Citation Recommendation: Approaches and Datasets

Michael Färber · Adam Jatowt

Abstract Citation recommendation describes the task of recommending citations for a given text. Due to the overload of published scientific works in recent years on the one hand, and the need to cite the most appropriate publications when writing scientific texts on the other hand, citation recommendation has emerged as an important research topic. In recent years, several approaches and evaluation data sets have been presented. However, to the best of our knowledge, no literature survey has been conducted explicitly on citation recommendation. In this article, we give a thorough introduction into automatic citation recommendation research. We then present an overview of the approaches and data sets for citation recommendation and identify differences and commonalities using various dimensions. Last but not least, we shed light on the evaluation methods, and outline general challenges in the evaluation and how to meet them. We restrict ourselves to citation recommendation for scientific publications, as this document type has been studied the most in this area. However, many of the observations and discussions included in this survey are also applicable to other types of text, such as news articles and encyclopedic articles.

1 Introduction

Citing sources in text is essential in many scenarios. Most prominently, citing has always been an integral part of academic research. Scientific works need to contain appropriate citations to other works for several reasons [155]. Most notably, all claims written by the author need to be backed up in order to ensure transparency, reliability, and truthfulness. Secondly, mentions of methods and data sets and further important domain-specific concepts need to be linked via references in order to help the reader to properly understand the text and to give attribution to the corresponding publications and authors (see Table 1). However, citing properly has become increasingly difficult due to the dramatically increasing number of scientific publications published each year [25,163,58] (see also Fig. 1). For instance, in the computer science domain alone, more than 100,000 new papers are published every year and three times more papers were published in 2010 than in 2000 [92]. A similar trend can be observed in other disciplines [94]. For instance, in the medical digital library database PubMed, the number of publications in 2014 (514k) was more than triple the amount published in 1990 (137k) and more than 100 times the amount published in 1950 (4k) [26]. Due to this phenomenon of information overload in science in the form of a "tsunami of publications," citing appropriate publications has become an increasing challenge for scientific writing.

As a consequence, approaches for citation recommendation have been developed. Citation recommendation refers to the task of recommending appropriate citations for a text passage within a document. For instance, given the phrase "and similarly, the emergence of GANs has led to significant improvements in human image synthesis" within a document, a citation recommendation system might insert two citations as follows: "and similarly, the emergence of GANs [1] has led to significant improvements in human image synthesis [2]." This would mean adding corresponding references to (1) a publication introducing generative adversarial networks (GANs), and (2) a publication backing up

Karlsruhe Institute of Technology (KIT), Kaiserstr. 89, 76133 Karlsruhe, Germany, E-mail: michael.faerber@kit.edu · Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan, E-mail: adam@dl.kuis.kyoto-u.ac.jp
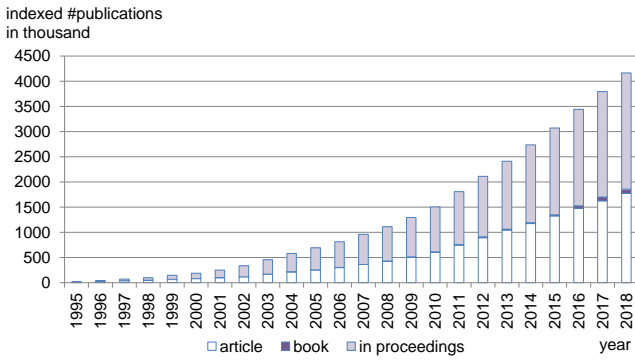
indexed #publications
in thousand

Fig. 1 Growth of scientific publications indexed in DBLP from 1995 until 2018. Data source: `http://dblp.org/statistics/recordsindblp`.

Table 1 Examples for in-text citations from Färber and Sampath [111].

Citation type	Example sentence
CONCEPT	"To this end, SWRL [14] extends OWL-DL and OWL-Lite with Horn clauses."
CLAIM	"In the traditional hypertext Web, browsing and searching are often seen as the two dominant modes of interaction (Olston & Chi, 2003)."
AUTHOR	"Gibson et al. [12] used hyperlink for identifying communities."

the statement concerning improvements in human image synthesis. Added references in such a scenario need to fit semantically to the context within the citing document and may be required to meet further constraints (e.g., concerning their recency).

Note that citation recommendation differs from paper recommendation [15,142]: *paper recommendation* aims to recommend documents to the user that are worthwhile to read and to investigate (particularly, in the context of a research topic). To that end, one or several papers [165,66,4,138] or the user's already clicked/bookmarked/written documents [8,95] can, for instance, be used for the recommendation. We can refer to [15, 11] for surveys on paper recommendation. *Citation recommendation*, by contrast, assists the user in substantiating a given text passage (e.g., written claim or scientific concept) within an input document by recommending publications that can be used as citations. The textual phrase to be backed up can vary in length – from one word up to a paragraph – and is called *citation context*. In some cases [71,103], the citation context needs to be discovered before the actual citation recommendation. While some existing works consider citation recommendation as a task of extending the set of known references for a given paper [82,83,64], we consider citation recommendation purely as a task for substantiating claims and concepts in the citation context. This

makes citation recommendation context-aware and very challenging, because the concept of relevance is much stricter than in ad hoc retrieval [144]. Consequently, citation recommendation approaches have been proposed using additional information besides the citation context for the recommendation, such as the author's name of the input document [48]. Evaluating a citation recommendation approach requires to verify if the recommended papers are relevant as citations for given citation contexts. For scalability reasons, usually the citations in existing papers and their citation contexts are used as ground truth (see Sec. 5.1).

Existing surveys focus only on related research areas of citation recommendation, but not explicitly on citation recommendation itself. Among the most closely related studies are the surveys on paper recommendation [15,11]. In these articles, the authors do not consider recommender systems for given citation contexts. Several surveys on other aspects of citation contexts have also been published. Alvarez et al. [7] summarize and discuss works on the identification of citation contexts, on the classification of each citation's role (called *citation function*), and on the classification of each citation's "sentiment" (called *citation polarity*). Ding et al. [44] focus on the content-based analyses of citation contexts, while White [164] considers primarily the classification of citations into classes, the topics covered by citation contexts, and the motivation of citing. Moreover, distantly related to this survey, surveys on the analysis of citing behavior [24,150] and surveys on works about the analysis of citation networks exist, for instance, for the purpose of creating better measurements of the scientific impact of researchers or communities [159]. Dedicated approaches and data sets for citation recommendation are not covered in all those works, nor is there any analysis of citation recommendation evaluations and evaluation challenges. This makes it necessary to consider citation recommendation separately and to use task-specific dimensions for comparing the approaches.

We make the following contributions in this survey:

1. We describe the process of citation recommendation, the scenarios in which it can be applied, as well as the advantages it has in general.
2. We systematically compare citation recommendation to related tasks and research topics.
3. We outline the different approaches to citation recommendation published so far and compare them by means of specifically introduced dimensions.
4. We give an overview of evaluation data sets and further working data sets for citation recommendation and show their limitations.

5. We shed light on the evaluation methods used so far for citation recommendation, we point out the challenges of evaluating citation recommendation approaches, and present guidelines for improving citation recommendation evaluations in the future.
6. We outline research directions concerning citations and their recommendations.

Several reader groups can benefit from this survey: non-experts can obtain an overview of citation recommendation; the community of citation recommendation researchers can use the survey as the basis for discussions of critical points in approaches and evaluations, as well as for getting suggestions for future research directions (e.g., research topic suggestions for PhD candidates); and finally, the survey can assist developers in choosing among the available approaches or data sets.

The rest of this article is structured as follows: in Section 2, we introduce the field of citation recommendation to the reader. In Section 3, we describe how we collected publications presenting citation recommendation approaches. We propose classification dimensions and compare the approaches by these dimensions. In Section 4, we give an overview of evaluation data sets and compare the data sets by corresponding dimensions. Section 5 gives a systematic overview of the evaluation methods that have been applied so far and of the challenges that emerge when evaluating citation recommendation approaches. Section 6 is dedicated to potential future work. The survey closes in Section 7 with a summary.

2 Citation Recommendation

2.1 Terminology

In the following, we define some important concepts of citation recommendation, which we use throughout the article. In order to have a generic task formalization, as we prefer, we do not restrict ourselves to scientific papers as a document type, but consider text documents in general.

The basic concept of citing is depicted in Fig. 2. A *citation* is defined as a link between a *citing document* and a *cited document* at a specific location in the citing document. This location is called the *citation marker* (e.g., "[1]") and the text fragment which should be supported by the citation is called the *citation context*. During processing, the citation context can be transformed into an *abstract representation*, such as an embedding vector [48, 21] or a translation model [74, 76]. This enables us to more accurately match the information in the citation context with the informa-
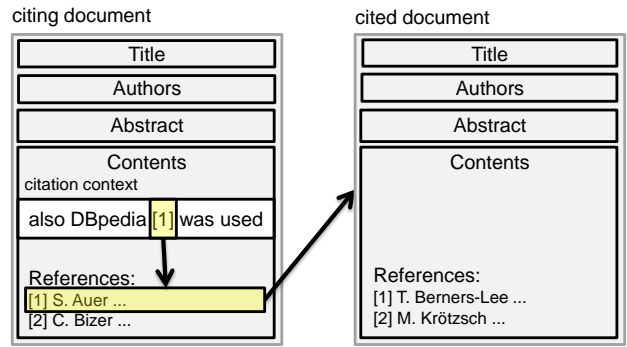

Fig. 2 Visualization of a citation in a scientific paper.

tion provided in the *"citable" documents* (also called *candidate cited documents*).

"References" and "citations" are often used interchangeably in the literature. However, we name in-text references, given by citation markers, *citations*. *References*, in contrast, are listed in the *reference section* of the citing document and describe links to other documents on a document level without context.

In the academic field, both the citing documents and the cited documents are usually scientific *papers*. We use the terms *paper*, *publication*, and *work* interchangeably in this article. The *authors* of scientific papers are usually *researchers*. We then use *researcher* and *scientist* interchangeably. Researchers who use a citation recommendation system become *users*.

Citing documents and cited documents consist of *content* and *metadata*. In the case of scientific papers, the paper's metadata typically consists of the title, the author information, an abstract, and other information, such as the venue in which the paper has been published.

Different citation context lengths can be used for citation recommendation. If only a fragment of an input text document is used as citation context (e.g., a sentence [69, 74] or a window of 50 words), we call it *local citation recommendation* or *context-aware citation recommendation*. If no specific citation context, but instead the whole input text document or the document's abstract is used for the recommendation (see, e.g., [144, 119, 151, 89]), we call it *global citation recommendation* or *non-context-aware citation recommendation* (following He et al. [72]). In the following sections, we will primarily focus on *local citation recommendation*, since only this variant targets the recommendation of papers for backing up single concepts and claims in a text fragment (i.e., assists the user in the actual citing process) and has not been addressed in other surveys, to the best of our knowledge.[1]

---

[1] It should be noted that it is also possible to design global context-aware citation recommendation approaches, i.e., ap-

2.2 Scenarios, Advantages, and Caveats of Citation Recommendation

In the "traditional" process of finding appropriate citations, the researcher needs to come up with candidate publications for citing on her own. The candidate papers that can be cited are either already known by her, are contained in a given document collection, or first need to be discovered. For the last option, the scientist typically uses widely used bibliographic databases, such as Google Scholar,[2] or domain-specific platforms such as DBLP[3] or PubMed.[4] The search for candidate papers to cite typically requires considerable time and effort as well as skills: the right keywords for querying need to be found, and the top n returned documents need to be manually assessed with regards to their relevance to the citing document and to the specific citation contexts.

The idea of citation recommendation is to enhance the citing process: The user provides the text she has written (with or without initial citations) to the recommender system. This system then presents to the user for specific segments of the input text all publications which were determined automatically as suitable citations. The user can investigate the recommendations in more detail and approve or disapprove them. Following this procedure, the tedious manual, separate search in bibliographic databases and paper collections can be considerably reduced (and maybe even skipped). The user does not need to think of meaningful keywords for searching papers any more. Last but not least, citing may become less dependent on the (often very limited) set of papers known to the current user.

We do not want to hide that citation recommendation can also entail problematic features if applied inadequately. Firstly, if citing becomes purely automated, the role of citations might change (e.g., instead of criticizing, citations might support a statement; see [155, 117, 154] for citation function schemes). The trust in citations might decrease, since machines (here: recommender systems) might not engender as much trust as experts who have dealt with the topic. We thus argue that a human-in-the-loop is still needed for citation recommendation. Secondly, if the recommendation models are trained on a fixed publication data set, instead

of removing citation biases, the recommender systems could introduce additional biases towards specific papers. Therefore, it must be ensured that a sufficiently large number of papers is indexed and that the new papers are indexed periodically. Caveats of citation recommendation are discussed in depth in Sec. 5.

Citation recommendation systems can be designed for several user groups:

(1) Expert Setting. In this setting, a researcher is familiar with her research area and is in the process of writing an expert text, such as a scientific publication (e.g., after having developed a novel approach or for conducting a survey in her research field). Recommendations of citations can still be beneficial for her, as such a user might still be unaware of publications in their field in the light of the "tsunami of publications" common in all scientific fields nowadays [25, 163, 58]. Citation recommendation systems might come up with recommendations which were not in the focus of the researcher if she cited in the traditional way, since the system might be able to bridge language barriers [87, 153] and also find publications which use synonyms or otherwise related concepts.

(2) Non-Expert Setting. We can think of several non-expert user types for which citation recommendation can be beneficial:

– A researcher needs to write a scientific text on a topic that is outside of her core research area and expertise (e.g., generic research proposals [20] and potential future work descriptions).
– A journalist in the science domain – e.g., authoring texts for a popular science magazine – needs to write an article on a certain scientific topic [130, 124]. We can assume that the journalist typically is not an expert on the topic she needs to write about. Having citations in the text helps to substantiate the written facts and make the text more complete and understandable.
– "Newcomers" in science, such as Masters students and PhD students in their early years, are confronted with the vast amount of citable publications and typically do not know any or all of the relevant literature in the research field yet [71, 174]. Getting citations recommended helps not only students in writing systematic and scientific texts, such as research project proposals (exposés), but also their mentors (e.g., professors).

In all these non-expert settings, the relevance of the recommended citations is presumably not so much determined by the timeliness of the publications, as in the expert setting, but instead more by the general importance and prominence of the publications. Thus, the

---

proaches which recommend citations for specific contexts (e.g., sentences) but which take the whole paper into account (e.g., to ensure an even greater understanding of the context or to diversify the recommendations). However, we are not aware of any such approach being published (see also Section 6 for potential future work).

[2] http://scholar.google.com/

[3] http://dblp.org/

[4] http://www.ncbi.nlm.nih.gov/pubmed/

relevance function for finding the most appropriate citations might vary from setting to setting.

Besides the pure topical relevance of recommended citations, also the fit from a social perspective might be essential. In recent decades, the citing behavior of scientists has been studied extensively in order to find good measurements for the scientific impact of scientists and their publications [24]. In this context, several biases in citing have been considered. Most notably, the hypothesis has been made that researchers tend to cite publications which they have written themselves or which have been written by colleagues [78]. Another hypothesis is that very prominent and highly cited works get additional citations only due to their prominence and visibility in the community (see, e.g., [164]). Citation recommendation systems can help in reducing biases by recommending citations which are the best fit for the author, the citation context with its argumentation, and the community.[5] Section 5.2 discusses citing bias in the context of citation recommendation in detail.

Overall, we can summarize the benefits of citation recommendation as follows:

1. Finding suitable citations should become more effective. This is because the match between the query (citation context) and the citable documents is more sophisticated than via manual matching (e.g., also considering synonyms, related topics, etc.). Furthermore, the recommender system typically covers a much larger collection of known publications than the set of documents known to the user.
2. Researchers are more (time-)efficient during the process of citing, as the number and extent of manual investigations (using bibliographic databases or own document collections) are reduced, and because recommendations are returned immediately.
3. The search for publications which can be cited becomes easier and more user-friendly ("citing for everyone"). As a consequence, citing is no longer just a "privilege" for experts, but potentially something for almost anyone.
4. By establishing a formal relevance function dealing with the issue of which papers are cited and what characteristics they have, the process is no longer left to chance. Hence, biases in citing behavior can be minimized.
5. Ideally, citation recommendation systems only recommend citations for valid statements and existing concepts, while unexaminable statements are not cited. Hence, citation recommendation implies an implicit fact checking process by showing sources to the user which support the written statements.

6. Advanced citation recommendation systems can, in addition, search for suitable, cite-worthy publications in other languages than the citing document (cross-linguality). They can also recommend publications under the special consideration of topic evolution over time, of current buzzwords, or in a personalized way, by incorporating user profiles.

