have generated a novel dataset that captures authors’ publication history and is enriched with different forms of paper citation knowledge, namely citation graphs, citation positions, citation contexts, and citation types. Through a number of empirical experiments on such dataset, we show that the exploitation of the extracted knowledge, particularly the type of citation, is a promising approach for recommending recently published papers that may not be cited yet. The dataset, which we make publicly available, also represents a valuable resource for further investigation on academic information retrieval and filtering.

Keywords: Research publication dataset, citation context, citation types, recommender systems

1. Introduction

Keeping up with the most recent scientific literature is a challenge for many researchers giving the continuous and increasing growth of academic publications. A recent report claims the existence of 33,100 active scholarly peer-reviewed scientific, technical and medical English-language journals in mid-2018 (plus a further 9,400 non-English-language journals), collectively publishing over 3 million papers a year. The report also states that the production of scientific publications is steadily increasing at a 4% yearly rate. Aiming to address this information overload problem, academic search engines like Google Scholar or PubMed have emerged in the last years. RS in particular have focused on suggesting relevant papers for a given user [Middleton et al., 2001; Torres et al., 2004], but have also addressed other relevant tasks, including recommending relevant papers for a particular snapshot of content, such as title and abstract [Bethard and Jurafsky, 2010; Huang et al., 2015], recommending relevant papers for a given publication [Ekstrand et al., 2010; Liang et al., 2011a; Khadka and Knoth, 2018], recommending relevant papers for a particular collection of publications [Ekstrand et al., 2010; Shimbo et al., 2007], and recommending relevant papers for an undergoing manuscript, i.e., a paper yet to be published [He et al., 2010; Strohman et al., 2007].

Most of academic paper recommender systems, however, do not take into consideration the time when papers were published, and do not address the real-world problem of recommending recently published papers [Ha et al., 2014]. In fact, traditional collaborative filtering systems are ineffective to address that problem, since they are not able to recommend the latest, most recent papers, which have not been rated or cited yet. On the other hand, there are approaches that have explored the use of citation knowledge to enhance recommendations, e.g., citation graph (i.e., relations between papers based on citations), citation position (e.g., section of the paper where a citation appears), and citation context (i.e., text around a citation), the exploitation of citation types, i.e., categories of citations based on their purpose –background, criticism, etc.– to enhance the recommendation of scientific publications, and particularly the recommendation of recent scientific publications, is still under explored.

Addressing this gap, we investigate how citation knowledge could be extracted and exploited to support users towards the discovery of recent and relevant scientific publications. To capture such knowledge, access to the users’ publications and the textual content of papers is needed. Existing datasets used for training and testing academic recommender systems do not provide either the entire authors’ publication histories or the manuscript texts, but just their metadata, e.g., title, abstract and keywords (see Section 2.4.). As part of our work, we have built and made publicly available a novel dataset, enriched with different forms of citation knowledge, to enable the implementation and evaluation of RS for the particular task of recommending recently published papers to users. Hence, we provide the following contributions:

- An in-depth analysis of existing categorisations and taxonomies of citation types.
- A unique dataset that includes authors’ publication history and entire textual contents of scientific publications.
- An enrichment of such dataset with four levels of citation knowledge: citation graph, citation position, citation context, and citation type.
- An evaluation of how citation information, and partic-
The goal of our work is therefore the creation of a new dataset and the datasets they used are no longer available. A key works have explored the recommendation of recent papers, From a survey of the literature, we observe that Very few (CTR) that combines collaborative filtering with topic mod-
ated problem of 'out-of-matrix prediction', i.e., recomed-
Wang et al. (Wang and Blei, 2011) worked on the re-
not provided.

details on where and how the data could be collected are
fields of data mining. The used dataset is not available, and
published between 1971 and 2001 by 5,891 academics, in the
authors experimented with a dataset of 6,241 papers pub-
tory. The top-k scored papers are then recommended. The
approach exploits the graph to compute prob-
ations that are relevant to her work or research. Even
though the time factor has been previously applied in the
recommender systems literature to classify user prefer-
ences into short- or long-term interests (Sugiyama and Kan,
2010), or to suggest papers to new users for which there is
no prior activity records (Cai et al., 2016; Hristakeva et al.,
2017), to the best of our knowledge, only Ha et al. (Ha et
al., 2014) have considered the scenario of recommending
new items, i.e., recent academic papers. The authors pro-
posed a Belief Propagation approach where an undirected
graph is built to capture relations between papers based
on citations. The approach exploits the graph to compute prob-
abilistic scores for an author based on her publication his-
tory. The top-k scored papers are then recommended. The
authors experimented with a dataset of 6,241 papers pub-
lished between 1971 and 2001 by 5,891 academics, in the
fields of data mining. The used dataset is not available, and
details on where and how the data could be collected are
not provided.
Wang et al. (Wang and Blei, 2011) worked on the re-
related problem of "out-of-matrix prediction", i.e., recom-
mending papers that have never been rated (although this
does not necessarily mean that they are new or recent). The
authors proposed a collaborative topic regression model
(CTR) that combines collaborative filtering with topic mod-
eling. Their evaluation was conducted on a selection of CiteULike data but the dataset is not available either.
From a survey of the literature, we observe that Very few
works have explored the recommendation of recent papers,
and the datasets they used are no longer available. A key
goal of our work is therefore the creation of a new dataset
that can serve the RS community to further investigate how
to design and develop systems capable of recommending
the most recent and relevant scientific publications.

