Scientometrics (2018) 114:335–357
https://doi.org/10.1007/s11192-017-2560-2

A content-based citation analysis study based on text categorization

Zehra Taşkın[1] · Umut Al[1]

Received: 16 August 2017 / Published online: 30 October 2017

Abstract Publications and citations are important components for measuring research performance. Academics receive incentives, tenures, or awards from the number of citations they receive; however, the use of citations for research/er evaluation purposes can give rise to unethical practices and manipulation. Consequently, it is necessary to change the current approach to the use of citations. The main aim of this study was to conduct a content-based citation analysis study for Turkish citations. To achieve this aim, 423 peer-reviewed articles, the associated 12,881 references, and 101,019 sentences published in library and information science literature in Turkey were thoroughly examined. The citations were divided into four main categories; citation meaning, citation purpose, citation shape, and citation array. Then, each category was further divided into sub-categories. A tagging process with inter-annotator agreement was conducted and citation categories for the citation sentences determined. Weka software was used to apply the text categorization methods. The automatic citation sentence classification achieved at least a 90% success rate for all citation classes, which proved that using computational linguistics to evaluate citation contexts developing new techniques was possible and gave more detailed results.

Keywords Content-based citation analysis · Qualitative research evaluation · Text categorization · Weka

✉ Zehra Taşkın
ztaskin@hacettepe.edu.tr

Umut Al
umutal@hacettepe.edu.tr

[1] Department of Information Management, Hacettepe University, 06800 Çankaya, Ankara, Turkey

Introduction

Scientific publications are important for publicizing research findings, with the relationships made between studies being critical in driving further research (Shum 1998, p. 19). In other words, the relationships are vital in spreading knowledge. The most fundamental element connecting research is the citation (Teufel 1999, p. 33), which academic authors use to support, elaborate on, or debunk, as these are seen to be critically related to their work. In this respect, citations are vital for establishing relationships between publications (Oppenheim 1996, p. 155). From this point of view, the basic function of citations is to establish a connection and relationship between the cited and citing publications (Smith 1981, p. 84). Citations are figuratively similar to frozen footprints in academic achievements (Cronin 1981, p. 16). With these footprints, it is possible to identify information from researchers who have come before, and find clues to subject development. In other words, knowledge is developed through citations, and if references are properly given, they provide a powerful and versatile tool for researchers. The development of scientific knowledge is the process of discovery, evaluation, consolidation, and reassessment (Cronin 1981, p. 20). For this reason, the network of links provided from citations between publications has great importance in academia.

The use of citations in performance evaluations has been the subject of discussions for many years, for which there have been two distinct views (Bornmann and Daniel 2008, p. 46). In one view, citation analysis is seen as an appropriate tool for rewards, identifying Nobel laureates, identifying prestige, academic rankings, peer assessments, and promotions (Cole 2000; Van Raan 2004). However, the alternate view is that citations should not be used for research/er evaluations for various reasons related to time, discipline, accessibility, and other factors (Cozzens 1985; Woolgar 1991). Considering all these factors, it is possible to improve citation-based evaluations. However, the important issue is "why do authors cite?" Garfield (1970, p. 85), the creator of citation indexes, listed the most popular citation motivations; to respect pioneers in a field, to give credit to related publications, to explain the methods and tools used, to provide background information, to correct their own or others' works, to criticize previous studies, and to verify data. Although these reasons answer the question as to why citations are made, the most related publications may not be cited, and irrelevant publications may be cited if resources are cited randomly (Kochen 1974, p. 74; Smith 1981, p. 84). For this reason, although the expectation is to strengthen the citation chain by citing the most relevant publications, it is possible to cite relatively less relevant articles. Price (1986, p. 58) noted that due to individual differences, authors do not always cite resources with the same consistency, completeness and honesty. The accuracy of Price's assessment can be seen clearly in today's practices. Nowadays, some authors do not cite their competitors or colleagues for strategic reasons. Editors or journal referees may request the addition of coercive citations from authors to increase the number of citations for these authors or for their journal (COPE 2012). Therefore, the number of citations can be easily manipulated using such practices.