2.3 Task Definition

In the following, we define local citation recommendation. By considering the whole document, abstract, or title as citation context, this definition can also serve as definition for global citation recommendation. The general architecture of a context-aware citation recommendation system is depicted in Fig. 3. State-of-the-art citation recommendation approaches are supervised learning approaches. Thus, we can distinguish between an *offline step* (or *training phase* in machine learning setups), in which a recommendation model is learned based on a collection of documents, and an *online step* (or *testing/application phase*), in which the recommendation model is applied to a new incoming text document. Note, however, that unsupervised learning approaches and rule-based approaches are also possible (although, to date, to the best of our knowledge, none such have been proposed). In that case, the learning phase in the offline step is eliminated and a given model (e.g., set of rules) can be directly applied (see Fig. 3).

In the following, we give an overview of the steps in case of supervised learning (using the symbols summarized in Table 2). Note that existing citation recommendation approaches use, to the best of our knowledge, content-based filtering techniques and are not based on other recommendation techniques, such as collaborative filtering or hybrid models. It is therefore not surprising that the approaches are mostly not personalized[6] (i.e., not incorporating user profiles). Hence, our task formalization does not consider personalization.

2.3.1 Offline Step

Input Input is a set of documents $D = \{d_1, ..., d_n\}$, which we call in the following the *citing documents*, with citations and references.[7]

---

[5] However, please also note the caveats of citation recommendation as outlined above and in Sec. 5.3.

[6] Exceptions are [174, 101], which also use the citing paper's author information besides the content.

[7] It should be noted that citation recommendation can be defined both on a citation context-level and on a document level. We here consider the task on a document level, because this enables us to have a more generic definition.

Table 2 Symbols used for formalizing *citation recommendation*, grouped by the *offline step* and the *online step*.

Symbol	Description
$D = \{d_1, ..., d_i, ..., d_n\}$	set of citing documents in the offline step
$R = \{r_1, ..., r_m, ..., r_M\}$	references of all citing documents D
$C_i = \{c_{i1}, ..., c_{ij}, ..., c_{iN}\}$	citation contexts from document d_i
$Z_i = \{z_{i1}, ..., z_{ij}, ..., z_{iN}\}$	abstract citation contexts from document d_i
Z	set of all abstract citation contexts of D
f	mapping function
g	mapping function
d	input document in the online step
R^d	references of document d
$C^d = \{c_1^d, ... c_k^d, ..., c_K^d\}$	potential citation contexts of document d
$Z^d = \{z_1^d, ..., z_k^d, ..., z_K^d\}$	abstract representations of potential citation contexts of document d
$R_{z_k^d}$	set of papers recommended for citation
d'	input document d enriched by recommended citations

Processing The processing of the input texts consists of the following steps:

(1) Reference Extraction. All references from the reference sections of all citing documents are extracted and stored in a global index R.

(2) Citation Context Extraction & Representation. First, all citation contexts $c_{ij} \in C_i$ from each citing document d_i need to be extracted. Then, these citation contexts are transformed into the desired representation form (e.g., embedding vectors, bag-of-entities, etc.) z_{ij}:

$$\forall d_i \in D \ \forall c_{ij} \in C_i : c_{ij} \rightarrow z_{ij}$$

(3) Model Learning. Given the output of the previous steps (the citing documents D, the cited documents R, and the abstract citation contexts Z), we can learn a mapping function f which maps each citation context representation z_{ij} and its citing document d_i to a reference (cited document) $r_m \in R$ as given by the training data:

$$\forall z_{ij} \in Z \ \forall d_i \in D \qquad f : (z_{ij}, d_i) \rightarrow r_m$$

Note that some approaches to citation recommendation might not use any other information from the citing documents besides the citation contexts, eliminating thus d_i as argument in the mapping function. In those cases, only the representation of the citation context z_{ij} is decisive (e.g., representation of a concept).

The mapping function f and the whole task can be formulated as a binary classification task (as also presented in [134]), especially in order to employ statistical models. Then, each citable document r_m is considered as a class and the task is to determine if (z_{ij}, d_i) should be in class r_m:

$$g(z_{ij}, d_i, r_m) \rightarrow [0, 1]$$

$[0, 1]$ is the probability of citing r_m given z_{ij} and d_i. As mentioned above, d_i might be optional for some approaches. In reality, g is often learned based on machine learning. However, one can also think of other ways to create g (e.g., rule-based approaches).

Output Output is the function g, given the abstract citation contexts Z, the citing documents D, and the cited documents R.

2.3.2 Online Step

Input Input is a text document d without citations and references (or only a few ones).

Processing Processing the document d consists of the following steps:

(1) Reference Extraction (optional): If d already contains citations and a reference section, the *references R^d* from d can be extracted and the corresponding representations can be retrieved from the database of cited papers R. These representations can be utilized for improving the citation recommendation within *Model Application*, e.g., for a better topical coherence among existing and recommended citations [91].

(2) Citation Context Extraction & Representation: First, if the existing *citations* in document d are to be used, the task is to extract and represent them in the same way as in the Offline Step. Then, all *potential citation contexts* $c_k^d \in C^d$ – i.e., contexts in d, which are judged as suitable for having a citation – are extracted from d and transformed into the same abstract representation form z_k^d as used in the *Model Learning*: $\forall c_k^d \in C^d : c_k^d \rightarrow z_k^d$. Note that, sometimes, an additional filtering step filters out all potential citation contexts which are not worth considering.

(3) Model Application: Here, the mapping function g, learned during the training, is applied on the potential citation context representations z_k^d of document d for recommending citations:

$$\forall z_k^d \in Z^d \qquad R_{z_k^d} = \{r_m \mid r_m \in R \wedge g(z_k^d, d, r_m) \geq \theta\}$$

R is thereby the global index of "citable" papers (gathered during the offline step). $R_{z_k^d}$ is the set of recommended cited papers. These papers were classified as *cited* with a likelihood of at least θ.
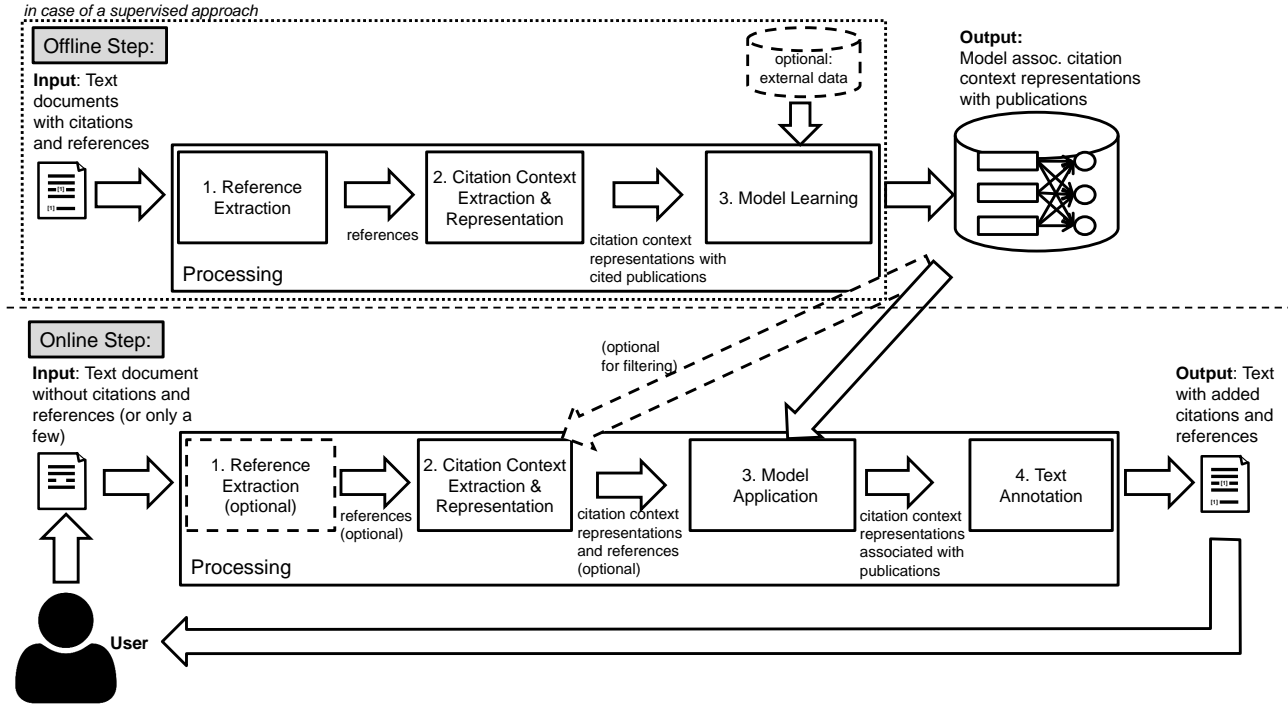
Fig. 3 Architecture of a prototypical citation recommendation system.

(4) Text Enrichment Given the document d and the set of recommendations $R_{z_k^d}$ for each citation context representation z_k^d, the running text of document d gets enriched by the recommended citations and the reference section of d gets enriched by the corresponding references.

Output Output is the annotated document d'.

2.4 Related Research Fields

2.4.1 Non-Scholarly Citation Recommendation

Also, outside academia, there is a demand for citing written knowledge. We can mention three kinds of documents, which often appear in such scenarios as citing documents: encyclopedic articles, news articles, and patents. Citation recommendation approaches developed for the scholarly field can in principle also be applied to such fields outside academia. Note, however, that each of the use cases might bring additional requirements and challenges. The scholarly domain is characterized by the use of a particular vocabulary, thus making it hard to apply models (e.g., embeddings) that were pre-trained on other domains (e.g., news). In contrast, documents in the non-scholarly field, such as news articles, often do not have a (dense) citation network. This

might make it harder to build metadata-based representations of the documents and to evaluate the recommender systems, because no co-citation network can be used for the evaluation (see the fuzzy evaluation metrics in Sec. 5.1). In the following, we outline specifically developed approaches for non-scholarly citation recommendation.

Encyclopedic articles as citing documents: The English Wikipedia is nowadays already very rich and quite complete in the number of articles included, but still lacks citations in the range of (at least) hundreds of thousands [80]. This lack of citations diminishes the potential of Wikipedia to be a reliable source of information. Since in Wikipedia mainly news articles are cited [56], several approaches have focused on developing methods for recommending news citations for Wikipedia [112, 113, 56, 55].

News articles as citing documents: Peng et al. [124] approach the task of citation recommendation for news articles. They use a combination of existing implicit and explicit citation context representations as well as 200 preselected candidate articles instead of hundreds of thousands per citation context.

Patents as citing documents: Authors of patents need to reference other patents in order to show the context in which the patent is embedded. Thus, approaches for patent citation recommendation have been proposed [109].

Table 3 Overview of tools for extracting in-text citations (i.e., references' metadata and citations' positions in the text) from scientific publications, sorted by publication year.

Tool	Approach	Input format	Output format	Extracts citation contexts (citation context length)	Extracts citing paper's abstract
CERMINE [158]	CRF	pdf	xml	yes (300 words)	yes
ParsCit [39]	CRF	txt	xml, txt	yes (200 words)	no
GROBID [104,105]	CRF	pdf	xml	no	yes
PDFX [38]	rule-based	pdf	xml	yes (300 words)	yes
Crossref pdf-extractor [40]	rule-based	pdf	xml, bib	no	no
IceCite [12]	rule-based	pdf	tsv, xml, json	no	yes
Science Parse [6]	CRF	pdf	json	yes	yes

2.4.2 Scholarly Data Recommendation

Scientists are not only confronted with an information overload regarding publications, but also regarding various other items, such as books, venues, and data sets. As a consequence, these items can also be recommended appropriately in order to assist the scientist in her work. Among others, approaches have been developed for recommending books [116], scientific events [90], venues [173] and reviewers [97] for given papers, patents [122], scientific data sets [139], potentially identical texts (by that means identifying plagiarism) [63], and newly published papers, via notifying functions [50].

2.4.3 Related Citation-based Tasks

In the following, we describe some citation-based tasks that are either strongly related to or an integral part of citation recommendation.

Citation network analysis Citation network analysis describes the task of analyzing the references between documents in order to make statements about the scientific landscape and to investigate quantitatively scientific publishing. Among others, citation network analysis has been performed to determine communities of researchers [172,43], to find experts in a domain [68], to know which researchers or publications have been or will become important, and to obtain trends in what is published over time [70]. Note that citation network analysis operates on the document level and generally does not consider the document's contents.

Citation context detection and extraction Each citation is textually embedded in a *citation context*. The citation context can vary in length, ranging typically from a part of a sentence to many sentences. As shown in several analyses [7,3], precisely determining the borders of the citation context is non-trivial. This is because several citations might appear in the same sentence and because citations can have different roles. While in some cases a claim made by the author needs to be backed up, in other cases a single concept (e.g., method, data set, or other domain-specific entity) needs to be referenced by a corresponding publication [111]. In conclusion, there seems to be no consistent single optimal citation context length [7,133,132]. Different citation context lengths have been used for citation recommendation (see Table 5).

To extract citation contexts and references from papers, specific approaches have been developed [156,157]. These approaches were developed for PDFs with a paper-typical layout. They are not only capable of extracting a paper's metadata, such as title, author information, and abstract, in a structured format, but also the references from the reference section, as well as linking the citation markers in the text to the corresponding references. Table 3 provides an overview of the existing publicly available implementations for extracting in-text citations from scientific papers. Note that we limited ourselves to implementations which were designed for scientific papers as input and which are still deployable; other PDF extraction tools are not considered by us (see [12,156,157] for an overview of further PDF-to-text tools). Furthermore, we excluded tools, such as Neural ParsCit [128], which do not output the positions of the citations in the text. Given these tools, we can observe the following: (1) All underlying approaches are a rule engine or a conditional random field. (2) Several tools (e.g., ParsCit) have the additional feature that they can extract not only the fulltext from the PDF documents, but also a citation context around the found citation markers. (3) Several tools (e.g., ParsCit) require plaintext files as input. Transforming PDF to plaintext is, however, an additional burden and leads to noise in the data. (4) The tools differ considerably in the process-

ing time needed for processing PDF files [12]. ParsCit and GROBID, which have been used most frequently by researchers, to our knowledge, are among the fastest.

Citation context characterization Citations can have different roles, i.e., citations are used for varying purposes. These reasons are also called *citation functions*. The citation function can be determined – to some degree automatically – by analyzing the citation context and by extracting features [155, 117, 154]. Similar tasks to the citation function determination are the polarity determination (i.e., if the author speaks in a positive, neutral, or negative way about the cited paper) [1, 61] and the determination of the citation importance [160, 36].

The general typical structure of publications has been studied and brought into a schema, such as the IM-RaD structure [141], standing for introduction, methods, results, and discussion. In [19], for instance, the authors find out that the average number of citations among the same sections in article texts is invariant across all considered journals, with the introduction and discussion accounting for most of the citations. Furthermore, apparently the age of cited papers varies by section, with references to older papers being found in the methods section and citations to more recent papers in the discussion. Although such insights have not been used for development of citation recommendation approaches yet, we believe that they can be beneficial for better approximating real human citing behavior.

Citation-based document summarization Citation-based document summarization is based on the idea that the citation contexts within the citing papers are written very carefully by the authors and that they reveal noteworthy aspects of the cited papers. Thus, by collecting all citation contexts and grouping them by cited papers, summaries and opinions about the cited papers can be obtained, opening the door for citation-based automatic survey generation and automatic related work section generation [2, 49, 114].

Citation matching and modeling Citation matching [123] deals with the research challenge of finding identical citations in different documents in order to build a coherent citation network, i.e., a global index of citations for a document collection.

Representing the metadata of both citing and cited papers in a structured way is essential for any citation-based task. Recently, several ontologies, such as FaBiO and CiTO [125], have been proposed for this purpose. Besides the metadata of papers, further relations and concepts can be modeled ontologically in order to facilitate transparency and advances in research [126].

3 Comparison of Citation Recommendation Approaches

Approaches to (local and global) citation recommendation have been published over the years, using diverse methods, and proposing many variations of the citation recommendation task, such as a recommendation across languages [153] or using specific metadata about the input text [134, 48]. However, no overview and comparison of these approaches has been presented in the literature so far. In the following, we give such an overview.

3.1 Corpus Creation

Following a similar procedure as in [15], we collect the papers for our comparison as follows:

1. On May 3, 2019, we searched in DBLP for papers containing "citation" and "rec*" in the title. This resulted in a set of 179 papers. We read those papers and manually classified each of them whether they present an approach to (local or global) citation recommendation or not.
2. In a further step, we also investigated all papers referenced by the so-far given relevant papers, and the ones that refer to these so-far given papers, and classify them as relevant or not.
3. To avoid missing any papers, we used Google Scholar as an academic search engine with the query keywords "citation recommendation" and "cite recommend," as well as the Google Scholar profiles from the authors of the so-far relevant papers. Based on that, we added a few more relevant papers to our corpus.[8]

Overall, 51 papers propose a novel, either global or local citation recommendation approach (see Table 4). Out of these, 17 present *local* citation recommendation approaches, that is, approaches that use a specific citation context within the input document (see Sec. 2.1 for the distinction between local and global citation recommendation). This means that only 33.3% of the approaches denoted by the corresponding authors as *citation recommendation approaches* are actually designed for using citation contexts as input and are therefore truly citation recommendation approaches (see Sec. 2.1).