2. State of the Art

Our work focuses on the task of recommending recent sci-
entific publications to users, and exploring the use of ci-
tation knowledge to do so. Therefore, in this section we
review related literature according to two aspects: publica-
tion time awareness and citation knowledge awareness. In
addition, we provide a review of existing categorisations of
citation types, as well as a review of existing datasets used
in the field of academic recommender systems.

2.1. Publication Time Awareness

We address the real world problem where a user is inter-
rested in receiving notifications of recently published pub-
llications that are relevant to her work or research. Even

does not necessarily mean that they are new or recent). The
authors proposed a Belief Propagation approach where an undirected
graph is built to capture relations between papers based
on citations. The approach exploits the graph to compute prob-
abilistic scores for an author based on her publication his-
tory. The top-k scored papers are then recommended. The
authors experimented with a dataset of 6,241 papers pub-
lished between 1971 and 2001 by 5,891 academics, in the
fields of data mining. The used dataset is not available, and
details on where and how the data could be collected are
not provided.
Wang et al. (Wang and Blei, 2011) worked on the re-
related problem of "out-of-matrix prediction", i.e., recom-
mending papers that have never been rated (although this
does not necessarily mean that they are new or recent). The
authors proposed a collaborative topic regression model
(CTR) that combines collaborative filtering with topic mod-
eling. Their evaluation was conducted on a selection of CiteULike data but the dataset is not available either.
From a survey of the literature, we observe that Very few
works have explored the recommendation of recent papers,
and the datasets they used are no longer available. A key
goal of our work is therefore the creation of a new dataset

http://www.citeulike.org/faq/data.a

2.2. Citation Knowledge Awareness

A number of works have explored the use of citations
knowledge to provide better recommendations, including:

- Citation graph, where nodes represent papers and
  edges represent relations between such papers based
  on their citations. Relations in the citation graph can
  be either directed –i.e., they capture the explicit source
  and target papers of the citations–, or undirected –i.e.,
  they do not consider which paper is the one citing
  and which paper is the one being cited (Torres et al.,
  2004).

- Citation position, or the particular section of the pa-
  per where the citation appears, e.g., introduction, lit-
  erature review, and conclusions (Chakraborty et al.,
  2016).

- Citation context, or text surrounding the citation.
  This text provides an indication of the semantic con-
  text in which the citation is mentioned (Sugiyama and
  Kan, 2013).

- Citation type, or category of citation, which mainly
  reflects the objective of the citation: background, crit-
icise, support, etc. (Liang et al., 2011b)

Proposed approaches that extract and exploit citation
knowledge have mainly focused on recommending a list of
relevant papers for a given target paper (He et al., 2010;
Khadka and Knoth, 2018; Ekstrand et al., 2010; McNee et
al., 2002; Chakraborty et al., 2016; Liang et al., 2011b),
or on recommending relevant papers for a fragment of text
(Bethard and Jurafsky, 2010; Huang et al., 2015). Fewer
works, in contrast, have focused on recommending rele-
vant papers for a given user (Sugiyama and Kan, 2013;
Sugiyama and Kan, 2010; Torres et al., 2004), and to the
best of our knowledge, none of them have addressed the
particular use case of recommending recently published pa-
pers.
Sugiyama et al. (Sugiyama and Kan, 2010) explored the
use of the citation graph to enrich user profiles, capturing
users’ research interests by considering not only their past
publications, but also the citations of such publications, and
the publications citing the users’ work. In an extension of
that work (Sugiyama and Kan, 2013), the authors also ex-
plor the use of citation context, or text around the cita-
tions, to enhance recommendations.

Citation positions, in combination with the citation graph,
were explored by Chakraborty et al. (Chakraborty et al.,
2016). They generated a directed graph linking papers
based on citations and enriched with information about the
positions of such citations, i.e., an edge representing a cita-
tion from paper $p_1$ to paper $p_2$ may be tagged with multiple
labels, if $p_1$ cites $p_2$ in several sections. This graph is then
used to provide recommendations (for a given query paper).
Also in the context of recommending papers for a given pa-
paper, Liang (Liang et al., 2011b) proposed a method that ex-
plor citation graphs including citation types. Three types

- Citation type
- Citation position
- Citation context
of citation were considered: Based-on –referring to when the citation highlights a paper that the current work is based on, Comparable –referring to when the citation highlights a paper that is used to reflect on similarities or differences, and General –referring to all citations that cannot be categorized as Based-on or comparable, such as those that provide background knowledge, for example.