One example of such a manipulation was revealed in the journal *Energy Education Science and Technology*, which was indexed in the Web of Science. As the result of doubts arising from the high self-citation rate, it was determined that most citations were from a "sister" journal, both of which had the same editor (Öztürk 2012; Kaplan 2014). Consequently, there were complaints about the ethics of such behavior (Al and Soydal 2012; Öztürk 2013), following which, the journal was removed from the Web of Science index in 2013. The interesting issue was that the Scientific and Technological Research Council of Turkey (TUBITAK) had given first rank incentives to social science researchers publishing in this journal (289 applications—151,624 Turkish Liras) before its removal (Kaplan 2013). This example is not unique in studies on citation manipulation. In another case, an academician, who was the editor in-chief of a geophysics journal and had peer-review roles at other geophysics journals, was involved in citation manipulation for the articles he had refereed (Davis 2017). After long discussions, which were interpreted as "citation cartels or editor gone rogue," the editor resigned from his job and an investigation was started (Oransky 2017). In 2016, 10 journals were suppressed from the Web of Science due to high self-citation rates, and citation-stacking was found for an additional three journals, which were also removed from the index (Title suppressions 2016). However, this is not surprising; as long as citations continue to be seen as an important criterion for the evaluation of research, researchers, or institutions, citation misuse will continue.

Despite the problems, manipulations, and criticisms in the past about counting citations, they are preferred by managers and decision makers who wish to make quick, effortless evaluations without the need for questionnaires or interviews (MacRoberts and MacRoberts 1996, p. 435; Smith 1981, p. 84). In this context, some databases (such as Web of Science and Scopus) are used as the main research evaluation information sources, with the authors being recognized as the most prominent[1] being given rewards or incentives by decision makers (Lerner and Wulf 2007, p. 634; Miller et al. 2013, p. 520). Using bibliometric methods, the most important authors, institutions, and countries in a field can be determined, scientific fields can be mapped, co-authorship analyses conducted, and the science effect evaluated.[2] However, citations should only be regarded as an indication that the citing author actually used the article and that its use transformed into a benefit (quality, value, or impact) (Smith 1981, p. 87).

To avoid the current equal evaluation of all citations, the aim of the present study is to design an evaluation model that can analyze both the semantic and syntactic citation structures to determine taxonomic citation categories that can replace traditional citation counting. This research is shaped around the hypothesis that "all citations are not equal", with the main objective being to design a tool that can assess the semantic and syntactic structures of Turkish citations to provide a content-based evaluation model for research evaluations. In this context, from a close analysis of Turkish academic texts, taxonomic citation categories were established using machine-learning processes to automatically detect these categories from high-volume texts. Therefore, considering the forgoing, the main research questions are as follows:

- How can Turkish citations be taxonomically categorized? Is it possible to create a classification scheme for these citations?
- Are there any differences between the taxonomic citation categories for the different sections in journal publications (introduction, methods, findings, etc.)?

[1] Websites such as Essential Science Indicators (http://wokinfo.com/products_tools/analytical/essentialscienceindicators/), Highly Cited Researchers (http://highlycited.com/) and ScienceWatch (http://archive.sciencewatch.com/) present rankings of authors, institutions and countries by using number of publications and citations.

[2] Numbers of citations are important indicators for tenures and incentives in Turkey. For example, authors who have received high citation rate for their publications are supported by Scientific Research Projects Coordination Unit of Hacettepe University to travel abroad for international conferences (Hacettepe Üniversitesi… 2015). In addition, numbers of citations to publications are important for tenures and academic promotions (Öğretim Üyeliğine Yükseltilme… 1982). There is a separate section for citations in "Academic Incentive Payment" given to academic staff working at state universities. Each citation is graded by using different evaluation elements such as position, number of authors, citations' origin etc. (Akademik 2016).