Note that we consider only papers presenting approaches to citation recommendation, and not those on data analysis (e.g., citation graph analysis). We also do not consider papers presenting approaches for recommending papers that do not use any text as the basis

---

8 [134, 100] are papers which are not indexed in DBLP, but which can be found on Google Scholar or Semantic Scholar.

Table 4 Approaches to global and local citation recommendation (CR).

Reference	Venue	Local CR
McNee et al. [110]	CSCW'02	
Strohman et al. [144]	SIGIR'07	
Nallapati et al. [119]	KDD'08	
Tang et al. [151]	PAKDD'09	
He et al. [72]	WWW'10	✓
Kataria et al. [89]	AAAI'10	✓
Bethard et al. [20]	CIKM'10	
He et al. [71]	WSDM'11	✓
Lu et al. [107]	CIKM'11	
Wu et al. [167]	FSKD'12	
He et al. [69]	SPIRE'12	✓
Huang et al. [74]	CIKM'12	✓
Rokach et al. [134]	LSDS-IR'13	✓
Liu et al. [101]	AIRS'13	✓
Jiang et al. [84]	TCDL Bulletin'13	
Zarrinkalam et al. [175]	Program'13	
Duma et al. [45]	ACL'14	✓
Livne et al. [103]	SIGIR'14	✓
Tang et al. [153]	SIGIR'14	✓
Ren et al. [131]	KDD'14	
Liu et al. [99]	JCDL'14	
Liu et al. [98]	CIKM'14	
Jiang et al. [85]	Web-KR'14	
Huang et al. [75]	WCMG'15	✓
Chakraborty et al. [35]	ICDE'15	
Hsiao et al. [73]	MDM'15	
Gao et al. [60]	FSKD'15	
Lu et al. [106]	APWeb'15	
Jiang et al. [86]	CIKM'15	
Liu et al. [100]	iConf'16	
Duma et al. [47]	LREC'16	
Duma et al. [46]	D-Lib'16	
Yin et al. [174]	APWeb'17	✓
Ebesu et al. [48]	SIGIR'17	✓
Guo et al. [65]	IEEE'17	
Cai et al. [29]	AAAI'18	
Bhagavatula et al. [21]	NAACL'18	
Kobayashi et al. [91]	JCDL'18	✓
Jiang et al. [87]	JCDL'18	
Han et al. [67]	ACL'18	✓
Jiang et al. [88]	SIGIR'18	
Zhang et al. [176]	ISMIS'18	
Cai et al. [28]	IEEE TLLNS'18	
Yang et al. [171]	JIFS'18	
Dai et al. [41]	JAIHC'18	
Yang et al. [170]	IEEE Access'18	✓
Mu et al. [118]	IEEE Access'18	
Jeong et al. [81]	arXiv'19	✓
Yang et al. [169]	IEEE Access'19	
Dai et al. [42]	IEEE Access'19	
Cai et al. [30]	IEEE Access'19	

for the recommendation, but instead use other information, such as the papers' metadata.
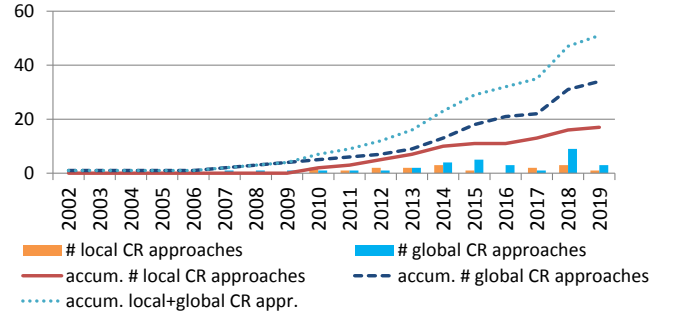
Fig. 4 Frequencies of citation recommendation (CR) approaches by publication year.

3.2 Corpus Characteristics

Table 4 lists all 51 papers on citation recommendation, together with the papers' venues and an indication of whether the described approach targets local or global citation recommendation. We can point out the following findings regarding the evolution of these approaches over time:

1. We can observe that approaches to citation recommendation have been published over the last 17 years (see Fig. 4). The task of *global* citation recommendation has attracted the interest of researchers at an earlier stage than local citation recommendation (first publication year 2002 [110] vs. 2010 [72]). Both the number of approaches to global citation recommendation and local citation recommendation has increased continuously. Overall, more approaches to global citation recommendation system have been published than approaches to local citation recommendation. However, note that the most recent publications on global citation recommendation have been published in very short time intervals at similar or same venues from partially identical authors (see Table 4).

2. Some precursor works on the general task of analyzing and predicting links between documents [37] have been published since 2000, while global citation recommendation has been targeted by researchers since 2002. Among others, there might be two major aspects that can explain the emergence of citation recommendation approaches at that time. Firstly, the number of papers published per year has increased exponentially. It became common in the 2000s to publish and to read publications online on the Web. Secondly, citations have become disproportionately more common over the years, that is, the number of citations has increased faster than the number of publications. Comparing the five-year periods 1999/2003 and 2004/2008 in [121], the number
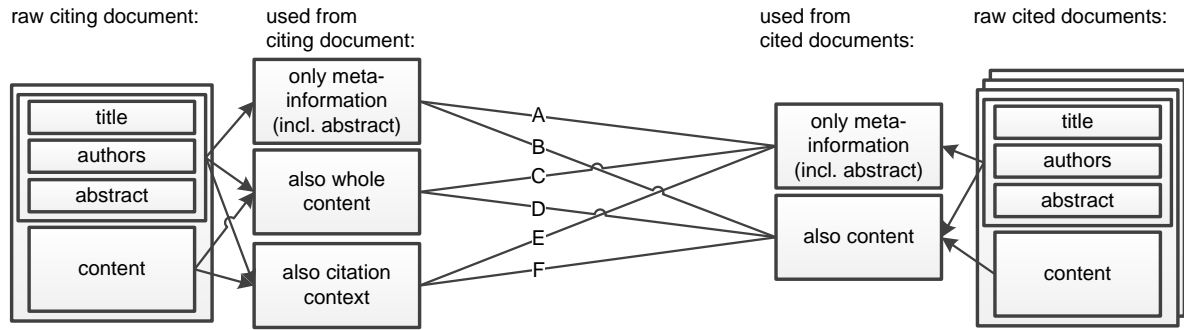
Fig. 5 Classification of citation recommendation approaches based on their set-up. The approaches are classified as follows: A: [110, 131, 85, 86, 73, 35, 60, 106, 65, 21, 87, 88, 29, 28, 171, 41, 118, 169, 30]: B: [20]; C: [175]; D: [144, 119, 151, 167, 84, 98, 100, 47, 46, 176, 42, 107, 99]; E: [72, 71, 74, 134, 101, 45, 153, 75, 174, 48, 170, 81]; F: [69, 103, 91, 67, 89] The numbers correspond to the references in the reference section.

of publications increased by 33%, while citations increased by 55%.

3. Before the content-based (local and global) citation recommendation approaches – as considered in this survey –, several systems had already been proposed that use purely the citation graph as basis for the recommendation. This "prehistory" of content-based citation recommendation is explainable by the fact that quantitative science studies such as bibliometrics have a long history, and were already quite established in the 2000s.

4. Having an appropriate and large collection of scientific papers as evaluation and training data is crucial and not easy to obtain, since – especially in the past – papers were often "hidden" behind paywalls of publishers. Therefore, it is not very surprising that several approaches [20, 86, 21, 87] consider only abstracts as citing documents instead of the papers' content. Citation recommendation then turns into reference recommendation for abstract texts.

5. Citation recommendation is located in the intersection of the research areas *information retrieval, digital libraries, natural language processing,* and *machine learning*. This is also reflected in the venues in which approaches to citation recommendation have been presented. Considering both global and local citation recommendation, SIGIR, IEEE Access, CIKM, and JCDL have been chosen most frequently as venues (5 times SIGIR, 5 times IEEE Access, 5 times CIKM, 3 times JCDL; together accounting for 35% of all papers). Particularly, IEEE Access has become popular as a venue for publishing citation recommendation approaches by a few researches in 2018 and 2019. Note that this journal's reviewing and publication process is designed to be very tight (one review round takes 7 days) and that IEEE has an article processing charge. Our paper corpus also contains a few publications from medium-ranked conferences,

such as AIRS [101]. It became apparent that these papers provide less comprehensive evaluations, but relatively high evaluation results (see the *evaluation metrics* paragraph in Section 3.3). Due to missing baselines, these results need to be taken with care.

6. Considering purely local citation recommendation, SIGIR (3 times) and ACL (2 times) occur most frequently as venue. The remaining venues occur only once.

Big picture. In Fig. 5, we present visually a "big picture" of the different settings in all citation recommendation approaches. We thereby differentiate between what data is used from the *citing* documents (either only metadata (incl. abstract), or metadata plus content, or metadata plus specific citation contexts), and what data is used from the *cited* documents (either only metadata, or metadata plus content). Note that approaches using the metadata or the content of the citing documents make up the group of global citation recommendation approaches, while approaches using specific citation contexts target local citation recommendation. Note also that approaches using only the metadata of the citing documents can be regarded as targeting both the expert setting and the non-expert setting (see Section 2.2), while the other approaches are designed primarily for the expert setting. The publications that propose the approaches sometimes do not point out in detail what data is used (e.g., whether the author information of the citing papers is also used), which makes a valid comparison infeasible. Thus, this "big picture" figure tries to provide a clear picture of what has been pursued so far. Notable, for instance, is that 23.5% (12 out of 51) of all approaches use citation contexts (less than the whole content) of the citing documents and only the metadata of the cited documents (see class E). In contrast, we can find only one approach that uses the whole content of the citing documents and only the metadata of the cited documents
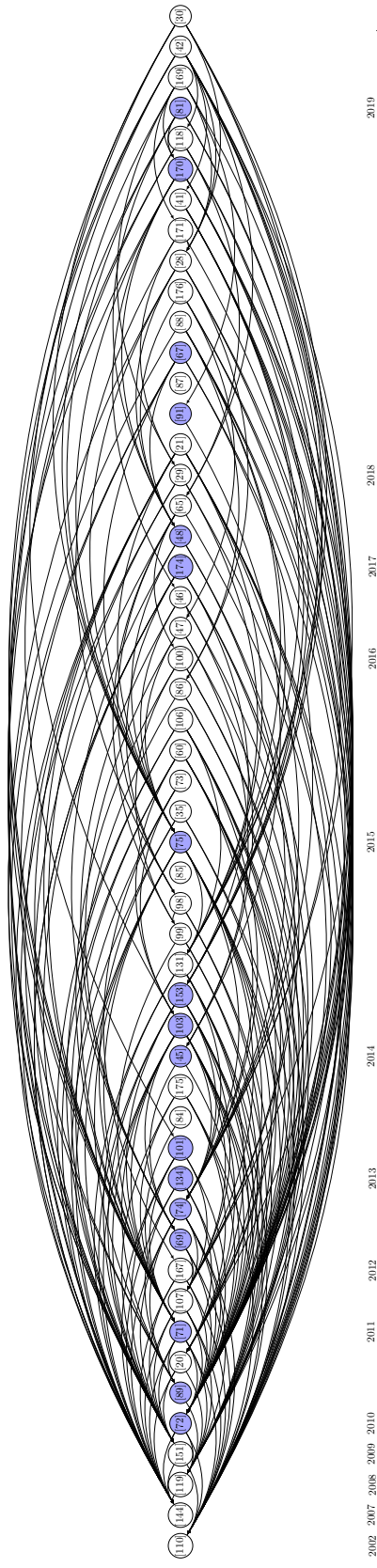
Fig. 6 Citations between papers presenting citation recommendation approaches. Local citation recommendation approaches are highlighted in blue.

(see class C). We can mention two potential reasons for this fact. Firstly, it can be difficult to obtain the publications' full texts (due to, among other reasons, limited APIs and copyright issues). Secondly, operating only with papers' metadata is also easier from a technical perspective.

Citation relationships. Fig. 6 shows the citation-relationships between papers with citation recommendation approaches. The papers are thereby ordered from left to right by publishing year. It is eye-catching that there is no continuous citing behavior along the temporal dimension, i.e., a paper in our set does not necessarily cite preceding papers in our set. However, in some cases we can explain this by the fact that publications were published within short time intervals. Consequently, the authors might not have been aware of other approaches which had either been published very recently or had not yet been published. Nevertheless, we can observe that authors of citation recommendation approaches do omit references to other citation recommendation approaches.

3.3 Comparison of Local Citation Recommendation Approaches

When comparing citation recommendation approaches, it is important to differentiate between approaches to local citation recommendation (making recommendations based on a small text fragment) and approaches to global citation recommendation. To understand that, consider a scenario in which a text document with 20 citation markers is given. In case of local citation recommendation, it is not uncommon to provide, for instance, three recommendations per citation context. However, a global citation recommendation system would provide only a list of 60 recommendations without indications where to insert the corresponding citation markers. In our mind, it is not reasonable to call this process context-aware citation recommendation and to evaluate the list of 60 recommendations in the same way as the 20 lists with 3 recommendations, since citations are meant to back up single statements and concepts on a clause level, i.e., being suitable only for specific contexts. Note also that global recommendation approaches in the context of paper recommendation are covered by existing surveys (see Introduction). This survey, in contrast, focuses on context-awareness, which, to date, has not yet been considered systematically. Thus, in this subsection, we compare only the 17 approaches to local citation recommendation.

In order to characterize and distinguish the different approaches from each other, we introduce the following dimensions:

1. What is the underlying approach and to which data mining technique is it associated?
2. What information is used for the user modeling, if any?
3. Is the set of candidate papers prefiltered before the recommendation?
4. What is used as the citation context (e.g., 1 sentence or 50 words before and after the citation marker)?
5. Is the citation context pre-specified in the evaluation or do cite-worthy contexts first need to be determined by the algorithm?
6. Is the content of the cited papers also needed (limiting the evaluation to corresponding data sets)?
7. Which evaluation data set is used (e.g., CiteSeerX or own data set)?
8. From which domain are the papers used in the evaluation (e.g., computer science)?
9. What are the used evaluation metrics?

Table 5 shows the classification of the approaches according to these dimensions. While in the following we point out the main findings per dimension, note that we also provide a description of the single approaches and their characteristics in an online semantic wiki.[9]

1. **Approach:** A variety of methods have been developed for local citation recommendation. We can group them into the following four groups:

   (a) **Hand-crafted feature based models [71, 134, 101, 45, 103].** All approaches in this group are based on features that were hand-crafted by the developers. Text similarity scores obtained between the citation context and the candidate papers are examples of text-based features. Remarkably, all features used for the approaches are kept comparably simple. Moreover, the approaches do not use additional external data sources, but rather statistics derived from the paper collection itself (e.g., citation count and text similarity). Relatively basic techniques used for the ranking of citations for the purpose of citation recommendation (e.g., logistic regression and linear SVM [101], or merely the cosine similarity of TF-IDF vectors [45]) seem to lead to already noteworthy evaluation results and, thus, can serve as strong baselines for the evaluations of other systems. Among the most complex presented methods are an ensemble of decision trees [71] and gradient boosted regression trees [103]. Note, however, that their superiority compared to simpler models is hard to judge due to differing evaluation settings, such as data sets and metrics.

In recent years, no novel approaches of this group have been published any more (latest one from 2014), likely due to the fact that (1) the obvious features have already been used and evaluated, and (2) recent approaches (e.g., neural networks) seem to outperform the hand-crafted feature-based models. Nevertheless, hand-crafted feature based models provide the following advantages: *1. Scalability:* Since both the computation of the features and the used classifier/regression model are kept rather simple, the citation recommendation approaches become very scalable and fast. *2. Explainability:* The described techniques are particularly beneficial when it comes to getting to know which features are most indicative for recommending appropriate citations. *3. Small data:* The models do not require huge data sets for training, but may already work well for small data sets (e.g., a few thousand documents). Existing approaches in this group use mainly lexical features and other bibliometrics-based features (e.g., citation count). Hand-crafted features focusing on the semantics and pragmatics of the citation contexts and of the candidate cited documents, are missing. In the future, one can envision a scenario in which claims or argumentation structures are extracted from the citation contexts and compared with the claims/argumentation structures from the citable documents.

   (b) **Topic modeling [72, 89].** Topic modeling is a way of representing text (here: candidate papers and citation contexts) by means of abstract topics, and thereby exploiting the latent semantic structure of texts. Topic modeling became popular, among others, after the publication of the LDA approach by Blei et al. in 2002 and was applied to local citation recommendation in 2010 [72, 89]. Using topic modeling in the context of citation recommendation means to adapt default topic modeling approaches, which work purely on plain text documents, in such a way that they can deal with both texts and citations. To this end, He et al. [72] use a probabilistic model based on Gleason's Theorem, while Kataria et al. [89] propose the LDA-variations Link-LDA and Link-PLSA-LDA.