While these works show how citation knowledge can help enhancing recommendation accuracy, they do not consider the particular use case of recommending the latest scientific publications to users. Moreover, while most reported approaches have focused on the use of the citation graph, citation contexts, and citation positions, very few works, such as (Liang et al., 2011b), have explored the exploitation of citation types for recommendation.

A key goal of our work is therefore to provide language resources that (i) will enable the investigation of recommender systems for the particular real-world scenario of recommending recent scientific publications to users, and (ii) will provide rich citation knowledge, in the form of citation graph, citation positions, citation contexts, and particularly citation types. Since citation types have been unexplored in the field of recommender systems, the next section aims to provide a comprehensive overview of taxonomies proposed in other fields, identifying those citation types that have been commonly used.

### 2.3. Citation Types

While citation types (categories) are, to the best of our knowledge, underexplored in the RS field, there has been a considerable amount of effort invested in the definition and exploitation of citation types in other areas, such as Scientometrics. For example, they have been used to study the evolution of scientific fields (Jurgens et al., 2018a), or to quantify the influence of research works (Valenzuela et al., 2015) [Hassan et al., 2017]. In this section, we discussed different citation categorizations and taxonomies existing in the literature. We aim to provide an in-depth analysis of the similarities and differences of such categorizations and to conduct an informed selection of citation types for our research. A summary of this analysis is provided in Table 1.

For instance, Nanba et al. (Nanba and Okumura, 1999) proposed three classes: Type B (the citations to base on other researchers’ theories), Type C (the citations to compare with related works or to point out their problems), and Type O (the citations other than types B and C). Jurgens et al. (Jurgens et al., 2018a) considered six classes (Background, Compare or Contrast, Motivation, Use, Extend and Future) for measuring the evolution of scientific fields. Valenzuela and Hassan (Valenzuela et al., 2015) [Hassan et al., 2018] referred to two classes (important –using and extending– and not important –related and comparison–) for quantifying the importance of cited references.

As can be observed in the figure, a significant amount of works have proposed different types of citations. However, on a closer inspection, similarities emerge among existing taxonomies. For instance, (Nanba and Okumura, 1999) Nanba et al., 2000 used the types Basis, Compare and Other. These citation types are very similar in terms of semantic meaning to the ones proposed by Liang (Liang et al., 2011b): Based-on, Comparable and General.

The figure shows our summary and analysis on the different categorizations of citations identified in the literature, and proposes a high-level categorization, which includes the following types:

- **Background**: This type refers to citations providing general scientific knowledge towards a domain. This class has been named differently in previous works as background, organic, general, assumptive, read alert, and neutral, among others.

- **Discuss, compare and contrast**: This type refers to papers that are cited to discuss the proposed work or provide comparisons against it (i.e., to support it, dispute it, correct it, etc.). Two key subtypes in this category are also identified in the literature: Criticize and Support. These subtypes are also called as negative or negational (for citations that are used to express criticism), and positive or affiliational (for citations that are used to support the work).

- **Use**: This type refers to papers that are cited because the current work uses them. It has been also referred in the literature as technical basis, useful, or method.

- **Extension**: This type refers to papers that are cited because the current work extends them. It is also referred to as expand, evolutionary, developmental, and based on.

- **Motivation**: This type refers to cited papers that motivate or inspire the current work.

- **Future**: This type refers to cited papers that inspire future work.

Identifying the particular type of a citation is a complex task. While some works have focused on manually labelling citations based on proposed categories (Teufel, 1999) [Moravcsik and Murugesan, 1975], others have attempted to develop methods to automatic categorization. These methods are based on the use of cue phrases (Teufel et al., 2006b) [Teufel et al., 2006a] Liang et al., 2011b, syntactic patterns (Meyers, 2013), rule-based heuristics (Abu-Jbara et al., 2013) [Garzone and Mercer, 2000] Nanba et al., 2000, or Machine Learning (Hassan et al., 2018) Jurgens et al., 2018a] Cohan et al., 2019). Since the high-level citation types identified in our work (see Table 1) are very similar to the ones identified by (Jurgens et al., 2018a), we apply their classifier in our work, which was trained with computer science papers, an more particularly Natural Language Processing (NLP) papers.