- How are positive, negative, and neutral citations used in Turkish literature, and is there a definable language for easy citation classification?
- Using the results of this work, is it possible to create a machine-learning model that can detect types of citations from Turkish texts?

As there have not been any previous machine-learning models for content-based citation analysis in Turkey and limited research focus elsewhere, the results of this research may assist decision makers and managers when making researcher evaluation decisions.

Literature review

Many studies have examined quality, effectiveness, usefulness, visibility, or other aspects of citations and several studies have also discussed timing, publishers or –by many researchers as misleading (Moravcsik and Murugesan 1975, p. 86). Garfield, the creator of citation indexes, emphasized that the use of citations to evaluate a paper was not wise, and argued that the citation frequency was a measure of the extent of a research activity rather than the significance of an author's work, stating that it was possible to measure the influence of a paper, not the author, from counting citations. For this reason, it is necessary to use other measures as well as citations for performance evaluations (Garfield 1973, p. 407). Goudsmith (1974, p. 28) felt that researchers may make citations to gather more citations for others or may not cite competitors within a citation reward system. A much older study also reported that (Ziman 1968, p. 58) citations may be given for politeness or political reasons and therefore could not be accepted as effective evaluation indicators. Oppenheim (1996, p. 156) claimed that the "the more papers you cite in your own article, the more likely it is that your article will subsequently be cited!" which could result in a greater number of citations that contribute little to the subject. On the contrary to concerns about increasing number of citations, Vinkler (1994, p. 499) claimed that numbers were not data and data were not indicators; however, it is not possible to maintain bibliometrics without relevant data, appropriate methods, or indicators.

Bibliometrics has been a main focus in many citation analyses from the late 1970s and 1980s with many criticisms being made regarding the negative or meaningless motivations for citations. An article published in 1979 (Garfield 1979) argued that negative and self-citations did not significantly influence citation analyses and further claimed that as negative citations were extremely rare in scientific publications, this did not affect citation analyses (Carter 1974). Another claim was that negative citations were also meaningful because science develops from criticism and extension (Garfield 1979, p. 362). Some early studies also claimed that erroneous publications were also valuable for their contributions to scientific literature (Cole and Cole 1971, p. 26); for instance, many Nobel laureates' pre-Nobel articles were at first rejected. In the same study, the assumption that methodological papers had the potential to attract more citations than others was emphasized; however, it was found that most methodological articles (73%) did not attract any citations when cited in large numbers (Garfield 1979, p. 363). Garfield also claimed that citation analyses could not measure an effect not defined by scientific authorities. In response to Garfield, Chubin (1980) argued about which scientific authorities should have the responsibilities for these kinds of evaluations. As academic competition has increased across the world and within academic circles within countries, studies on the meaning/meaninglessness of counting citations has intensified since these very early papers.

In the 1980s, because of the paucity of research, it was not considered meaningful to evaluate developing countries using citation indexes in the same way as in the English content citation indexes (Arunachalam and Manorama 1988, pp. 394, 406). However, with the changes in the use of citation indexes and the increase in academic publications in developing countries, there has been a commensurate regional expansion in these indexes (Testa 2008). Numbers, which can easily be obtained from citation indexes, are being used to evaluate academic performances or determine a university position in scientific communities. Although there has been a great deal of discussion on the drawbacks of only using number of citations to evaluate researchers, countries where significant progress in science has yet to be made, continue to judge the value of academics on the number of citations (Tonta 2014, pp. 16–17; Van Raan 2005).