   Note that topic modeling per se is computationally rather expensive and may require more resources than approaches of the group (a). Moreover, conceptually it might be designed rather for longer texts, and, thus, more suitable for global citation recommendation (where it has been applied in [151, 119]). In the series of citation rec-

---

Table 5 Overview of local scientific citation recommendation approaches, listed chronologically by publication date (considering year and month).

Paper	Year	Group	Approach	User model	Pre-filter length	Citation context length	Citation place-holders needed	Cited papers' content needed	Evaluation data set	Domain	Evaluation metrics
[72]	2010	b	Probabilistic model (Gleason's Theorem)	–	–	50 words before and after	yes	no	CiteSeerX	Computer science	recall, co-cited prob., nDCG, runtime
[89]	2010	b	Topic model (adapt. LDA)	–	–	30 words before and after	yes	yes	CiteSeerX	Computer science	RKL
[71]	2011	a	Ensemble of decision trees	–	–	50 words before and after	no	no	CiteSeerX	Computer science	precision, co-cited probability, nDCG
[69]	2012	c	Machine translation	–	–	1 sentence	yes	yes	CiteSeer & CiteULike	Computer science	F1, precision, recall, F1; Bpref, MRR
[74]	2012	c	Machine translation	–	–	1-3 sentences	yes	no	CiteSeer & CiteULike	Computer science	MAP
[134]	2013	a	Ensemble of supervised ML techniques	author	top 500	50 words before and after	yes	no	Own dataset	Computer science	recall, MAP
[101]	2013	a	SVM	author	–	on average 13.4 words	yes	no	Own dataset	Computer science	recall, MAP
[45]	2014	a	cos similarity of vectors (TF-IDF based)	–	–	5-30 words before and after	yes	depending on variant	Part of ACL Anthology	Comput. linguistics	accuracy
[103]	2014	a	Regression trees (gradient boosted)	author	top 500	50 words before and after	no	yes	Own dataset	Computer science	nDCG
[153]	2014	d	Learning-to-rank	–	–	sentence plus sentence before and after	yes	no	Own dataset	Computer science	recall, MAP, MRR
[75]	2015	d	Neural network (feed-forward)	–	variable	sentence plus sentence before and after	yes	no	CiteSeer	Computer science	MAP, MRR, nDCG
[174]	2017	d	Neural network (CNN)	–	variable	sentence plus sentence before and after	yes	no (but title + abstract)	own (same as in [101])	Computer science and technology	MAP, recall
[48]	2017	d	Neural network (CNN + RNN)	author	top 2048	50 words before and after	yes	no	RefSeer	Computer science	Recall, nDCG, MAP, MRR
[91]	2018	d	cos. similarity of paper embeddings	–	–	1 sentence	yes	yes	Own dataset (from ACM library)	Computer science	nDCG
[67]	2018	d	Dot product of 2 paper embeddings	–	–	50 words before and after	yes	yes	NIPS, ACL-ANT, CiteSeer + DBLP	Computer science	recall, MAP, MRR, nDCG
[170]	2018	d	Neural network (LSTM)	author, venue	–	5 sentences before and after	yes	yes	AAN + DBLP	Computer science	recall, MAP, MRR
[81]	2019	d	Neural network (feed-forward)	–	–	50 words before and after	yes	no	AAN + own dataset	Computer science	MAP, MRR, recall

ommendation approaches, topic modeling has been applied within a relatively short time interval (2010 only for local citation recommendation; 2008 and 2009 in case of global citation recommendation) and has been replaced first by machine translation models (group (c)) and later by neural network-based approaches (group (d)).

(c) **Machine translation [69, 74].** The authors of [69, 74] apply the concept of machine translation to local citation recommendation. These approaches had been published also within a short time frame, namely only in 2012. Using machine translation might appear surprising at first. However, the developed approaches do not translate words from one language into another, but merely "translate" the citation context into the cited document (written in the same language, but maybe with a partially different vocabulary). In this way, the vocabulary mismatch problem is avoided. The first published approach using machine translation for citation recommendation was designed for global citation recommendation [107]. Here, the words in the citing document are translated to the words in the cited document. This requires the cited documents' content to be available. Approaches to local citation recommendation follow: In [69], the translation model uses several positions in the citable document for translations. However, this makes the approach computationally very expensive. The last approach in this group [74] translates the citing document merely into the identifiers of the cited documents and does not use the cited documents' content any more. By doing that, the authors obtain surprisingly high evaluation results. Note that machine translation is a statistical approach and requires a large training data set. However, in the published papers and their evaluations, rather small data sets (e.g., 3,000 and 14,000 documents in [74] and 30,000 documents in [69]) are used. Moreover, high thresholds for the translation probability may be set to make the machine translation approach feasible [74].

(d) **Neural networks [153, 75, 174, 48, 91, 67].** This group contains not only many approaches to local citation recommendation (6 out of 17, that is, 35%), but also the most recent ones: here, papers have been published since 2014. Due to the large field of neural network research in general, the architectures proposed here also vary considerably. Although there are also relatively generic neural network architectures applied [174, 153],

we can observe a tendency in increasing complexity of the approaches. Approaches are either specifically designed for texts with citations (e.g., [48, 91]) or consider texts with citations as a special case of hyperlinked documents [67]. In the first subgroup are approaches using convolutional neural networks [48] and special attention mechanisms, such as for authors [48]. In the latter subgroup is an approach which uses two vector representations for each paper. Note that the approaches in this approach group do not incorporate any user model information, but work purely on the sequence of words. An exception is [48] which exploits the citing document's author information.

When it comes to deciding whether neural networks should be used in a productive system, one should note that neural networks need to be trained on large data sets. In recent years, large paper collections have been published (see Sec. 4). However, also the infrastructure, such as GPUs, needs to be available. Moreover, considerable approximations need to be applied to keep the approach feasible. This includes the negative sampling strategy [174, 75, 91, 67]. But also a pre-filtering step before the actual citation recommendation approach is often performed, which reduces the set of candidate papers significantly [75].

Han et al. [67], who propose one of the most recent citation recommendation systems and who evaluate their approach on data sets with real-world sizes, report recall@10 values of 0.16/ 0.32/ 0.21 and nDCG@10 values of 0.08/0.21/0.13 for the data sets NIPS, ACL-Anthology, and Cite-Seer+DBLP data. This shows that the results depend considerably on the data set and on the pre-processing steps (e.g., whether PDF-to-text conversion is performed). Overall, it can be assumed that the novel approaches to citation recommendation published in the near future will mainly be based on neural networks, too.

Overall, existing approaches are primarily based on implicit representations of the cited statements and concepts (e.g., embeddings of citation contexts [91, 67]), but not on fine-grained explicit representations of statements or events. One reason for that might be the missing research on the different citation types besides the citation function, and the current relatively low performance of fact extraction and event extraction methods from text.

2. **User model:** As outlined in Sec. 2, approaches to citation recommendation can optionally incorpo-

rate user information, such as the user name, the venue that the input text should be submitted to, or keywords which categorize the input text explicitly. Overall, we can observe that most approaches (12 out of 17, i.e., 71%) do not use any user model. Five approaches are dependent on the author name of the citing document.[10]

3. **Prefilter:** By default, all candidate papers need to be taken into account for any citation recommendation. This often results in millions of comparisons between representation forms and, thus, turns out to be unfeasible. To escape from that, the proposed methods often incorporate a pre-filtering step as a step before the actual recommendation, in which the set of candidate papers is drastically reduced. For instance, before applying a neural network-based approach for a precise citation recommendation, the top 2048 most relevant papers are retrieved from the paper collection via BM25 [48]. In 30% (5 out the 17) of the considered papers, the authors mention such a step (see Table 5). While three authors implement a certain numerical value as threshold [134, 103, 48],[11] others use flexible thresholds such as the word probabilities [75, 174].[12]

4. **Citation context length:** The size of the citation context varies from approach to approach. Typically, 1–3 sentences [69, 74, 153, 75, 174, 91] or a window of up to 50 words [72, 89, 71, 134, 45, 103, 48, 67, 81] is used. Investigations on the citation context length suggest that there is no one ideal citation context length [7].

5. **Citation placeholders:** The citation placeholders, i.e., the places in which a citation should be recommended, and therefore also the citation context, are typically already provided a priori for evaluating the single approaches (exceptions are [71, 103]). The main reason for this fact is presumably that the past approaches focus on the citation recommendation task itself and see the identification of "cite-worthy" contexts as a separate task. Determining the cite-worthiness, which is similar to determining the citation function, is not tackled in the approaches. However, there have been separate attempts at solving this task [54, 148] (and related: [3]). Also, with respect to performing the evaluation, having a flexible citation context makes it very tricky to compare the approaches in offline eval-

uations with the citation contexts and their citations from the ground truth. Single attempts such as [71, 103], solve it, however, for instance, by using only those citation contexts and associated citations which overlap with the found citation contexts to a considerable degree.

6. **Cited papers' content needed:** The approaches to citation recommendation differ in the characteristic of whether they incorporate the content of the *cited* documents or not. Incorporating the contents means that all cited documents need to be available in the form of full text. This is often a limitation, since any paper published somewhere could be referenced by authors; the cited documents are, thus, often not in any ready collection of citing documents. For instance, in the CiteSeer data set of [119], only 16% of the cited documents are also citing documents; this is similar to the arXiv CS data set [52] and unarXiv data set [136]. Not incorporating the content, on the other hand, leads to a less fine-grained recommendation and the vision of even a single fact-based recommendation is illusive. Considering the approaches to local citation recommendation, we cannot recognize a clear trend concerning the aspect of used content: both approaches using the cited papers' content and not using it have been proposed in recent years.

7. **Evaluation data set:** In general, a variety of data sets have been used in the publications. Most frequently (in 8 out of 17, i.e., 47% of the cases), versions of the CiteSeer data set (i.e., CiteSeer, CiteSeerX, RefSeer) have been applied, because this data set has been available since the early years of citation recommendation research and because it is relatively large. However, even the approaches in recent years are often evaluated on newly created data sets. As Sec. 4.1 is dedicated to data sets used for citation recommendation, we can refer to this section for more details.

8. **Domain:** Independent of which data set has been applied, all data sets cover the computer science or computational linguistics domain. We can assume that this is because (1) the papers in those domains are relatively easy to obtain online, and because (2) the papers are understandable by the authors of these approaches, allowing them to judge at first sight whether the recommendations are suitable.

9. **Evaluation metrics:** Concerning the usage of evaluation metrics and the interpretation of evaluation scores, the following aspects are especially noteworthy:

(a) *Varying metrics:* The metrics used across the papers vary considerably; most frequently, recall,

---

[10] The two global citation recommendation approaches [99, 21] allow the user to disclose more information about her optionally.

[11] Examples in the case of global citation recommendation are [144, 21].

[12] Concerning global citation recommendation, we can refer here to [99, 86].

MAP, nDCG, and MRR are used (10/9/7/7 out of 17 times). This variety makes it hard to compare the effectiveness of the approaches.

(b) *Varying data sets:* Since largely systems have been evaluated on varying data sets and with varying document filtering criteria, we can hardly compare the systems' performance overall. For instance, the recent approaches [75,48] report both nDCG@10 scores of 0.26.[13]

(c) *Varying k:* Even if the same metrics are used in different papers, and maybe when even the same data sets are used, for considering the top k returned recommendations, different k values are considered, with a great variance from $k = 1$ up to $k = 200$. Especially high values like $k = 100$ [91] or $k = 80$ [81] seem to be unrealistic as no user-friendly system would presumably expect the user to check so many recommendations.

(d) *Missing baselines:* It can be observed that the considered papers do not reference all prior works (see also Fig. 6) and that previously proposed approaches are not used sufficiently as baselines in the evaluations, although the papers propose solutions for the same research problem. This applies to papers on local citation recommendation and global citation recommendation. For instance, [87] does not cite [153], although both tackle the cross-language citation recommendation problem. This issue was already observed for papers on paper recommendation in [15].

(e) *Varying citation recommendation tasks:* The system's performance strongly depends on the kind of citation recommendation which is pursued. Given not only a citation context as input, but also the metadata of the citing paper, such as the authors, the venue, etc., then the nDCG@10 score can be 0.62 as in [103] instead of around 0.26 as in [75,48].[14]

In total, it is very hard to compare the effectiveness of the approaches (1) if different metrics are used and with different top k values, (2) if different evaluation data sets are used, (3) if the approaches do not use existing systems as baselines, and (4) if the differences in the task set-up are not outlined. Considering the above-discussed approaches, we can observe this phenomenon to a high degree.

---

[13] In case of global citation recommendation, see [100] with an nDCG@10 score of 0.21.

[14] Moreover, global citation recommendation systems using only the papers' abstracts perform differently to the ones based on the papers' full text. This can be illustrated by the fact that Liu et al. [100] use an abstract as input and obtain MAP@all of 0.16, while the same authors in [99] obtain a MAP@all score of 0.64 when using the full text.

3.4 System Demonstrations

While a relatively large amount of approaches to citation recommendation have been published, only *RefSeer* [76] and *CITEWERTs*[15] [53] have been presented as systems for demonstration purposes. *RefSeer* is based on the model proposed by He et al. [72] and uses CiteULike as the underlying document corpus. It recommends one citation for each sentence in the input text. *CITEWERTs*, in contrast, is the first system which not only recommends citations but also identifies cite-worthy contexts in the input text beforehand. This makes the system more user-friendly, since it hides unnecessary recommendations, and it reduces the number of costly recommendation computations. Besides these systems, to the best of our knowledge, only paper recommendation systems exist, i.e., systems that do not use any citation context, but, for instance, only use a citation graph [77]. *TheAdvisor* [93], *FairScholar* [9] are further examples of paper recommender system demonstrations. *Google Scholar*,[16] *Mendeley*[17], *Docear* [16], and *Mr. DLib* [17] also provide a functionality for obtaining paper recommendations.

4 Data Sets for Citation Recommendation

In this section, we give an overview of data sets which can be used in the context of citation recommendation. Section 4.1 presents data sets containing papers' content, while Section 4.2 outlines data sets containing purely metadata about papers.

4.1 Corpora Containing Papers' Content

4.1.1 Overview of Data Sets

There exist several corpora which provide papers' content and, hence, can serve as a gold standard for automatic evaluations. Table 6 gives an overview of the data sets which are considered by us. Note that we only consider data sets here that are not outdated and that are still available (either online or upon request from the author). Hence, old data sets, such as the Rexa data base [144] or the initial CiteSeer database [62], are not included.[18]

---

[15] http://citewerts.citation-recommendation.org/

[16] http://scholar.google.com/

[17] http://mendeley.com/

[18] *CiteULike* (http://www.citeulike.org/), a popular data set for paper recommendation, is not included in our list, since the full text of the papers is not available.

Generally, we can differentiate between two corpora sets: firstly, the CiteSeer data sets, available in different versions, have been explicitly created for citation-based tasks. They already provide the citation contexts of each citing paper and can be described as follows:

- **CiteSeerX (complete) [32]**: Referring to the CiteSeerX version of 2014, the number of indexed documents exceeded 2M. The CiteSeerX system crawls, indexes, and parses documents that are openly available on the Web. Therefore, only about half of all indexed documents are actually scientific publications, while a large fraction of the documents are manuscripts. The degree to which the findings resulting from the evaluations based on CiteSeerX also hold for the actual citing behavior in science is therefore unknown to some degree.
- **CiteSeerX cleaned by Caragea et al. [32]**: The raw CiteSeerX data set contains a lot of noise and errors as outlined by Roy et al. [135]. Thus, in 2014, Caragea et al. [32] released a smaller, cleaner version of it. The revised data set resolves some of the noise problems and in addition links papers to DBLP.
- **RefSeer [75]**: RefSeer has been used for evaluating several citation recommendation approaches [75, 48]. Since it contains the data of CiteSeerX as of October 2013 without further data quality improvement efforts, RefSeer is on the same quality level as CiteSeerX.
- **CiteSeerX cleaned by Wu et al. [168]**: According to Wu et al. [168], the cleaned data set [32] still has relatively low precision in terms of matching CiteSeerX papers with papers in DBLP. Hence, Wu et al. have published another approach for creating a cleaner data set out of the raw CiteSeerX data, achieving slightly better results on the matching of the papers from CiteSeerX and DBLP.