### 2.4. Existing Datasets

Various datasets have been used in the past to investigate academic recommender systems. A comprehensive list of publicly available datasets is given in Table 2. For each dataset, the table shows a brief description of the type of data and time period when the data were collected, the number of users, items and ratings in the dataset, and whether
the full text $PDF_{av}$ and publication history of the users $UPH_{av}$ are available.

As we can observe, most of the existing datasets do not provide the full texts of publications, and hence, citation position, context and type cannot be extracted from them. Similarly, many datasets do not provide the authors’ publication history, and thus knowledge about the users, particularly their preferences (i.e., publications and references), cannot be easily captured. To address these limitations we have built a new dataset that includes both the full textual content of papers, and the authors’ publication history. The building process of our dataset is described next.

## 3. Dataset Building

We explain here the construction of the dataset, as well as how citation knowledge has been extracted from it. The dataset is accessible from [https://ordo.open.ac.uk/s/7dd3396c9d9ec0082b0c](https://ordo.open.ac.uk/s/7dd3396c9d9ec0082b0c) and will be publicly available after notification.

### 3.1. Collecting Data

As previously explained, we aimed to build a new dataset that (1) could serve to investigate the particular recommendation scenario of discovering the most recent academic papers relevant for a target user, and (2) provides the full textual content of papers in addition to their metadata, so that fine grained citation knowledge could be extracted and exploited in information retrieval and filtering tasks. Since we were interested in exploring the usage of citation knowledge for recommendation purposes, we needed to ensure that there are sufficient items (papers) cited by other items within the dataset. Following this requirement, we gathered the publication history of authors working on the same field (e.g., publishing in the same conference), since they are likely to cite each other’s publications. Specifically, we selected the ACM Conference Series on Recommender Systems (RecSys) and collected data for

https://recsys.acm.org/
1,931 authors who have published in the conference from the first RecSys conference in 2007 until the twelfth RecSys edition in 2018. The complete publication histories of these authors were collected from the well-known computer science bibliography data provider DBLP\footnote{https://www.theguardian.com/}

\footnote{https://dblp.uni-trier.de/}. Note that the publication history of an author contains not only her RecSys papers, but also papers published in other venues (journals, conferences, etc.).

As shown in Figure 1, an author’s publication history contains metadata of each of the author’s papers, including their titles, abstracts, publication dates and venues. The metadata also includes a URL to the corresponding Google Scholar page of each paper, which we used to gather the PDF file of the paper.

Out of the 80,808 papers crawled from DBLP, we obtained textual content for 35,473 of them (about 44%). We note that, while initiatives like open access and pre-prints enabled full access to scientific publications, many of their papers are hidden behind pay-walls and thus are not publicly accessible\footnote{https://orcid.org/}. Out of the 35,473 papers for which we have textual content, we could obtain citation types for 21,924 of them using the approach proposed by Jurgens et al.\footnote{https://github.com/kermitt2/grobid}. Then, to ensure that we had sufficient historical data to capture user preferences, we discarded all the authors for which we obtained less than 4 publications, keeping a total of 1,102 authors.

Afterwards, we divided the dataset into training and test sets by observing the publication time distribution, and selected as breaking date the 1st of January 2018 (see Figure 2). All papers published before that date were considered part of the training set. All papers published after that date were considered part of the test set. Lastly, we kept those authors having at least 60% of the data in the training set, and at least 10% of the data in the test set. The final dataset consists of 446 authors and 9,399 academic papers, from which 7,786 belong to the training set and 1,613 represent the test set.

### 3.2. Modelling Authors, Papers and Citations

Figure 1 shows the different features that are captured for users, publications and citations, as well as the different relations between them. For each user, we consider four different identifiers, including: the internal identifier within the dataset, the ORCID identifier\footnote{https://en.wikipedia.org/wiki/}

\footnote{https://www.theguardian.com/higher-education-network/2018/may/21/scientists-access-journals-researcher-article} and both, the DBLP and the Google Scholar URLs. In addition, we capture name, last name, website and affiliation.

For each publication, we also extracted multiple identifiers, including the internal identifier within the dataset, the DBLP URL (from where metadata about the paper were extracted) and the Google Scholar URL (from where the PDF file of the paper was downloaded –if available–). Metadata about the publication includes title, abstract, publication date, and publication venue.

To extract citation knowledge, we parsed the available PDF files using the GROBID parser\footnote{https://github.com/kermitt2/grobid} and the classifier of citation types provided by Jurgens et al.\footnote{https://jurgens.com/}.
into six main types: background, compare and contrast, use, extension, motivation and future. This classifier was trained with papers from the ACL Anthology corpus\footnote{\url{http://ranksys.org}} a corpus of scholarly publications about Computational Linguistics. To evaluate the performance of this classifier when identifying citation types from the RS field we conducted a manual assessment of 100 randomly selected citations. Based on 3 annotators (all of them with computer science background), displaying a moderate agreement (0.62 Fleiss kappa\cite{Fleiss1973}), we identified a 69% average correct classification, which is in line to the one reported by the authors\cite{Jurgens2018a}. The 1,071 distinct referenced publications were classified 9,933 times as background, 3,032 as CompareOrContrast, 225 times as Extend, 449 times as Use 83 times as Motivation and 39 times as Future.