It has been suggested that citation analysis studies are inadequate to measure scientific development as many authors may read and refer to randomly selected publications in the field (MacRoberts and MacRoberts 1996, pp. 436), and it was found in one study (Simkin and Roychowdhury 2003, p. 269) that only about 20% of cited papers were actually fully read. Simkin and Roychowdhury in a later study claimed that while comparative studies concentrated on counting citations, this was pointless if academics had not read the cited publications (Simkin and Roychowdhury 2006, p. 172), especially as many authors were found to copy the cited references made by others in the research preparation process. This could be seen as a demonstration of the least effort theory (Zipf 1949, p. 1), which is the desire to reach the best output with minimal effort. Wetterer (2006) reported on an example of citation copying in a study on big ants living in Madeira, whereby the author when citing subject matter made a mistake in translating the information from German to English. Consequently, the same mistake was repeated in most subsequent article citations. The most important reason for repeating erroneous information in new studies is that the authors wish to use the already translated article without referring to the original article. In addition to not sighting the original citation source, various other problems have also been found such as biased citations, references to secondary sources, changing the citation motivation by field or time, using rejected findings, and citing biased data (MacRoberts and MacRoberts 1996, p. 436–438).

Another approach for citation analysis is that a publication that attracts at least one citation is more likely to be cited again than a non-cited article, primarily because academics and scientists tend to follow the paths created by past studies. Although this "citation pearl growing" tendency is a known approach (Markey and Cochrane 1981, p. 19), citations lose their meaning when authors have not sighted the original articles. This citation method tends to confirm the belief that "at least one cited publication will also be cited in the future", which resembles a sentence from Matta Bible's 25th chapter, "unto every one that hath shall be given, and he shall have abundance" (Matthew 25:29 2004; Simkin and Roychowdhury 2006, p. 181). Based on this verse, Robert Merton, who introduced the concept of the Matthew Impact in 1968, implied that reputed or distinguished researchers would have more credibility and dignity than a researcher whose name was unheard of, even if their works were similar (Merton 1968, p. 59). Merton claimed that this Matthew Effect directly influenced individual researchers' careers and rewards. Similar to the Matthew Effect, the "success breeds success" approach emphasized the fact that older researchers had more advantages than younger researchers (Cozzens 1985, p. 149). There have been several other theories proposed about the factors affecting the number of citations [e.g., Stigler's Law of Eponymy (Stigler 1980) and Ortega Hypothesis (Cole and Cole 1972)].

In a study criticizing the counting of citations, Oppenheim argued that these analyses were not reliable, as not all citations were equal (Oppenheim 1996, p. 157), claiming that counting citations was not sufficient to measure "impact" or "quality" as everything was reduced to numbers, and even the mistakes made when citing could affect results by 10–20%. One important problem in citation analysis has been incorrect publications. For example, it was discovered that only seven of the 55 publications produced by Darsee were valid, 40 were questionable, and eight were definitely fraudulent. However, 198 citations are gathered between 1982 and 1990; in other words, author can gather a high number of citations, even if the research is questionable. Further, 86% of the citations to Darsee's works only mentioned or confirmed his studies (Oppenheim 1996, p. 158). This situation is no different for retracted articles. In a content-based citation analysis study of retracted articles, the citations for the top five most-cited retracted articles in 2015 were examined (Halevi and Bar-Ilan 2016). Despite the expectation that the majority of these citations were going to be negative, there was no mention that these articles had been retracted, some citations were positive, and the articles were still fully accessible on the publishers' websites free of charge due to their potential to gather citations. In another article on retracted articles, it was found that many works were still being cited many years after they had been retracted (Al and Soydal 2015, p. 32) and nearly half (45%) had been made after the retractions. Another study attempted to explain the reasons for citations being made after retraction and found influencing factors such as unclear publisher websites, the presence of pirated websites, the use of older versions on the web, and author intentions to hide the retractions. It was concluded that these issues created significant problems for academic rigor (Silva and Dobránszki 2017, p. 1653).

In recent years, several studies have been conducted on how citations are used in researcher evaluations to avoid evaluations that are completely focused on quantity. When evaluating scientific studies or providing incentives to researchers, going beyond quantitative evaluations as decision makers is necessary to assess the contribution of the researchers to the field as part of the evaluation process (Al and Soydal 2014, p. 40); therefore, it is important to focus on what has been written about rather than counting the number of articles that have been published.