Then, there are collections of scientific publications, with and without provided metadata, for which citation contexts are not explicitly provided. However, in those cases, the citation contexts can be extracted by appropriate tools based on the papers' content, making these corpora also applicable as ground truth for offline evaluations. They are listed alphabetically in the following:

- **ACL Anthology Network (ACL-AAN) [129]**: ACL-AAN is a manually curated database of citations, collaborations, and summaries in the field of Computational Linguistics. It is based on 18k papers. The latest release is from 2016. ACL-AAN has been used as an evaluation data set for many tasks.
- **ACL Anthology Reference Corpus (ACL-ARC) [22]**:[19] ACL-ARC is a widely used corpus of schol-

arly publications about computational linguistics. There are different versions of it available. ACL-ARC is based on the ACL Anthology website and contains the source PDF files (about 11k for the February 2007 snapshot), the corresponding content as plaintext, and metadata of the documents taken either from the website or from the PDFs.
- **arXiv CS [52]**: This data set, used by [54,51], was obtained by utilizing all arXiv.org source data of the computer science domain and transforming the LATEX files into plaintext by an own implemented TEXparser. As far as possible, each reference is linked to its DBLP entry.
- **CORE**:[20] CORE collects openly available scientific publications (originating from institutional repositories, subject repositories, and journal publishers) as data basis for approaches concerning search, text mining, and analytics. As of October 2019, the data set contains 136M open access articles. CORE has been proposed for citation-based tasks for several years. However, to the best of our knowledge, it has not yet been used for evaluating or deploying any of the published citation recommendation systems.
- **Scholarly Paper Recommendation Dataset 2 (Scholarly Data Set)**:[21] This data set contains about 100k publications of the ACM Digital Library and has been used for evaluating paper recommendation approaches [147,146].
- **unarXiv [136]**: This data set is an extension of the *arXiv CS* data set. It consists of over one million full text documents (about 269 million sentences) and links to 2.7 million unique papers via 29.2 million citation contexts (having 15.9 million unique references). All papers and citations are linked to the Microsoft Academic Graph.

4.1.2 Comparison of Evaluation Data Sets

Table 6 shows the mentioned data sets categorized by different dimensions. We can outline the following highlights with respect to these dimensions:

Size of data set The considered data sets differ considerably in their sizes: they range from small (below 100k documents; see ACL-ARC and ACL-AAN) to very large (over 1M documents; see CiteSeerX complete). Note thereby that the cleanliness of the provided papers' contents does not necessarily depend on the overall size of the data set: for instance, ACL-AAN and ACL-ARC are quite noisy, as they contain rather old

[19] http://acl-arc.comp.nus.edu.sg/.
[20] http://core.ac.uk/.
[21] http://www.comp.nus.edu.sg/~sugiyama/SchPaperRecData.html.

Table 6 Overview of data sets applicable to citation recommendation.

	Size of data set	Citation context available, size	Metadata of citing paper (structured)	Metadata of cited paper (structured)	Full text of all citing papers	Full text of all cited papers	Abstract of citing paper	Abstract of cited paper	Full citation graph	Cleanliness	Links	Usage
CiteSeerX complete	very large	yes, 400 chars	yes (noisy)	yes (noisy)	yes	no	yes	not all	no (but large)	no	no	[119][151][72][89][71][74][134][75]
CiteSeerX cleaned by Caragea et al.	large	yes, 400 chars	yes (noisy)	yes (noisy)	no	no	yes	not all	no (but large)	no	DBLP	
RefSeer	large	yes, 400 chars	yes (noisy)	yes (noisy)	no	no	yes	not all	no	no	no	[48]
CiteSeerX cleaned by Wang et al.	large	yes, 400 chars	yes (noisy)	yes (noisy)	no	no	yes	not all	no (but large)	no	DBLP	
ACL-AAN	small	no (extractable)	yes	no (extractable)	yes (noisy)	no	(extractable)	not all	no	no	no	[45][67][170][81]
ACL-ARC	small	no (extractable)	yes	no (extractable)	yes (noisy)	no	(extractable)	not all	no	no	no	[20]
arXiv CS	medium	yes, 1 sentence	yes	yes	yes	no	(extractable)	not all	no	yes	DBLP	
CORE	very large	no (part. extractable)	yes	no	partially	no	yes	not all	no (but large)	yes	no	
Scholarly Dataset 2	medium	no (extractable)	no (extractable)	no (extractable)	yes	no	(extractable)	not all	no	yes	DBLP	
unarXiv	large	yes, 3 sentences	yes	yes	yes	no	(extractable)	not all	no	yes	MAG	

Table 7 Overview of papers' metadata data sets applicable to citation recommendation.

	Size of data set	Abstract of citing paper	Abstract of cited paper	Full citation graph	Cleanliness	Links
AMiner DBLPv10	large	partially	partially	yes	yes	DBLP
AMiner ACMv9	large	yes	yes	yes	yes	DBLP (but no URIs)
Microsoft Academic Graph	very large	no	no	yes	yes	no
Open Academic Graph	very large	yes	yes	yes (open access papers)	yes	DBLP (but no URIs)
PubMed	large	no	partially	yes	yes	no

publications, which are hard to parse. However, clean metadata of the cited papers is available for those data sets.

Availability of citation context CiteSeerX, arXiv CS, and the unarXiv data set provide explicitly extracted citation contexts of the citations in the documents. In case of the different versions of CiteSeerX, a fixed window of 400 characters has been chosen around the citation markers. In the case of arXiv CS and unarXiv, the content is provided sentence-wise, so that all sentences annotated with citation identifiers can be used as citation context. The corpora which contain the publications contents in their original form – namely, ACL-AAN, ACL-ARC, CORE, and Scholarly – do not provide citation contexts. However, these contexts could be extracted without much effort by using appropriate tools from the source PDF files.

Structured metadata of citing papers For all the presented corpora, structured metadata of all the citing papers is provided. An exception is Scholarly, which only consists of PDF files. Hence, the metadata needs to be extracted by oneself with the corresponding tools. Note that the metadata is clean only for those corpora for which the information has been entered manually at some point. For CiteSeerX, all information, including the metadata of citing papers, has been extracted from the publications (mainly PDFs). Hence, this framework is independent of external data. However, as a tradeoff, the extracted metadata is to some extent noisy and inaccurate (missing information or wrongly split strings etc.) [135].

Structured metadata of cited papers Only the CiteSeer data sets as well as arXiv CS and unarXiv provide this information per se. In the case of CORE, it is planned that publications will be linked to the Microsoft Academic Graph. Consequently, structured metadata of cited papers will be retrievable from this data set.[22] For the other corpora containing publications' content, the metadata of the cited papers can be obtained by extracting the information from the publications' reference sections via the appropriate tools. However, note that it does not only require the parsing via an appropriate information extraction tool, but also the reconciliation of the data (i.e., building a global database of publications' metadata). The task of how to find out if two referenced papers are actually the same and, hence, should have the same identifier is non-trivial and is known as *citation matching*.

Paper content of citing papers Some approaches, such as sequence-to-sequence approaches, require the complete contents of all citing papers. In the complete CiteSeerX data set, all citing papers' contents are still available. Also the paper collections Scholarly, arXiv CS, unarXiv, ACL-ARC, and ACL-AAN (and CORE to some degree) contain the papers' full texts. However, in case of Scholarly and ACL-AAN, the original data sets do not contain the contents as plaintext, so that one first needs to run appropriate transformation approaches.

Paper content of cited papers All considered data sets do not provide the full texts of all cited papers. This is not surprising, as papers typically cite papers without any restrictions and, thus, from various publishers.

Abstract of citing papers Since the abstract of papers belongs to the metadata, it is quite easily obtainable for both citing papers and cited papers. Furthermore, it already summarizes the main points of each paper (although typically not sufficiently for a detailed and precise recommendation) and can be used for obtaining a better representation of the paper, and, hence, for improving the recommendation of papers based on citation contexts overall. Regarding the citing papers, all data sets either provide the abstract already in an explicitly given form (see the CiteSeerX data set and partially CORE) or contain the original publications (as PDF or similar formats), so that the abstract can be extracted from them (see Scholarly, arXiv CS, unarXiv, ACL-ARC, ACL-AAN).

Abstract of cited papers Having as much information as possible about what the cited papers are dealing with is crucial for a good citation recommendation. In this context, the abstracts of cited papers are very useful and are used by several approaches [72, 71, 103, 86, 107, 45, 174, 21, 87]. However, none of the data sets contain abstracts for all cited papers.

Full citation graph In a full citation graph (also called *citation network*), not only the citations of the citing papers are represented, but the citations of any paper of a given document collection. Such a graph can be used for obtaining a good representation of the papers (see paper embeddings [48, 63]) and to compute similarities among papers. None of the considered corpora provides such an extended citation graph.[23] As an alternative, one can think of linking papers from one corpus with papers of a metadata corpus (see Section 4.2).

---

[22] As of November 4, 2019, the webpage mentions links to the Microsoft Academic Graph. However, no corresponding information can be found in the data set.

[23] Note, however, that data sets such as unarXiv and CORE link to the Microsoft Academic Graph providing citation information.

Cleanliness The situation is mixed in this regard: the metadata of the papers is of good quality, especially if it originates from corresponding, dedicated databases instead of being extracted solely from the publications themselves (see ACL-AAN, ACL-ARC, arXiv CS, and unarXiv vs. the CiteSeerX data sets). The papers' content is typically provided via information extraction methods, meaning that the quality is not that high, particularly if the papers were hard to parse and process, e.g., due to being very old (see the papers of ACL-ARC and ACL-AAN vs. Scholarly, which contains newer papers) or due to special formating in the publications, such as formulas in the text (see CiteSeerX data sets vs. the arxiv CS and unarXiv data sets, where formulas were detected and removed).

Links to bibliographic data sets Having publications linked to external bibliographic data sets such as DBLP allows the use of interlinked information for paper representations and for search. Corpora of scientific papers have often been created in the area of computer science, since there are many publications available online. As a consequence, the most widely used bibliographic database for computer science, DBLP, has been used as a reference of interlinking. More precisely, the cleaned versions of CiteSeerX and the arXiv CS data set provide links to DBLP. unarXiv provides links to the Microsoft Academic Graph, as it covers not only computer science papers, but also many other disciplines.

4.2 Corpora Containing Papers' Metadata

Besides corpora including papers' content, data sets exist that contain metadata about publications; typical metadata include the citation relations between papers and the titles, venues, publication years, and abstracts of the publications. Although no content is usually provided, the metadata can be regarded as an explicit, structured representation of the papers and, hence, can be used as a valuable representation of the papers, e.g., for learning embedding vectors based of them (see, e.g., [48,59]). Due to their extensive sizes, the following data sets are in our view particularly suitable for citation recommendation:[24]

- **AMiner DBLPv10**[25] [152]: This data set contains over 3M papers and 25.2M citation relationships, making it a large citation network data set. Since DBLP was used as data source, the data is very clean.
- **AMiner ACMv9**[26] [152]: This data set has the same structure as AMiner DBLPv10, but was constructed from 2.4M ACM publications, with 9.7M citations.
- **Microsoft Academic Graph:**[27] This data set can be considered as an actual knowledge graph about publications and associated entities such as authors, institutions, journals, and fields of study. Direct access to the MAG is only provided via an API. However, dump versions have been created.[28] Prior versions of the MAG are known as the *Microsoft Academic Search data set*, based on a the project *Microsoft Academic Search* which retired in 2012.
- **Open Academic Graph:**[29] This data set is designated to be an intersection of the Microsoft Academic Graph and the AMiner data. In many cases, the DBLP entries for computer science publications ought to be retrievable.
- **PubMed:**[30] PubMed is a database of bibliographic information with a focus on life science literature. As of October 2019, it contains 29M citations and abstracts. It also provides links to the full-text articles and third-party websites if available (but no content).

Table 7 shows the mentioned data sets categorized by various dimensions. The same dimensions are used as for comparing the corpora in Section 4.1, except the ones which are homogeneous among the metadata data sets (e.g. *availability of citation context, paper content of citing papers*). Due to page limitations, we omit a textual comparison of the mentioned metadata data sets.

5 Evaluation Methods and Challenges

In this section, we first discuss the different ways of evaluating citation recommendation approaches. Secondly,

---

[24] The data set *Mendeley DataTEL* is not listed, as it has not been available to us after several requests. Further data sets, such as CORA (`https://relational.fit.cvut.cz/dataset/CORA`), have not been shortlisted due to their small size. We have also not listed bibliographic databases like DBLP here, as they contain neither the papers' contents nor information about the citations between papers. Also Springer's SciGraph does not contain any citation information yet. Bibliographic databases, such as *Scopus* and *Web*

of Science, are dedicated information retrieval platforms, but do not officially support bulk downloads.

[25] `https://aminer.org/citation`.

[26] `https://aminer.org/citation`.

[27] `https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/`.

[28] `https://kddcup2016.azurewebsites.net/` and `http://ma-graph.org/`.

[29] `https://www.openacademic.ai/oag/`.

[30] `https://www.nlm.nih.gov/databases/download/pubmed_medline.html`.

we point out important challenges related to evaluating citation recommendation approaches. Afterwards, we provide the reader with guidelines concerning what aspects to consider for evaluating future recommender systems.

5.1 Evaluation Methods for Citation Recommendation

Generally, we can distinguish between offline evaluations, online evaluations, and user studies. In offline evaluations, no users are involved and the evaluation is performed automatically. Online evaluations measure the acceptance rates of recommendations in deployed recommender systems. User studies are used for measuring the user satisfaction through explicit user ratings.

For *offline evaluations*, the following evaluation methods have been applied so far for citation recommendation:

1. **Strict "citation re-prediction:"** This evaluation method has been used by almost all approaches to local citation recommendation (15 out of 17; see [72, 89, 69, 74, 134, 101, 45, 153, 75, 174, 48, 91, 67, 170, 81]). The evaluation is performed as follows: an approach is evaluated by assessing which of the citations that have been recommended by the system are also in the original publications. We can therefore call this method "re-prediction." This evaluation method scales very well, but ignores several evaluation challenges, such as the relevance of alternative citations, and the cite-worthiness of contexts (see Section 5.2). Hence, the evaluation metrics used for strict citation re-prediction, such as normalized discounted cumulative gain (nDCG), mean average precision (MAP) and mean reciprocal rank (MRR), might reflect the reality in the sense of the citing behavior observed in the past, but not the desired citing behavior.

2. **"Relaxed citation re-prediction:"** In order to allow papers to be recommended which are not written as citations by the authors of the papers, but which are still relevant, and on the other hand, to keep the evaluation still automatic and scalable, sometimes a relaxation of the strict re-prediction method is applied. In the set of considered approaches, the following methods have been applied by both He et al. [71] and Livne et al. [103]:

   (a) The *relative co-cited probability* metric is designed as a modified accuracy metric and based on the assumption that papers which are frequently co-cited are relevant to each other. Hence, if not the

actual cited paper, but a co-cited paper[31] is recommended, this paper is also considered as a hit to some degree. The relative co-cited probability is the ratio to which recommended papers are either directly cited or are co-citations of actual citations. In the latter case, the co-cited paper is only scored gradually.

   (b) The regular nDCG score is used for measuring the correct ranking of items. Modifying this score is based on the idea that if the actual paper is not standing on the intended position, but there is another paper there, which is also relevant (here, again determined by the co-citations), then this should also be judged as correct to some degree. More specifically, the authors use the average relative co-cited probability of r with all original citations of d to obtain a citation relevance score of r to d. Then the documents in D are sorted with respect to this relevance score and each document is assigned a score on a 5 point scale regarding its relevance. Finally, the average nDCG score over all documents is calculated based on these scores.

A more comprehensive, but not very scalable way to evaluate approaches is to rely on *online evaluations* [18]. None of the considered approaches has been evaluated in this way so far. Also no *user studies* for citation recommendation systems are known to us.[32]

5.2 Challenges of Evaluating Citation Recommendation Approaches

In the previous subsection we learned that it is hard to apply traditional evaluation metrics for offline evaluations of citation recommendation systems. We now point out further challenges when it comes to determining the performance of citation recommendation systems. In Section 5.3, we then propose steps for approaching some of these challenges.

5.2.1 Fitness of Citations

Training and evaluating a citation recommendation system based on an existing paper collection used as ground truth is tricky, since the citing behavior encoded in the citations of these considered papers is taken as ground truth. This becomes a problem when the original citing behavior is not favorable and adaptations are desired.

---

[31] B is a co-cited paper of A, if both A and B are cited by a third paper C.

[32] For paper recommendation, a few manual evaluations exist [15]. However, paper recommendation is out of our scope.

In the past, several analyses of scientific citing behavior have been published [150]. We can reuse these for characterizing the different aspects of citing biases in the context of evaluating citation recommendation. We thereby group citing biases along the attributes of the *citable* publications:

1. **Content Understanding:** Authors of citing papers may differ in their expertise, knowledge level, and working style when selecting citations (cf. professor vs. masters student). The suitability of the content of citable papers is therefore often judged differently.

   Furthermore, authors of citing papers might perform literature investigations and reviews in a rather sloppy way [79] and read, for instance, mainly titles and abstracts of documents only. However, titles and abstracts may deceive users about the true claims and contributions of papers. Moreover, the selection of citations can be biased by the style of the titles and abstracts (see, e.g., [27, 145]). Also the writing style of the fulltext of the citable papers has some influence on citing, as it reflects the perceived quality of the paper [102].