### 3.3. Modelling User Preferences

We distinguished between two main ways of capturing user preferences. On the one hand, we considered that a user has a preference (positive rating) for all the papers she has authored. Authored papers encapsulate the user’s interest in terms of research areas, topics, methods, etc., and constitute a relevant source of information to build her profile. On the other hand, we considered that a user has a preference for all the papers she has authored as well as the papers she has cited, since cited papers also encapsulate research that the user considers relevant in relation to her work. By doing so, we explore the use of the citation graph to enrich user profiles.

Figure\footnotemark[2] illustrates these two user preference models. In the left part of the figure, we show a rating matrix $R_P$ relating authors (rows) and papers (columns) where a cell has a value 1 if the corresponding user authored the associated paper, and 0 otherwise. In the right part of the figure, we show a rating matrix $R_{PC}$ where a cell has a value 1 if the user authored or cited the paper, and 0 otherwise. In addition to these rating matrices, we also considered two enriched versions of $R_{PC}$: (i) $R_{PCX}$, where $X$ stands for context, and $R_{PCXT}$, where $T$ stands for type. In case of $R_{PCX}$ for every rating based on citations, it captures the position (section in the paper) and the text around a citation. In the case of $R_{PCXT}$, for every rating based on citations, it captures the citation type and the text around a citation. In the experiments, the above four matrices were split into training and test sets according to a target time, in particular, the 1st of January 2018. The final dataset splits are captured in Table\footnotemark[3]

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{P}^{training}$</td>
<td>446 users, 7, 786 items and 9, 348 ratings</td>
</tr>
<tr>
<td>$R_{P}^{test}$</td>
<td>446 users, 1, 613 items and 2, 033 ratings</td>
</tr>
<tr>
<td>$R_{PC}^{training}$</td>
<td>446 users, 7, 786 items and 1, 104 ratings</td>
</tr>
<tr>
<td>$R_{PC}^{test}$</td>
<td>446 users, 1, 613 items and 2, 126 ratings</td>
</tr>
<tr>
<td>$R_{PCX}^{training}$</td>
<td>446 users, 7, 786 items and 13, 104 ratings, and citation context and position information</td>
</tr>
<tr>
<td>$R_{PCX}^{test}$</td>
<td>446 users, 1, 613 items and 2, 126 ratings, and citation context and position information</td>
</tr>
<tr>
<td>$R_{PCXT}^{training}$</td>
<td>446 users, 7, 786 items and 13, 104 ratings, and citation context and type information</td>
</tr>
<tr>
<td>$R_{PCXT}^{test}$</td>
<td>446 users, 1, 613 items and 2, 126 ratings, and citation context and type information</td>
</tr>
</tbody>
</table>

### Table 3: Capturing user preferences: training and the test sets

### 4. Evaluation

To assess the created dataset and investigate whether citation knowledge does indeed help enhancing the recommendation of recent scientific publications we present here a series of experiments comparing a novel hybrid recommendation method that explores the previously captured citation knowledge, against various baselines (standard recommendation methods). In this section, we describe the followed evaluation set-up and obtained results.

#### 4.1. Evaluation Setting

Next we present our proposed recommendation method, the baselines considered for comparison, and the metrics used for evaluation. The evaluations were conducted through the RiVal framework\cite{Said2014a}.

#### 4.1.1. Baselines

We experimented with the following baselines: (i) content-based filtering\cite{Lops2011a}, (ii) Matrix Factorisation (MF)\cite{Koren2009a}, (iii) user-based collaborative filtering, (iv) item-based collaborative filtering\cite{Ricci2010}, Factorisation Machine (FM)\cite{Rendle2010}, and the Random method from RankSys framework\footnote{\url{http://ranksys.org}}. In addition, we evaluated ItemRank\cite{Gori2007}, a popular graph-based recommendation method. It is important to note that certain baselines, such as collaborative filtering cf, are not able to provide recommendations in the proposed scenario where new items (for which no ratings are available) are recommended to users. Among the evaluated baselines, the best performing one was the content-based cb method. Hence, for simplicity, we report results against cb in Table\footnotemark[4]. Content-based methods
recommend items (academic papers) to a user that are ‘similar’ to those she positively rated, i.e., authored or cited. The similarity between users and items is computed based on profiles built from textual information. User preferences and item attributes correspond to text features; in our case, keywords extracted from the titles and abstracts of the papers. Recommendations are generated by means of user and item similarities in the text feature space.