Studies evaluating citations by content rather than quantity have been conducted since the 1950s and can be generally divided into four types; (a) evaluations based on a syntactic approaches that examine the position of the citation in the text, (b) studies that evaluate the semantic relationships between the cited and citing publications (such as positive, negative, and neutral citations), (c) studies that investigate citation frequency in a single study, and (d) studies that classify citation motivations. In almost all these studies, the first question asked has been "why do authors cite?" New generation citation analyses, which are based on the reasons for the author's citations, are known as "content-based citation analysis" (Ding et al. 2014, p. 1820), which in general can be divided into semantic and syntactic approaches. Semantic approaches are concerned with how the citation is made and syntactic approaches examine where the citation is made in the text.

Thirty to forty years ago, content-based citation analyses were often not generalized because of the sample sizes and techniques used. However, today, with the rapid developments in computational linguistics, it has become easier to apply content-based citation analyses to publications because of the open access to full-text documents, the ability to process large-scale texts, and the development of various analysis algorithms (Teufel 1999, p. 38). With the development of machine-learning techniques and the ability to use computers for analyses, content-based citation analyses are now being implemented using machine-learning techniques. There have been various computational linguistics techniques developed for citation analysis, such as citation recommendation systems (Liu et al. 2013), data mining from citations (Schneider and Borlund 2005), information retrieval (Aljaber et al. 2010; Fu and Aliferis 2010; Liu et al. 2014; Ritchie 2008), sentiment analysis (Athar 2011; Cavalcanti et al. 2011; Tandon and Jain 2012; Yu 2013), citation categorization (Bertin 2008), and citation summarization (Elkiss et al. 2008; Tandon and Jain 2012). Of these, automated text categorization techniques were used in this study to automatically classify citation sentences.

In the literature, content-based citation analysis studies have been tested by using computational techniques with various applications since the 2000s (Angrosh et al. 2010; Athar 2014; Ding et al. 2013; Dong and Schäfer 2011; Maricic et al. 1998; Xu et al. 2015; Sendhilkumar et al. 2013; Suppe 1998; Teufel et al. 2006; Zhu et al. 2015). The results of these studies are compared in detail with the results of our study in the Findings section.

Methodology

The methodology is presented step-by-step in this section to assist others in using these content-based analysis techniques in their own fields. The main process phases are shown in Fig. 1.

Identifying the dataset

A concern in citation evaluation studies in the literature has been the need for field-based assessments. Just as each individual has different information seeking behavior, citation styles specific to specific areas also differ, and therefore must be evaluated in the light of these differences (Cano 1989, p. 284). Within this context, Turkish articles published in *Türk Kütüphaneciliği (Turkish Librarianship-TL)* and *Bilgi Dünyası (Information World-IW)* journals, which are essential journals for Turkish LIS publications, constituted the main dataset. As both journals are open access, there was no problem accessing the collections. Only peer-reviewed articles were considered and all other document types were excluded.

Data collection and processing

Within the scope of the study, all peer-reviewed articles published in *TL* and *IW* were saved in pdf format using an optical character recognition process to scan the documents as images. Then, all articles were converted to txt files and UTF-8 character encoding was selected to identify the special Turkish characters. All txt files were used for the background structure of the developed database.

After collection, all articles were given smart identity numbers with the structure *journal name* + *year* + *volume* + *issue* information; for example, article number TK201031 represents the first article published in TL in volume three, issue 1 in 2010. After these processes, a MySql-based relational database was designed to collect the data for the CBCA process to ensure the metadata, references, and full-text data were kept in a standard structure. After the database was created, to facilitate the data analyses, an interface design was implemented.