2. **Authors:** It is quite common to cite publications written by oneself, called self-citations, [5, 78] or written by colleagues, advisors, and friends [161], with an element of preferential bias. Although analyses have shown that this is not per se harmful [150], a citation recommender system ought to be designed independent of any bias. Furthermore, the user of a citation recommendation system might be interested particularly in works she does not yet know. There are also cases in which the authors of the citing and cited document do not know each other, but in which the author of the citing document still favors specific persons as authors of the cited documents. Most notably, sometimes citation cartels exist in the scientific communities, which first of all cite papers within sub-communities [57]. Furthermore, it has been observed that even the country a person comes from, the race, and the gender play a role in the selection of citations [149]. A bias towards citing authors who act as the referee or reviewer of the citing document in a peer-review process is also plausible [162].

3. **Venue and Paper Type:** It is obvious that the venue is an influential factor in selecting appropriate citations for a given text. Highly rated conferences and journals might get higher levels of attention and are privileged compared to lower rated conferences, workshops, and similar publication formats [31, 163]. A bias can go so far that a relatively weak publication in a prestigious journal receives a high number of citations only due to the centrality of the journal [31]. Papers in interdisciplinary journals are more likely to be cited [10]. Last but not least, it should be noted that, in the frame of the widely performed peer-reviewing process, especially papers that were published in the same venue as the citing paper are more often selected as citations [166].

   Many venues have introduced a page limit for submitted papers. As a consequence, authors often choose to cut several citations which would be relevant and important for understanding the content.

4. **Timeliness:** The temporal dimension concerning citing behavior is, to the best of our knowledge, relatively unexplored in the context of citation recommendation. On the one hand, due to the acceleration in the publishing rate of scientific contributions, authors of citing papers might target citing especially recent papers. On the other hand, older papers have more citations and are easier to find. Note also that the reasons for citing specific publications can change over time [34].

5. **Accessibility and Visibility:** During the citing process, researchers are limited by their capabilities for finding appropriate publications for citing. In particular, they typically cite only papers to which they have fulltext access. However, a considerable amount of researchers have limitations in this regard, such as having no license for accessing papers of specific publishers (e.g., ACM or Springer) and paper collections. Consequently, the set of citable papers is narrowed down considerably. Hence, either not all concepts and claims in the citing paper can be backed up by citations or they cannot be backed up by optimal citations.

   Papers are also embedded in the social interactions and dissemination processes of researchers. Most notably, the claim that prominent publications get cited more is comprehensible and well-studied, even though more relevant alternative publications might exist for citation [164]. Prominent papers are papers which already have a high number of citations, or papers written by authors who are well known in the field and who also have a high aggregated citation count. We can refer in this context to the studies on the so-called *Matthew effect* [14] and on the *Google Scholar effect* [137]. Particularly prominent papers are called *landmark papers* and *citation classics* [140]. They are characterized by the fact that they are often added as citations in a ritualized way and self-enforce their citing.

   Last but not least, it cannot be neglected that nowadays many publications are disseminated via social networks and other channels. Research on these as-

pects in the context of citing behavior has been performed only to a limited extent [96].

6. **Discipline:** Firstly, researchers naturally work within scientific communities and disciplines, with the consequence that they are often exclusively familiar with works published in their discipline or field and that it is difficult for them to discover papers from other fields (due to different venues, terminology, etc.). Hence, citations tend to be limited by the affiliation to the discipline (or even research field).

Secondly, the citing behavior also changes from discipline to discipline. Comparing the citing behavior across disciplines, and, hence, comparing also citation recommendation systems trained and tested on different disciplines, is challenging. For instance, disciplines differ in (1) the number of articles published, (2) the number of co-authors per paper, (3) the relevance of the publication type (e.g., journal, conference, book) for publishing, and (4) the age of cited papers [108]. These aspects have a direct influence on the relevance function of any citation recommendation model. Investigations and evaluations on the context of citation recommendation approaches are missing so far, however. As stated in Section 3, evaluations on citation recommendation have been performed mainly on corpora containing only computer science publications.

5.2.2 Cite-Worthiness of Contexts

Citation recommendation systems typically consider predefined citation contexts for their prediction. However, in reality, typically not only the provided citation contexts are cite-worthy, but also further contexts. Among others, one reason for missing citations is the page restriction which authors need to fulfill for submitting papers to venues.[33] In the past, there have been a few approaches for assessing the cite-worthiness of potential citation contexts automatically, however, only in the sense of a binary classification task [54,23,148,53]. Although there are single works on characterizing the citation context, such as on the citation function, the citation importance, and the citation polarity (see Section 2.4), these aspects are not considered in citation recommendation approaches so far. In particular the type of citation, given as the citation function or in the

form of another classification, such as whether the citation backs up a single concept or a claim, seems to be a notable aspect to be considered.

5.2.3 Scenario Specificity

As outlined in Section 2.2, citation recommendation systems can be applied in different scenarios, differing in particular in (1) the user type (see expert vs. nonexpert setting) and (2) in the type and length of input text. Considering these nuances during evaluation makes a comparison of approaches difficult. However, it is necessary, as the comparison would be unfair otherwise. For instance, citation recommendation systems using only text from an abstract perform differently than ones based on a paper's full text (see the MAP@all score of 0.16 [100] vs. 0.64 [99]). In contrast to that, the difference in the usability of systems for different user types can be assessed via online evaluations and user studies.

5.3 Discussion

Based on the given observations, we propose the following suggestions for an improved evaluation of citation recommendation systems:

Concerning **offline evaluations** In the main, nDCG, MRR, MAP, and recall have been used as the evaluation metric in existing offline evaluations. We recommend using them for the top k recommendations with a rather low value for k (e.g., $k = 5$ or $k = 10$) as in [74, 103,75,67], since it is in our view realistic to return only very few recommendations to the user per citation context (and not using e.g., nDCG@50, and nDCG@75 as in [72] or MAP@100 as in [153]). Tang and Zhang [151] agree with us that it is hard to specify for each citation context how many recommended citations should be returned and notes that for simplicity, the average number of citations per paper could be set as k (e.g., 11 in [151]), if the whole input document is considered. Common evaluation metrics used for citation recommendation reflect the reality only in the sense of the citing behavior observed in the past, but not alternatively valid citations. So far, only a few citation recommendation systems have been evaluated based on alternative offline evaluation metrics (see "relaxed citation re-prediction" in Section 5.1). For instance, the precision metric is softened and papers are also assessed as a hit if they are only related to the cited publications in the citation graph. We argue that such metrics need to be taken with care in the light that citation recommendation aims to back up specific claims and concepts.

---

[33] The San Francisco Declaration on Research Assessment (DORA; http://www.ascb.org/dora/) from 2012 targets the improvement of ways in which the outputs of scientific research are evaluated, and was signed by over 13,000 researchers and institutions. In this declaration, it is proposed that authors should not be restricted by page limitations for references any more, or at least should have reduced restrictions. The reality, however, still looks different.

*Concerning **online evaluations and user studies***
As outlined in Section 5.1, user studies and online evaluations are so far missing in the context of citation recommendation, while offline evaluations predominate. The situation is therefore similar to the situation in the field of paper recommendation [15]. Similar to [18], we recommend performing user studies and online evaluations as necessary steps in the future. This might be particularly fruitful (1) for determining a reasonable ratio of citations per document (cf. cite-worthiness of contexts), and (2) for assessing the relevance of alternative citations, which can be even more relevant than the original citations.[34] Differentiating and automatically determining different levels of relevance seems to be necessary to address this issue, as outlined by [143]. Studies on the importance and grading of citations are rare (see Section 2.4.3), and, to the best of our knowledge, there are no user assessment studies on assessing alternative papers in the context of (personalized or unpersonalized) citation recommendation.

*Concerning **citing biases*** In order to minimize the biases in the citing behavior, the corpora used for training and testing might need to be changed. For instance, only those publications might be considered for which a high degree of fairness can be guaranteed. Single publications could be classified in this respect and might receive a confidence value concerning biases [127].

To not introduce a citing bias via recommending specific papers, citation recommendation systems should use large paper collections (see Sec. 4) and the information which recommendation algorithm and candidate papers are used, should be made available to the user.

*Concerning **scenario specificity*** Similar to paper recommender systems [15], the evaluation results of citation recommendation approaches are often not reproducible, since the data sets are not available and/or many important details of the implementation are omitted in the papers due to constraints such as page limitations [13]. Therefore, we recommend making evaluation data sets, the implementation of the system, and the calculation of evaluation metrics as transparent as possible. Also the targeted scenario (see Section 2.2) and use case characteristics should be clearly visible.

6 Potential Future Work

There are still many variations of the architectures and of the input and output of citation recommendation sys-

tems which have not been considered yet. More specifically, we can think of the following adaptations to enhance and improve citation recommendation:

– Topically diversifying recommended citations [35];
– Recommending papers which state similar, related, or contrary claims as the ones in the citation contexts (i.e., recommending not only papers with identical claims);
– Inserting a sufficient (optimal) set of citations; this could be useful in the presence of paper size limitation, which may be imposed, for example, by conferences. A citation recommendation system should then prioritize important citation contexts that cannot be left without the insertion of citations, while perhaps skipping other less important ones in order to keep the paper size within the limits;
– Given an input text with already present citations, suggesting newer/better ones to update some obsolete/poor citations;
– Combating the cold-start problem for freshly published papers which are not yet cited, hence no training data is available on them;
– Incorporating information on social networks among researchers and considering knowledge sharing platforms; such data can offer additional (often timely) hints on the appropriateness of papers to be cited in particular citation contexts;
– Focusing on specific user groups, which have a given pre-knowledge in common (see our listed scenarios in Sec. 2.2);
– Studying the influences of citing behavior on citation recommendation systems and developing methods for minimizing citing biases in citation recommendation such as biases arising from researchers belonging to the same domains, research groups, or geographical areas (cf. Section 5.2);
– Developing global context-aware citation recommendation approaches, i.e., approaches that recommend citations in a context-aware way, yet still consider the entire content of a paper;
– Recommending citations refuting an argument (using argumentation mining);
– Designing domain-specific citation recommendation approaches and evaluating generic approaches on different disciplines (outside computer science).

Besides these concrete future works, we can think of the following visions in the long term, which embrace a new process of citing in the future:

1. One can envision that, in the future, citation recommendation approaches could better capture the semantics of the citation context, with the result that actual fact-based citation recommendation would

---

[34] The fact that other documents are more relevant as citations can also be observed for Wikipedia, see [56].

have good chance to become reality. This suggests the opportunity of obtaining precise citation recommendations, since both the claims in the citation context and the claims in the candidate cited documents are represented explicitly in a semantically-structured form. In this sense, citation recommendation systems might be capable of not only citing publications, but also any knowledge (in particular, facts and events) available on the Web. This vision becomes particularly feasible in light of the Linked Open Data (LOD) cloud and is in line with research on LOD-based recommender systems [120].

2. One can envision that the working style of researchers would dramatically change in the next few decades [33, 115]. As a result, we might think not only of citation recommendation as considered in this article, but one based on the expected or potential characteristics of scientific publishing. For instance, one can imagine that publications will not be published in PDF format any more, but in either an annotated and more structured version of it (with information about the hypotheses, the methods, the data sets, the evaluation set-up, and the evaluation results), or in the form of a flexible publication format (e.g., subversioning system), in which authors can subsequently change the content, especially the citations, since over the time citations might become obsolete or new citations might become relevant.

7 Conclusions and Outlook

In this survey, we gave a profound overview of the research field of citation recommendation. To that end, we firstly introduced citation recommendation via outlining possible scenarios and via a description of the task. We saw that the approaches to context-aware citation recommendation can be grouped into hand-crafted feature-based models, topic models, machine translation models, and neural network models. The approaches do not only differ with respect to the underlying method, but also with respect to the provided input data. More specifically, the considered set-ups differ in the use of a user model, the prefiltering of candidate papers, the length of the citation context, whether citation placeholders are provided, and whether the content of cited papers is needed. Concerning the evaluation, the approaches are evaluated based on very diverse metrics and different data sets, making it hard to assess the validity and advance of single approaches. Moreover, approaches are often compared to existing approaches to a limited extent.

We also considered the data sets that can be used for deploying and evaluating citation recommendation.

We distinguished between corpora containing papers' content and corpora providing papers' metadata. Here we learned that several corpora exist, especially in the field of computer science. However, the data sets differ considerably in their size and in their quality (e.g., noise due to information extraction).

Concerning the challenges of evaluating citation recommendation and the evaluation methods used so far, we found out that biases in the citing behavior have largely been ignored, as well as the "worthiness" to cite at all or in specific circumstances. Assessing citation recommendations might also depend on the scientific discipline and on the concrete use case. Approaches have been evaluated rather unilaterally and not across disciplines.

Upcoming approaches on citation recommendation are likely to be based on more advanced techniques of machine learning, such as variants of recurrent neural networks. In the long term, one can envision that citation recommendation approaches can better capture the semantics of the citation context, with the result that actual fact-based citation recommendation becomes reality. Given the likely continuation and proliferation of the "tsunami" of publications and of citations in the years and decades to come, we can assume that citation recommendation will become an integral component of a researcher's working environment.

References

1. Amjad Abu-Jbara, Jefferson Ezra, and Dragomir R. Radev. Purpose and Polarity of Citation: Towards NLP-based Bibliometrics. In *Proceedings of the 2013 Conference of the North American Chapter of the Association of Computational Linguistics: Human Language Technologies*, NAACL-HLT'13, pages 596–606, 2013.
2. Amjad Abu-Jbara and Dragomir R. Radev. Coherent Citation-Based Summarization of Scientific Papers. In *Proceedings of the 2011 Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT'11, pages 500–509, 2011.
3. Amjad Abu-Jbara and Dragomir R. Radev. Reference Scope Identification in Citing Sentences. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT'12, pages 80–90, 2012.
4. Shahbaz Ahmad and Muhammad Tanveer Afzal. Combining Co-citation and Metadata for Recommending More Related Papers. In *Proceedings of the 15th International Conference on Frontiers of Information Technology*, FIT'17, pages 218–222, 2017.
5. Dag W. Aksnes. A macro study of self-citation. *Scientometrics*, 56(2):235–246, 2003.
6. AllenAI. Science Parse. `https://github.com/allenai/science-parse`, 2019. Accessed: 2020-04-06.
7. Myriam Hernández Alvarez and José M. Gómez. Survey about citation context analysis: Tasks, techniques, and

resources. *Natural Language Engineering*, 22(3):327–349, 2016.

8. Anas Alzoghbi, Victor Anthony Arrascue Ayala, Peter M. Fischer, and Georg Lausen. Pubrec: Recommending publications based on publicly available meta-data. In *Proceedings of the LWA 2015 Workshops: KDML, FGWM, IR, and FGDB*, pages 11–18, 2015.

9. Ankesh Anand, Tanmoy Chakraborty, and Amitava Das. FairScholar: Balancing Relevance and Diversity for Scientific Paper Recommendation. In *Proceedings of the 39th European Conference on IR Research*, ECIR'17, pages 753–757, 2017.

10. Anupama Annalingam, Hasitha Damayanthi, Ranil Jayawardena, and Priyanga Ranasinghe. Determinants of the citation rate of medical research publications from a developing country. *SpringerPlus*, 3(1):140, 2014.

11. Xiaomei Bai, Mengyang Wang, Ivan Lee, Zhuo Yang, Xiangjie Kong, and Feng Xia. Scientific Paper Recommendation: A Survey. *IEEE Access*, 7:9324–9339, 2019.

12. Hannah Bast and Claudius Korzen. A Benchmark and Evaluation for Text Extraction from PDF. In *Proceedings of the 17th Joint Conference on Digital Libraries*, JCDL'17, pages 99–108, 2017.

13. Jöran Beel, Corinna Breitinger, Stefan Langer, Andreas Lommatzsch, and Bela Gipp. Towards reproducibility in recommender-systems research. *User Model. User-Adapt. Interact.*, 26(1):69–101, 2016.

14. Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: An Introductory Overview. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics*, ISSI'09, pages 230–241, 2009.

15. Jöran Beel, Bela Gipp, Stefan Langer, and Corinna Breitinger. Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4):305–338, 2016.

16. Jöran Beel, Bela Gipp, Stefan Langer, and Marcel Genzmehr. Docear: an academic literature suite for searching, organizing and creating academic literature. In *Proceedings of the 2011 Joint International Conference on Digital Libraries*, JCDL'11, pages 465–466, 2011.

17. Jöran Beel, Bela Gipp, Stefan Langer, Marcel Genzmehr, Erik Wilde, Andreas Nürnberger, and Jim Pitman. Introducing Mr. DLib: a machine-readable digital library. In *Proceedings of the 2011 Joint International Conference on Digital Libraries*, JCDL'11, pages 463–464, 2011.