More formally, an item \( i_n \)’s profile consists of a vector \( \mathbf{v}_{i_n} = [v_{i_n,1}, v_{i_n,2}, ..., v_{i_n,L}] \in \mathbb{R}^L \) where \( v_{i_n,l} \) denotes the relative weight (relevance) of feature \( f_l \) for \( i_n \), and \( L \) is the number of existing features. To compute the weights \( w_{m,l} \), we used TF-IDF (Jones, 1972).

Similarly, a user \( u_m \)'s profile is represented as a vector \( \mathbf{v}_{u_m} = [v_{u_m,1}, v_{u_m,2}, ..., v_{u_m,L}] \in \mathbb{R}^L \), where \( v_{u_m,l} \) denotes the relative weight (relevance) of feature \( f_l \) for \( u_m \), and \( L \) is computed by aggregating the titles of all the papers that have a rating associated to the user, i.e., all the papers for which the user has expressed an interest. The recommendation score of an item \( i \) for a target user \( u \) is then computed as the cosine similarity \( \text{score}(u,i) = \cos(u,i) \). We refer to this method as \( \text{cb} \).

The textual features utilised to build user profiles for \( \text{cb} \) varies according to the available citation knowledge (see Section 3.3). For \( R_P \), a user’s profile is built by considering the titles of the papers the user authored. For \( R_{PC} \), a user’s profile is built by considering the titles of the papers she authored and the papers she cited. Lastly, for \( R_{PCX} \) and \( R_{PCXT} \), a user’s profile is built by considering the titles of the papers she authored, the titles of the papers she cited, and the citation contexts, i.e., the texts around citations within the papers.

### 4.1.2. Proposed Hybrid method

Next, we present a series of hybrid recommendation methods, \( \text{hyb} \), that jointly exploit the content of the papers and the user-item ratings (see Table 4) to provide personalised recommendations. Hybrid methods (Burke, 2002) aim to mitigate the disadvantages of individual approaches by combining the strengths of various methods, in general, content-based and collaborative filtering. In our case, we aim to mitigate the ineffectiveness of \( \text{cf} \) when recommending the latest scientific publications by combining it with \( \text{cb} \) and exploiting the extracted citation knowledge.

The proposed approach, \( \text{hyb} \), is based on the item-based \( \text{cf} \) nearest neighbour heuristic\(^{13}\), where content-based features are used to compute item similarities. In item-based \( \text{cf} \) algorithms (Sarwar et al., 2001), similarities between items are used to estimate scores for a (user, item) pair. In our case, item (paper) profiles are generated based on textual features, which vary with respect to the available citation knowledge (see Section 3.3): for \( R_P \), item profiles are build from the title of the item; for \( R_{PC} \), item profiles are build from the title of the item and the titles of the cited papers; and for \( R_{PCX} \) and \( R_{PCXT} \), item profiles are build from the title of the item, the titles of the cited papers, and the citations’ context (i.e., texts around citations). We formulate our \( \text{hyb} \) method in Equation 1:

\[
\hat{r}_{u,i} = \frac{\sum_{i' \in N(i)} \text{Sim}(i,i') \cdot \sum_{l \in n} w_{i',l} \cdot \text{Sim}(i',l)}{\sum_{i' \in N(i)} |\text{Sim}(i,i')|}
\]

where \( \hat{r}_{u,i} \) is the preference score to be predicted for the target user \( u \) and item \( i \), \( \text{Sim}(i,i') \) is the similarity between the interacted item \( i' \) and an item \( i \) from \( i \)'s neighbourhood \( N(i) \). Cosine similarity is used to measure the similarity between items. Finally, \( r_{u,i'} \) is the preference (rating) given by user \( u \) to an item \( i' \). We also use different sizes of neighbourhoods, specifically 5, 10, 15 and 20.

To investigate the relevance of the different types of citations, we further modified our hybrid method by incorporating a weight, \( w_{u,i'} \) in the eq. 1 heuristic, that reflects the strength of an item \( i' \) for a user \( u \) based on the different citation types, named as \( \text{hybType} \). Equation 2 represents \( \text{hybType} \) hybrid method. We formulate this method \( \text{hybType} \) in Equations 2 and 3:

\[
\hat{r}_{u,i} = \frac{\sum_{i' \in N(i)} \text{Sim}(i,i') \cdot r_{u,i'} \cdot w_{u,i'}}{\sum_{i' \in N(i)} |\text{Sim}(i,i')|}
\]

where the strength (weight) \( w_{u,i'} \) is computed by considering all the instances where \( i' \) is cited by \( u \). Note that an item \( i' \) may be cited by \( u \) in several publications, and with different citation type of the same publication. Then, the weight is normalised by the total number of instances. More formally, the strength \( w_{u,i'} \) of item \( i' \) for user \( u \) is calculated as:

\[
w_{u,i'} = \frac{\sum_{n=1}^{n} (w_{jbg} + w_{jcom} + w_{jmot} + w_{jaux} + w_{jext} + w_{jfor})}{|w_{u,i'}|}
\]

where \( n_{u,i'} \) is the number of times \( i' \) is cited by \( u \) in her papers, and \( w_{jbg}, w_{jcom}, w_{jmot}, w_{jaux}, w_{jext}, w_{jfor} \) reflect the number of times \( i' \) is cited as background, compareOr-Contrast, motivation, use, extension, or future, respectively.