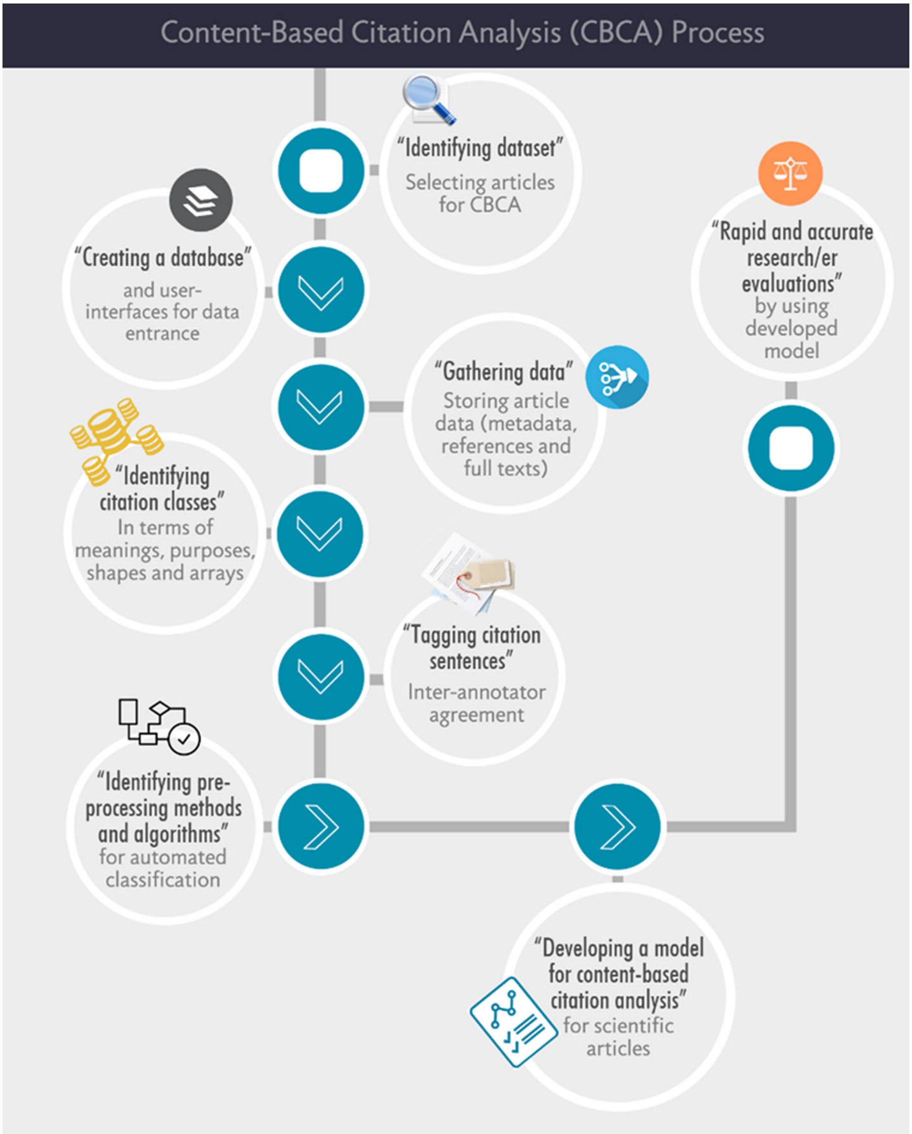

Fig. 1 Main phases for the content-based citation analysis process

Designing data interfaces

As it was important in this study to collect all data for each article, the metadata, references, and full texts were all stored in the database; therefore a three-level data entry structure was developed at the interface.

The first level determined the basic article elements; author names, titles, abstracts, and keywords. The interface was able to automatically determine the main fields using key terms at the beginning or ending of the texts such as the keywords placed between the *"keywords"* and the *"introduction"*. On this first level, the data entry operators were able to correct incorrect classifications using the interface editing tools. The second level was reserved for collecting the cited references from the articles. This process was also automated and determined by the machine, which examined the "references" and any other titles that were similar to "references." The third level was designed to determine all sentences and the IMRAD (Introduction, Methodology, Research, and Discussion) sections in the articles. The interface divided the articles into the sentences using a period (.) sign; special uses of the period such as in titles (Ph.D., Dr.) or numbers (1., X.) were also identified by the interface. Using the interface's dropdown menu, the data entry operators determined the main paragraph sections based on the IMRAD structure.