18. Jöran Beel and Stefan Langer. A Comparison of Offline Evaluations, Online Evaluations, and User Studies in the Context of Research-Paper Recommender Systems. In *Proceedings of the 19th International Conference on Theory and Practice of Digital Libraries*, TPDL'15, pages 153–168, 2015.

19. Marc Bertin, Iana Atanassova, Yves Gingras, and Vincent Larivière. The invariant distribution of references in scientific articles. *Journal of the Association for Information Science and Technology*, 67(1):164–177, 2016.

20. Steven Bethard and Dan Jurafsky. Who Should I Cite: Learning Literature Search Models from Citation Behavior. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management*, CIKM'10, pages 609–618, 2010.

21. Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar. Content-Based Citation Recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT'18, pages 238–251, 2018.

22. Steven Bird, Robert Dale, Bonnie J. Dorr, Bryan R. Gibson, Mark Thomas Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir R. Radev, and Yee Fan Tan. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, LREC'08, 2008.

23. Hamed Bonab, Hamed Zamani, Erik G. Learned-Miller, and James Allan. Citation Worthiness of Sentences in Scientific Reports. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR'18, pages 1061–1064, 2018.

24. Lutz Bornmann and Hans-Dieter Daniel. What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1):45–80, 2008.

25. Lutz Bornmann and Rüdiger Mutz. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11):2215–2222, 2015.

26. Brian Kurilla. Can too much science be a bad thing? Growth in scientific publishing as a barrier to science communication. http://geekpsychologist.com/can-too-much-science-be-a-bad-thing-growth-in-scientific-publishing-as-a-barrier-to-science-communication/, 2015. Accessed: 2019-06-19.

27. R. K. Buter and Anthony F. J. van Raan. Non-alphanumeric characters in titles of scientific publications: An analysis of their occurrence and correlation with citation impact. *Journal of Informetrics*, 5(4):608–617, 2011.

28. Xiaoyan Cai, Junwei Han, Wenjie Li, Renxian Zhang, Shirui Pan, and Libin Yang. A three-layered mutually reinforced model for personalized citation recommendation. *IEEE Trans. Neural Netw. Learning Syst.*, 29(12):6026–6037, 2018.

29. Xiaoyan Cai, Junwei Han, and Libin Yang. Generative Adversarial Network Based Heterogeneous Bibliographic Network Representation for Personalized Citation Recommendation. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, AAAI'18, pages 5747–5754, 2018.

30. Xiaoyan Cai, Yu Zheng, Libin Yang, Tao Dai, and Lantian Guo. Bibliographic Network Representation Based Personalized Citation Recommendation. *IEEE Access*, 7:457–467, 2019.

31. Callaham, M and Wears, RL and Weber, E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Journal of the American Medical Association*, 287(21):2847–50, 2002.

32. Cornelia Caragea, Jian Wu, Alina Maria Ciobanu, Kyle Williams, Juan Pablo Fernández Ramírez, Hung-Hsuan Chen, Zhaohui Wu, and C. Lee Giles. CiteSeer x : A Scholarly Big Dataset. In *Proceedings of the 36th European Conference on IR Research*, ECIR'14, pages 311–322, 2014.

33. Fabio Casati, Fausto Giunchiglia, and Maurizio Marchese. Liquid publications: Scientific publications meet the web. Technical report, University of Trento, 2007.

34. Donald Owen Case and Georgeann M. Higgins. How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the*

American Society for Information Science, 51(7):635–645, 2000.

35. Tanmoy Chakraborty, Natwar Modani, Ramasuri Narayanam, and Seema Nagar. DiSCern: A diversified citation recommendation system for scientific queries. In *Proceedings of the 31st IEEE International Conference on Data Engineering*, ICDE'15, pages 555–566, 2015.

36. Tanmoy Chakraborty and Ramasuri Narayanam. All Fingers are not Equal: Intensity of References in Scientific Articles. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, EMNLP'16, pages 1348–1358, 2016.

37. David A. Cohn and Thomas Hofmann. The Missing Link – A Probabilistic Model of Document Content and Hypertext Connectivity. In *Advances in Neural Information Processing Systems 13*, NIPS'00, pages 430–436, 2000.

38. Alexandru Constantin, Steve Pettifer, and Andrei Voronkov. PDFX: fully-automated PDF-to-XML conversion of scientific literature. In *Proceedings of the 2013 ACM Symposium on Document Engineering*, DocEng'13, pages 177–180, 2013.

39. Isaac G. Councill, C. Lee Giles, and Min-Yen Kan. ParsCit: An open-source CRF reference string parsing package. In *Proceedings of the International Conference on Language Resources and Evaluation*, LREC'08, 2008.

40. CrossRef Labs. pdf-extract. `https://github.com/CrossRef/pdfextract`, 2015. Accessed: 2018-01-31.

41. Tao Dai, Li Zhu, Xiaoyan Cai, Shirui Pan, and Sheng Yuan. Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network. *J. Ambient Intelligence and Humanized Computing*, 9(4):957–975, 2018.

42. Tao Dai, Li Zhu, Yifan Wang, Hongfei Zhang, Xiaoyan Cai, and Yu Zheng. Joint model feature regression and topic learning for global citation recommendation. *IEEE Access*, 7:1706–1720, 2019.

43. Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.

44. Ying Ding, Guo Zhang, Tamy Chambers, Min Song, Xiaolong Wang, and Chengxiang Zhai. Content-based Citation Analysis: The Next Generation of Citation Analysis. *Journal of the Association for Information Science and Technology*, 65(9):1820–1833, 2014.

45. Daniel Duma and Ewan Klein. Citation Resolution: A method for evaluating context-based citation recommendation systems. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL'14, pages 358–363, 2014.

46. Daniel Duma, Ewan Klein, Maria Liakata, James Ravenscroft, and Amanda Clare. Rhetorical classification of anchor text for citation recommendation. *D-Lib Magazine*, 22(9/10), 2016.

47. Daniel Duma, Maria Liakata, Amanda Clare, James Ravenscroft, and Ewan Klein. Applying Core Scientific Concepts to Context-Based Citation Recommendation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, LREC'16, 2016.

48. Travis Ebesu and Yi Fang. Neural Citation Network for Context-Aware Citation Recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'17, pages 1093–1096, 2017.

49. Aaron Elkiss, Siwei Shen, Anthony Fader, Günes Erkan, David J. States, and Dragomir R. Radev. Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the Association for Information Science and Technology*, 59(1):51–62, 2008.

50. Daniel Faensen, Lukas Faulstich, Heinz Schweppe, Annika Hinze, and Alexander Steidinger. Hermes: A Notification Service for Digital Libraries. In *Proceedings of the Joint Conference on Digital Libraries*, JCDL'01, pages 373–380, 2001.

51. Michael Färber, Ashwath Sampath, and Adam Jatowt. PaperHunter: A System for Exploring Papers and Citation Contexts. In *Proceedings of the 41th European Conference on Information Retrieval*, ECIR'19, 2019.

52. Michael Färber, Alexander Thiemann, and Adam Jatowt. A High-Quality Gold Standard for Citation-based Tasks. In *Proceedings of the International Conference on Language Resources and Evaluation*, LREC'18, 2018.

53. Michael Färber, Alexander Thiemann, and Adam Jatowt. CITEWERTs: A System Combining Cite-Worthiness with Citation Recommendation. In *Proceedings of the 40th European Conference on Information Retrieval*, ECIR'18, pages 815–819, 2018.

54. Michael Färber, Alexander Thiemann, and Adam Jatowt. To Cite, or Not to Cite? Detecting Citation Contexts in Text. In *Proceedings of the 40th European Conference on Information Retrieval*, ECIR'18, pages 598–603, 2018.

55. Besnik Fetahu, Katja Markert, and Avishek Anand. Automated News Suggestions for Populating Wikipedia Entity Pages. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, CIKM'15, pages 323–332, 2015.

56. Besnik Fetahu, Katja Markert, Wolfgang Nejdl, and Avishek Anand. Finding News Citations for Wikipedia. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, CIKM'16, pages 337–346, 2016.

57. Iztok Fister, Iztok Fister, and Matjaž Perc. Toward the Discovery of Citation Cartels in Citation Networks. volume 4, page 49, 2016.

58. Santo Fortunato, Carl T Bergstrom, Katy Börner, James A Evans, Dirk Helbing, Staša Milojević, Alexander M Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, et al. Science of science. *Science*, 359(6379):eaao0185, 2018.

59. Soumyajit Ganguly and Vikram Pudi. Paper2vec: Combining Graph and Text Information for Scientific Paper Representation. In *Proceedings of the 39th European Conference on IR Research*, ECIR'17, pages 383–395, 2017.

60. Zequn Gao. Examining influences of publication dates on citation recommendation systems. In *Proc. of the 12th International Conference on Fuzzy Systems and Knowledge Discovery*, FSKD'15, pages 1400–1405, 2015.

61. Souvick Ghosh, Dipankar Das, and Tanmoy Chakraborty. Determining sentiment in citation text and analyzing its impact on the proposed ranking index. *CoRR*, abs/1707.01425, 2017.

62. C. Lee Giles, Kurt D. Bollacker, and Steve Lawrence. CiteSeer: An Automatic Citation Indexing System. In *Proceedings of the 3rd ACM International Conference on Digital Libraries*, DL'98, pages 89–98, 1998.

63. Bela Gipp. *Citation-based Plagiarism Detection – Detecting Disguised and Cross-language Plagiarism using Citation Pattern Analysis*. Springer, 2014.

64. Marco Gori and Augusto Pucci. Research paper recommender systems: A random-walk based approach. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, WI'06, pages 778–781, 2006.

65. Lantian Guo, Xiaoyan Cai, Fei Hao, Dejun Mu, Changjian Fang, and Libin Yang. Exploiting Fine-Grained Co-Authorship for Personalized Citation Recommendation. *IEEE Access*, 5:12714–12725, 2017.

66. Matthias Hagen, Anna Beyer, Tim Gollub, Kristof Komlossy, and Benno Stein. Supporting Scholarly Search with Keyqueries. In *Proceedings of the 38th European Conference on IR Research*, ECIR'16, pages 507–520, 2016.

67. Jialong Han, Yan Song, Wayne Xin Zhao, Shuming Shi, and Haisong Zhang. hyperdoc2vec: Distributed Representations of Hypertext Documents. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, ACL'18, pages 2384–2394, 2018.

68. Seyyed Hadi Hashemi, Mahmood Neshati, and Hamid Beigy. Expertise Retrieval in Bibliographic Network: A Topic Dominance Learning Approach. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*, CIKM'13, pages 1117–1126, 2013.

69. Jing He, Jian-Yun Nie, Yang Lu, and Wayne Xin Zhao. Position-Aligned Translation Model for Citation Recommendation. In *Proceedings of the 19th International Symposium on String Processing and Information Retrieval*, SPIRE'12, pages 251–263, 2012.

70. Qi He, Bi Chen, Jian Pei, Baojun Qiu, Prasenjit Mitra, and C. Lee Giles. Detecting Topic Evolution in Scientific Literature: How Can Citations Help? In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM'09, pages 957–966, 2009.

71. Qi He, Daniel Kifer, Jian Pei, Prasenjit Mitra, and C. Lee Giles. Citation Recommendation without Author Supervision. In *Proceedings of the Forth International Conference on Web Search and Web Data Mining*, WSDM'11, pages 755–764, 2011.

72. Qi He, Jian Pei, Daniel Kifer, Prasenjit Mitra, and C. Lee Giles. Context-aware Citation Recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW'10, pages 421–430, 2010.

73. Bo-Yu Hsiao, Chih-Heng Chung, and Bi-Ru Dai. A Model of Relevant Common Author and Citation Authority Propagation for Citation Recommendation. In *Proceedings of the 16th IEEE International Conference on Mobile Data Management*, MDM'15, pages 117–119, 2015.

74. Wenyi Huang, Saurabh Kataria, Cornelia Caragea, Prasenjit Mitra, C. Lee Giles, and Lior Rokach. Recommending Citations: Translating Papers into References. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM'12, pages 1910–1914, 2012.

75. Wenyi Huang, Zhaohui Wu, Liang Chen, Prasenjit Mitra, and C. Lee Giles. A Neural Probabilistic Model for Context Based Citation Recommendation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, AAAI'15, pages 2404–2410, 2015.

76. Wenyi Huang, Zhaohui Wu, Prasenjit Mitra, and C. Lee Giles. RefSeer: A citation recommendation system. In *Proceedings of the 14th Joint Conference on Digital Libraries*, JCDL'14, pages 371–374, 2014.

77. Tin Huynh, Kiem Hoang, Loc Do, Huong Tran, Hiep Phuc Luong, and Susan Gauch. Scientific Publication Recommendations Based on Collaborative Citation Networks. In *Proceedings of the International Conference on Collaboration Technologies and Systems*, CTS'12, pages 316–321, 2012.

78. Ken Hyland. Self-citation and self-reference: Credibility and promotion in academic publication. *Journal of the Association for Information Science and Technology*, 54(3):251–259, 2003.

79. Emi Ishita, Yasuko Hagiwara, Yukiko Watanabe, and Yoichi Tomiura. Which Parts of Search Results do Researchers Check when Selecting Academic Documents? In *Proceedings of the 18th on Joint Conference on Digital Libraries*, JCDL'18, pages 345–346, 2018.

80. Kris Jack, Pablo López-García, Maya Hristakeva, and Roman Kern. {{citation needed}}: Filling in Wikipedia's Citation Shaped Holes. In *Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval*, BIR'14, pages 45–52, 2014.

81. Chanwoo Jeong, Sion Jang, Hyuna Shin, Eunjeong Park, and Sungchul Choi. A Context-Aware Citation Recommendation Model with BERT and Graph Convolutional Networks. *CoRR*, abs/1903.06464, 2019.

82. Haofeng Jia and Erik Saule. An Analysis of Citation Recommender Systems: Beyond the Obvious. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM'17, pages 216–223, 2017.

83. Haofeng Jia and Erik Saule. Local Is Good: A Fast Citation Recommendation Approach. In *Proceedings of the 40th European Conference on IR Research*, ECIR'18, pages 758–764, 2018.

84. Zhuoren Jiang. Citation Recommendation via Time-series Scholarly Topic Analysis and Publication Prior Analysis. *TCDL Bulletin*, 9(2), 2013.

85. Zhuoren Jiang, Xiaozhong Liu, and Liangcai Gao. Dynamic Topic/Citation Influence Modeling for Chronological Citation Recommendation. In *Proceedings of the 5th International Workshop on Web-scale Knowledge Representation Retrieval & Reasoning*, Web-KR@CIKM'14, pages 15–18, 2014.

86. Zhuoren Jiang, Xiaozhong Liu, and Liangcai Gao. Chronological Citation Recommendation with Information-Need Shifting. In *Proceedings of the 24th Internat. Conference on Information and Knowledge Management*, CIKM'15, pages 1291–1300, 2015.

87. Zhuoren Jiang, Yao Lu, and Xiaozhong Liu. Cross-language citation recommendation via publication content and citation representation fusion. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, JCDL'18, pages 347–348, 2018.

88. Zhuoren Jiang, Yue Yin, Liangcai Gao, Yao Lu, and Xiaozhong Liu. Cross-language Citation Recommendation via Hierarchical Representation Learning on Heterogeneous Graph. In *Proc. of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR'18, pages 635–644, 2018.

89. Saurabh Kataria, Prasenjit Mitra, and Sumit Bhatia. Utilizing Context in Generative Bayesian Models for Linked Corpus. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, AAAI'10, 2010.

90. Ralf Klamma, Manh Cuong Pham, and Yiwei Cao. You Never Walk Alone: Recommending Academic Events Based on Social Network Analysis. In *Proceedings of the*

First International Conference on Complex Sciences, Complex'09, pages 657–670, 2009.

91. Yuta Kobayashi, Masahi Shimbo, and Yuji Matsumoto. Citation Recommendation Using Distributed Representation of Discourse Facets in Scientific Articles. In *Proceedings of the 2018 Joint International Conference on Digital Libraries*, JCDL'18, pages 243–251, 2018.

92. Onur Küçüktunç, Erik Saule, Kamer Kaya, and Ümit V. Çatalyürek. Diversifying Citation Recommendations. *ACM Transactions on Intelligent Systems and Technology*, 5(4):55:1–55:21, December 2014.

93. Onur Küçüktunç, Erik Saule, Kamer Kaya, and Ümit V. Çatalyürek. TheAdvisor: A Webservice for Academic Recommendation. In *Proceedings of the 13th Joint Conference on Digital Libraries*, JCDL '13, pages 433–434, 2013.

94. Peder Olesen Larsen and Markus Von Ins. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3):575–603, 2010.

95. Shuchen Li, Peter Brusilovsky, Sen Su, and Xiang Cheng. Conference paper recommendation for academic conferences. *IEEE Access*, 6:17153–17164, 2018.