### 4.2. Evaluation Results

We summarise the results obtained from our experiment in Table 4. The foremost conclusion we draw from the experiment is that the incorporation of citation knowledge helps improving the performance of recommendation methods. Between the rating matrices \( R_P \) and \( R_{PC} \) (where \( R_{PC} \) reduces the sparsity because of the additional citation information added to it), \( \text{cb} \) and our \( \text{hyb} \) method show improvements with all the evaluation metrics at the cut-off points of 5 and 10. The improvement continues in all the sizes of neighbourhoods, i.e., 5, 10, 15 and 20. This shows that the incorporation of user preferences based on the citation graph, allows finding valuable relationships between items and users that are productively exploited.

When incorporating citation knowledge in the form of citation context, captured in the \( R_{PCX} \) matrix, we notice further increment in the performance of the hybrid methods, but a decrease in the performance of the \( \text{cb} \) method. The saturation in the textual features between \( R_{PC} \) and \( R_{PCX} \) may be the cause of this drop in performance.

Exploiting citation knowledge in the form of citation types (\( \text{hybType} \)), captured in the \( R_{PCXT} \) matrix, outperformed all other methods (i.e., \( \text{hyb} \) and \( \text{cb} \)) for all the evaluation.

\(^{13}\)We also tested the user-based \( \text{cf} \) heuristic, but discarded it due to its non competitive performance results.
metrics. This shows that citation type is a useful feature for recommending recent papers to users. In particular, the performance of our hypType method is higher when the citation belongs to a Future or Extension categories. This corroborates the intuition that citations that are related to future research and to work that is being extended or enhanced, represent relevant pointers for new directions in a research field. As can be seen in table 4, precision, recall and nDCG@5 values are higher when citations belong to Future category, while nDCG, precision and recall@5 values are higher when the citation type is Extension. In this context, whereas higher precision and recall values show the methods were able to find relevant and new items, the higher nDCG value shows that novel documents are appearing earlier in the recommendation lists. The Motivation type has also positive impact on the recommendation. Citations that motivate the work are helpful on reflecting problems that may be targeted not only in the present, but also in future scientific publications.

5. Discussion and Future Work

In this work, we have addressed the problem of recommending recent scientific publications to users, and investigated the extraction and exploitation of citation knowledge for such purpose. In doing so we have generated a novel dataset capturing (i) users’ preferences (in the form of publication and citation history) and, (ii) different notions of citation knowledge including the citation graph, citation positions, citation contexts, and citation types. To the best of our knowledge, this is the first available dataset capturing the above information, and hence a unique resource to enable the investigation of an important real-world problem, the personalised recommendation of recent scientific publications to users.

A particular interesting form of citation knowledge captured in this dataset is citation types, including background, compare or contrast, use, extension, motivation and future. Citation types are not only useful to enhance paper recommendation (as shown in our evaluation), but can also enable further refinement of the recommendation process based on the users’ intention. For instance, users with an interest on how to use or apply certain algorithms or techniques, may receive recommendations for papers cited under the type ‘use’ or ‘extension’. Refining recommendations based on the users’ intent, in addition to the users’ publication and citation history, is one of our future lines of work. We also aim to develop recommendation methods based on the combination of different notions of citation knowledge, particularly citation context, type and position, since these three elements capture fine-grained patterns of the intent with which papers are cited. Despite the timeliness and potential of this dataset we also acknowledge several limitations:

Data sampling: This dataset is representative of the area of RS. Complementing it with papers from different fields will help to assess whether the obtained findings are specific to the RS research field or are representative of other areas as well.

Data annotation: Although we tested that the classifier provided by [Jurgens et al., 2018] achieved comparable results when identifying citation types for our collected papers, the average obtained accuracy is 69%, indicating the presence of noise in the classification process. In addition, the PDF parser used as part of this classifier could only extract citation information for 61% of the collected papers, reducing the completeness of the data. Providing more effective citation type classification methods is an interesting research question for the analysis, search and recommendation of scientific publications.

Data filtering: In order to capture users’ preferences we discarded from our dataset all authors for which we obtained less than 4 publications. Our dataset therefore does not capture the scenario of recommending recent papers to users for which no preferences have been gathered.

Despite the above limitations, our work provides a unique resource that will help the RS community to further investigate the problem of recommending recent scientific publications to users. We also hope that the use of this dataset will stimulate discussions across related disciplines (scientometrics, information retrieval, etc.) on the usefulness of citation knowledge to effectively target the problem of information overload in the scientific world.

Table 4: Experiment results of the baselines and proposed hybrid recommendation methods. A gray scale is used to highlight better (dark gray) and worst (white) values for each ranking metric. For every metric, the best values are highlighted in bold.

<table>
<thead>
<tr>
<th>matrix</th>
<th>method</th>
<th>p@5</th>
<th>p@10</th>
<th>r@5</th>
<th>r@10</th>
<th>p@5</th>
<th>p@10</th>
<th>r@5</th>
<th>r@10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hyb5</td>
<td>0.039</td>
<td>0.045</td>
<td>0.073</td>
<td>0.08</td>
<td>0.053</td>
<td>0.064</td>
<td>0.065</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>hyb10</td>
<td>0.043</td>
<td>0.045</td>
<td>0.073</td>
<td>0.08</td>
<td>0.053</td>
<td>0.064</td>
<td>0.065</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>hyb15</td>
<td>0.043</td>
<td>0.045</td>
<td>0.073</td>
<td>0.08</td>
<td>0.053</td>
<td>0.064</td>
<td>0.065</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>hyb20</td>
<td>0.043</td>
<td>0.045</td>
<td>0.073</td>
<td>0.08</td>
<td>0.053</td>
<td>0.064</td>
<td>0.065</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>CB</td>
<td>0.047</td>
<td>0.052</td>
<td>0.083</td>
<td>0.095</td>
<td>0.056</td>
<td>0.046</td>
<td>0.071</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>Hybrid5</td>
<td>0.042</td>
<td>0.047</td>
<td>0.075</td>
<td>0.081</td>
<td>0.055</td>
<td>0.044</td>
<td>0.074</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>Hybrid10</td>
<td>0.046</td>
<td>0.052</td>
<td>0.082</td>
<td>0.09</td>
<td>0.055</td>
<td>0.047</td>
<td>0.095</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>Hybrid15</td>
<td>0.042</td>
<td>0.047</td>
<td>0.075</td>
<td>0.081</td>
<td>0.055</td>
<td>0.044</td>
<td>0.074</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>Hybrid20</td>
<td>0.047</td>
<td>0.052</td>
<td>0.083</td>
<td>0.095</td>
<td>0.056</td>
<td>0.047</td>
<td>0.077</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>hybType5</td>
<td>0.066</td>
<td>0.078</td>
<td>0.101</td>
<td>0.126</td>
<td>0.067</td>
<td>0.062</td>
<td>0.111</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>hybType5 (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)</td>
<td>0.066</td>
<td>0.078</td>
<td>0.101</td>
<td>0.126</td>
<td>0.067</td>
<td>0.062</td>
<td>0.111</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>hybType5 (0, 0, 0, 1, 0)</td>
<td>0.064</td>
<td>0.075</td>
<td>0.096</td>
<td>0.121</td>
<td>0.06</td>
<td>0.058</td>
<td>0.106</td>
<td>0.161</td>
</tr>
<tr>
<td></td>
<td>hybType5 (0, 1, 0, 0, 0, 0, 0, 0, 0, 0)</td>
<td>0.067</td>
<td>0.079</td>
<td>0.101</td>
<td>0.126</td>
<td>0.067</td>
<td>0.062</td>
<td>0.111</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>hybType5 (0, 1, 0, 0, 0, 0, 0, 0, 0, 0)</td>
<td>0.046</td>
<td>0.055</td>
<td>0.069</td>
<td>0.089</td>
<td>0.04</td>
<td>0.039</td>
<td>0.074</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>hybType5 (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)</td>
<td>0.041</td>
<td>0.048</td>
<td>0.063</td>
<td>0.078</td>
<td>0.037</td>
<td>0.034</td>
<td>0.063</td>
<td>0.104</td>
</tr>
</tbody>
</table>

In order to capture users’ preferences we discarded from our dataset all authors for which we obtained less than 4 publications. Our dataset therefore does not capture the scenario of recommending recent papers to users for which no preferences have been gathered. Despite the above limitations, our work provides a unique resource that will help the RS community to further investigate the problem of recommending recent scientific publications to users. We also hope that the use of this dataset will stimulate discussions across related disciplines (scientometrics, information retrieval, etc.) on the usefulness of citation knowledge to effectively target the problem of information overload in the scientific world.
6. References


Teufel, S. (1999). Argumentative zoning information extraction from scientific text.