96. Jennifer Lin and Martin Fenner. Altmetrics in Evolution: Defining & Redefining the Ontology of Article-Level Metrics. *Information Standards Quarterly*, 25(2):20–26, 2013.

97. Xiang Liu, Torsten Suel, and Nasir D. Memon. A Robust Model for Paper Reviewer Assignment. In *Proceedings of the Eighth ACM Conference on Recommender Systems*, RecSys'14, pages 25–32, 2014.

98. Xiaozhong Liu, Yingying Yu, Chun Guo, and Yizhou Sun. Meta-Path-Based Ranking with Pseudo Relevance Feedback on Heterogeneous Graph for Citation Recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM 2014, pages 121–130, 2014.

99. Xiaozhong Liu, Yingying Yu, Chun Guo, Yizhou Sun, and Liangcai Gao. Full-text based Context-Rich Heterogeneous Network Mining Approach for Citation Recommendation. In *Proceedings of the Joint Conference on Digital Libraries*, JCDL'14, pages 361–370, 2014.

100. Xiaozhong Liu, Jinsong Zhang, and Chun Guo. Citation Recommendation via Proximity Full-Text Citation Analysis and Supervised Topical Prior. In *Proceedings of the iConference 2016*, 2016.

101. Ya'ning Liu, Rui Yan, and Hongfei Yan. Guess What You Will Cite: Personalized Citation Recommendation Based on Users' Preference. In *Proceedings of the 9th Asia Information Retrieval Societies Conference*, AIRS'13, pages 428–439, 2013.

102. Ziming Liu. Citation theories in the framework of international flow of information: New evidence with translation analysis. *Journal of the American Society for Information Science*, 48(1):80–87, 1997.

103. Avishay Livne, Vivek Gokuladas, Jaime Teevan, Susan T. Dumais, and Eytan Adar. CiteSight: supporting contextual citation recommendation using differential search. In *Proceedings of the 37th International Conference on Research and Development in Information Retrieval*, SIGIR '14, pages 807–816, 2014.

104. Patrice Lopez. GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications. In *Proceedings of the 13th European Conference on Digital Libraries*, ECDL'09, pages 473–474, 2009.

105. Patrice Lopez and Laurent Romary. GROBID - Information Extraction from Scientific Publications. *ERCIM News*, 2015(100), 2015.

106. Wen-Yang Lu, Yu-Bin Yang, Xiao-Jiao Mao, and Qi-Hai Zhu. Effective Citation Recommendation by Unbiased Reference Priority Recognition. In *Proceedings of the 17th Asia-Pacific Web Conference*, APWeb'15, pages 536–547, 2015.

107. Yang Lu, Jing He, Dongdong Shan, and Hongfei Yan. Recommending citations with translation model. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management*, CIKM'11, pages 2017–2020, 2011.

108. Michael Mabe and Adrian Mulligan. What Journal Authors Want: Ten Years of Results from Elsevier's Author Feedback Programme. *New Review of Information Networking*, 16(1):71–89, 2011.

109. Parvaz Mahdabi and Fabio Crestani. Query-driven mining of citation networks for patent citation retrieval and recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM'14, pages 1659–1668, 2014.

110. Sean M. McNee, Istvan Albert, Dan Cosley, Prateep Gopalkrishnan, Shyong K. Lam, Al Mamunur Rashid, Joseph A. Konstan, and John Riedl. On the Recommending of Citations for Research Papers. In *Proceeding on the ACM 2002 Conference on Computer Supported Cooperative Work*, CSCW'02, pages 116–125, 2002.

111. Michael Färber and Ashwath Sampath. Determining the Linguistic Types of Citations. In *Proceedings of the 22nd International Conference on Theory and Practice of Digital Libraries*, TPDL'18, 2019.

112. Arunav Mishra. Linking Today's Wikipedia and News from the Past. In *Proceedings of the 7th PhD Workshop in Information and Knowledge Management*, PIKM'14, pages 1–8, 2014.

113. Arunav Mishra and Klaus Berberich. Leveraging Semantic Annotations to Link Wikipedia and News Archives. In *Proceedings of the 38th European Conference on IR Research*, ECIR'16, pages 30–42, 2016.

114. Saif Mohammad, Bonnie J. Dorr, Melissa Egan, Ahmed Hassan Awadallah, Pradeep Muthukrishnan, Vahed Qazvinian, Dragomir R. Radev, and David M. Zajic. Using Citations to Generate surveys of Scientific Paradigms. In *Proceedings of the 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL-HLT'09, pages 584–592, 2009.

115. Paolo Montuschi and Alfredo Benso. Augmented reading: The present and future of electronic scientific publications. *IEEE Computer*, 47(1):64–74, 2014.

116. Raymond J. Mooney and Loriene Roy. Content-based Book Recommending Using Learning for Text Categorization. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, DL'00, pages 195–204, New York, NY, USA, 2000. ACM.

117. Michael J Moravcsik and Poovanalingam Murugesan. Some Results on the Function and Quality of Citations. *Social studies of science*, 5(1):86–92, 1975.

118. Dejun Mu, Lantian Guo, Xiaoyan Cai, and Fei Hao. Query-focused personalized citation recommendation with mutually reinforced ranking. *IEEE Access*, 6:3107–3119, 2018.

119. Ramesh Nallapati, Amr Ahmed, Eric P. Xing, and William W. Cohen. Joint latent topic models for text and citations. In *Proceedings of the 14th ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining, KDD'08, pages 542–550, 2008.

120. Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, and Markus Zanker. Linked Open Data to support Content-based Recommender Systems. In *Proceedings of the 8th International Conference on Semantic Systems*, I-SEMANTICS '12, pages 1–8, 2012.

121. NSF. Science and Engineering Indicators 2014. `https://www.nsf.gov/statistics/seind14/`, 2014. Accessed: 2019-06-19.

122. Sooyoung Oh, Zhen Lei, Wang-Chien Lee, Prasenjit Mitra, and John Yen. CV-PCR: A Context-guided Value-driven Framework for Patent Citation Recommendation. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*, CIKM'13, pages 2291–2296, 2013.

123. Hanna Pasula, Bhaskara Marthi, Brian Milch, Stuart J. Russell, and Ilya Shpitser. Identity Uncertainty and Citation Matching. In *Advances in Neural Information Processing Systems 15: Proceedings of the Neural Information Processing Systems Conference*, NIPS'02, pages 1401–1408, 2002.

124. Hao Peng, Jing Liu, and Chin-Yew Lin. News Citation Recommendation with Implicit and Explicit Semantics. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, ACL'16, 2016.

125. Silvio Peroni and David M. Shotton. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Journal of Web Semantics*, 17:33–43, 2012.

126. Vayianos Pertsas and Panos Constantopoulos. Scholarly Ontology: modelling scholarly practices. *International Journal on Digital Libraries*, 18(3):173–190, 2017.

127. Evaggelia Pitoura, Panayiotis Tsaparas, Giorgos Flouris, Irini Fundulaki, Panagiotis Papadakos, Serge Abiteboul, and Gerhard Weikum. On Measuring Bias in Online Information. *SIGMOD Record*, 46(4):16–21, 2017.

128. Animesh Prasad, Manpreet Kaur, and Min-Yen Kan. Neural ParsCit: a deep learning-based reference string parser. *International Journal on Digital Libraries*, 19(4):323–337, 2018.

129. Dragomir R. Radev, Pradeep Muthukrishnan, Vahed Qazvinian, and Amjad Abu-Jbara. The ACL anthology network corpus. *Language Resources and Evaluation*, 47(4):919–944, 2013.

130. James Ravenscroft, Amanda Clare, and Maria Liakata. HarriGT: A Tool for Linking News to Science. In *Proceedings of ACL'18 System Demonstrations*, pages 19–24, 2018.

131. Xiang Ren, Jialu Liu, Xiao Yu, Urvashi Khandelwal, Quanquan Gu, Lidan Wang, and Jiawei Han. ClusCite: Effective Citation Recommendation by Information Network-Based Clustering. In *Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining*, KDD'14, pages 821–830, 2014.

132. Anna Ritchie. *Citation context analysis for information retrieval*. PhD thesis, University of Cambridge, UK, 2009.

133. Anna Ritchie, Stephen Robertson, and Simone Teufel. Comparing Citation Contexts for Information Retrieval. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM'08, pages 213–222, 2008.

134. Lior Rokach, Prasenjit Mitra, Saurabh Kataria, Wenyi Huang, and Lee Giles. A Supervised Learning Method for Context-Aware Citation Recommendation in a Large Corpus. In *Proceedings of the Large-Scale and Distributed Systems for Information Retrieval Workshop*, LSDS-IR'13, pages 17–22, 2013.

135. Dwaipayan Roy, Kunal Ray, and Mandar Mitra. From a Scholarly Big Dataset to a Test Collection for Bibliographic Citation Recommendation. In *Proceedings of Scholarly Big Data Workshop*, 2016.

136. Tarek Saier and Michael Färber. Bibliometric-Enhanced arXiv: A Data Set for Paper-Based and Citation-Based Tasks. In *Proceedings of the 8th International Workshop on Bibliometric-enhanced Information Retrieval*, BIR'19, pages 14–26, 2019.

137. Alexander Serenko and John Dumay. Citation classics published in Knowledge Management journals. Part II: studying research trends and discovering the Google Scholar Effect. *Journal of Knowledge Management*, 19(6):1335–1355, 2015.

138. Ritu Sharma, Dinesh Gopalani, and Yogesh Meena. Concept-Based Approach for Research Paper Recommendation. In *Proceedings of the 7th International Conference on Pattern Recognition and Machine Intelligence*, PReMI'17, pages 687–692, 2017.

139. Ayush Singhal, Ravindra Kasturi, Vidyashankar Sivakumar, and Jaideep Srivastava. Leveraging Web Intelligence for Finding Interesting Research Datasets. In *Proceedings of the 2013 International Conferences on Web Intelligence*, WI'13, pages 321–328, 2013.

140. Henry Small. On the shoulders of Robert Merton: Towards a normative theory of citation. *Scientometrics*, 60(1):71–79, 2004.

141. Luciana B Sollaci and Mauricio G Pereira. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *Journal of the medical library association*, 92(3):364, 2004.

142. Laura Steinert. *Beyond Similarity and Accuracy – A New Take on Automating Scientific Paper Recommendations*. PhD thesis, University of Duisburg-Essen, Germany, 2017.

143. Trevor Strohman, W. Bruce Croft, and David Jensen. Recommending Citations for Academic Papers. Technical report, 2007.

144. Trevor Strohman, W. Bruce Croft, and David D. Jensen. Recommending Citations for Academic Papers. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'07, pages 705–706, 2007.

145. Sinisa Subotic and Bhaskar Mukherjee. Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science*, 40(1):115–124, 2014.

146. Kazunari Sugiyama and Min-Yen Kan. Exploiting potential citation papers in scholarly paper recommendation. In *Proceedings of the 13th Joint Conference on Digital Libraries*, JCDL '13, pages 153–162, 2013.

147. Kazunari Sugiyama and Min-Yen Kan. A comprehensive evaluation of scholarly paper recommendation using potential citation papers. *International Journal on Digital Libraries*, 16(2):91–109, 2015.

148. Kazunari Sugiyama, Tarun Kumar, Min-Yen Kan, and R. C. Tripathi. Identifying Citing Sentences in Research Papers Using Supervised Learning. In *Proceedings of the 2010 International Conference on Information Retrieval & Knowledge Management*, CAMP'10, pages 67–72. IEEE, 2010.

149. Iman Tahamtan, Askar Safipour Afshar, and Khadijeh Ahamdzadeh. Factors affecting number of citations: a

comprehensive review of the literature. *Scientometrics*, 107(3):1195–1225, 2016.

150. Iman Tahamtan and Lutz Bornmann. Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics*, 12(1):203 – 216, 2018.

151. Jie Tang and Jing Zhang. A Discriminative Approach to Topic-Based Citation Recommendation. In *Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, PAKDD'09, pages 572–579, 2009.

152. Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'08, pages 990–998, 2008.

153. Xuewei Tang, Xiaojun Wan, and Xun Zhang. Cross-language Context-aware Citation Recommendation in Scientific Articles. In *Proceedings of the 37th International Conference on Research and Development in Information Retrieval*, SIGIR '14, pages 817–826, 2014.

154. Simone Teufel, Advaith Siddharthan, and Dan Tidhar. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP'07, pages 103–110, 2006.

155. Simone Teufel, Advaith Siddharthan, and Dan Tidhar. An annotation scheme for citation function. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 80–87, 2009.

156. Dominika Tkaczyk, Andrew Collins, Paraic Sheridan, and Jöran Beel. Evaluation and comparison of open source bibliographic reference parsers: A business use case. *CoRR*, abs/1802.01168, 2018.

157. Dominika Tkaczyk, Andrew Collins, Paraic Sheridan, and Jöran Beel. Machine Learning vs. Rules and Out-of-the-Box vs. Retrained: An Evaluation of Open-Source Bibliographic Reference and Citation Parsers. In *Proceedings of the 18th Joint Conference on Digital Libraries*, JCDL'18, pages 99–108, 2018.

158. Dominika Tkaczyk, Pawel Szostek, Mateusz Fedoryszak, Piotr Jan Dendek, and Lukasz Bolikowski. CERMINE: automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition*, 18(4):317–335, 2015.

159. R. Todeschini and A. Baccini. *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research*. Wiley, 2016.

160. Marco Valenzuela, Vu Ha, and Oren Etzioni. Identifying Meaningful Citations. In *Scholarly Big Data: AI Perspectives, Challenges, and Ideas*, SBD'15, 2015.

161. Peiling Wang and Dagobert Soergel. A Cognitive Model of Document Use During a Research Project. Study I. Document Selection. *Journal of the American Society for Information Science*, 49(2):115–133, 1998.

162. Peiling Wang and Marilyn Domas White. A Cognitive Model of Document Use During a Research Project. Study II. Decisions at the Reading and Citing Stages. *Journal of the American Society for Information Science*, 50(2):98–114, 1999.

163. Mark Ware and Michael Mabe. The STM Report: An overview of scientific and scholarly journal publishing. 2015.

164. Howard D. White. Citation Analysis and Discourse Analysis Revisited. *Applied Linguistics*, 25(1):89–116, 2004.

165. Howard D. White. Bag of Works Retrieval: TF*IDF Weighting of Co-cited Works. In *Proceedings of the Third Workshop on Bibliometric-enhanced Information Retrieval*, BIR'16, pages 63–72, 2016.

166. Allen W. Wilhite and Eric A. Fong. Coercive citation in academic publishing. *Science*, 335(6068):542–543, 2012.

167. Hao Wu, Yu Hua, Bo Li, and Yijian Pei. Enhancing citation recommendation with various evidences. In *Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery*, FSKD'12, pages 1160–1165, 2012.

168. Jian Wu, Athar Sefid, Allen C. Ge, and C. Lee Giles. A Supervised Learning Approach To Entity Matching Between Scholarly Big Datasets. In *Proceedings of the Knowledge Capture Conference*, K-CAP'17, pages 41:1–41:4, 2017.

169. Libin Yang, Zeqing Zhang, Xiaoyan Cai, and Lantian Guo. Citation recommendation as edge prediction in heterogeneous bibliographic network: A network representation approach. *IEEE Access*, 7:23232–23239, 2019.

170. Libin Yang, Yu Zheng, Xiaoyan Cai, Hang Dai, Dejun Mu, Lantian Guo, and Tao Dai. A LSTM Based Model for Personalized Context-Aware Citation Recommendation. *IEEE Access*, 6:59618–59627, 2018.

171. Libin Yang, Yu Zheng, Xiaoyan Cai, Shirui Pan, and Tao Dai. Query-oriented citation recommendation based on network correlation. *Journal of Intelligent and Fuzzy Systems*, 35(4):4621–4628, 2018.

172. Tianbao Yang, Rong Jin, Yun Chi, and Shenghuo Zhu. Combining Link and Content for Community Detection: A Discriminative Approach. In *Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining*, KDD'09, pages 927–936, 2009.

173. Zaihan Yang and Brian D. Davison. Venue Recommendation: Submitting Your Paper with Style. In *Proceedings of the 11th International Conference on Machine Learning and Applications*, ICMLA'12, pages 681–686, 2012.

174. Jun Yin and Xiaoming Li. Personalized Citation Recommendation via Convolutional Neural Networks. In *Proceedings of the First International Joint Conference on Web and Big Data*, APWeb-WAIM'17, pages 285–293, 2017.

175. Fattane Zarrinkalam and Mohsen Kahani. SemCiR: A citation recommendation system based on a novel semantic distance measure. *Program*, 47(1):92–112, 2013.

176. Ye Zhang, Libin Yang, Xiaoyan Cai, and Hang Dai. A Novel Personalized Citation Recommendation Approach Based on GAN. In *24th International Symposium on Foundations of Intelligent Systems*, ISMIS'18, pages 268–278, 2018.