After completing this three-level data entry for all articles in the dataset, 12,881 references and 101,019 sentences from 423 articles were stored in the database. The main taxonomic citation classes were then determined before the content-based citation analysis process.

Identifying the citation classes

The taxonomic citation categories are shown in Fig. 2.

- *Meaning* The most discussed topics in the content-based citation analysis literature were divided into positive, negative, and neutral citations. While some researchers have argued that negative citations can develop disciplines in a positive way (Carter 1974; Cole and Cole 1971, p. 26; Garfield 1979, p. 362), others have claimed that negative and positive citations are not the same (Chubin 1980; Spiegel-Rösing 1977; Voos and Dagaev 1976). Citations that do not add any value to scientific works have also been criticized (Moravcsik and Murugesan 1975). Therefore, it was important to assess the citations for meaning, for which three sub-classes; positive, negative, and neutral; were determined.

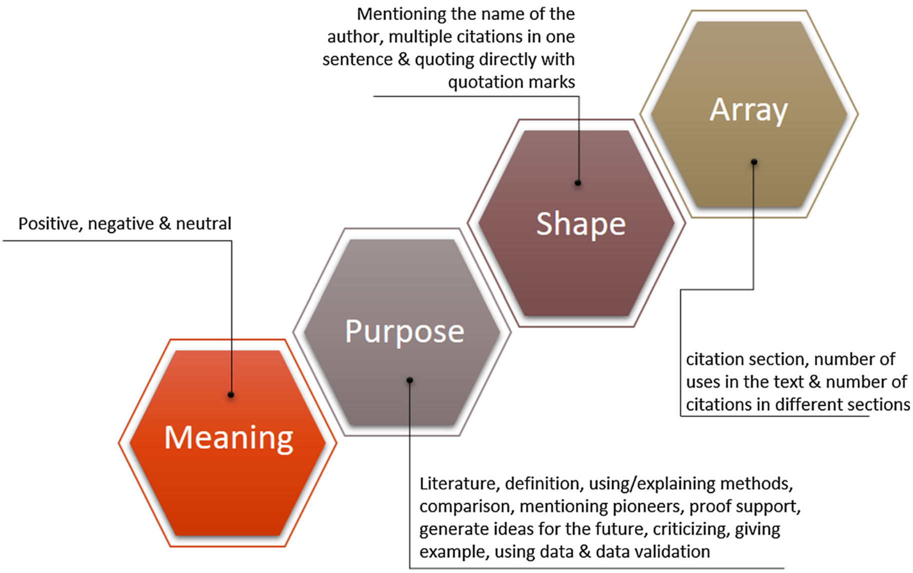

Fig. 2 Taxonomic citation categorizations

- *Purpose* The citation purpose refers to the elements the authors highlight when using a citation, for which five citation sub-classes were determined; literature, definition, data, method, and data validation. While there are obviously a greater number of purpose sub-classes such as comparison, mentioning pioneers, proof support, generating ideas for the future, criticizing, and giving examples, for text categorization simplicity, the five most important purpose sub-classes were determined.
- *Shape* Works are cited in various ways; mentioning the name of the author, quoting directly with quotation marks, or multiple citations in one sentence (Bonzi 1982, p. 211); therefore, it is important to classify citations by shape, for which three main classes were defined for this study. However, citations could fall into more than one class; for example, an author may cite a name and give a quote from the article.
- *Array* If a publication was cited more than once in an article (Herlach 1978, p. 310), was cited in the findings, or was based on the method developed in the cited article (Maricic et al. 1998, pp. 530–540), these citations were considered to be more valuable than others. Therefore, the citations were also categorized by their array, the subgroups for which were; citation section, number of uses in the text, and number of citations in the different sections.

To determine the answers to the research questions, evaluations were based on the above classes.

Tagging the citation sentences

During the tagging phase, a tagging interface was designed and training was provided for the operators on the tagging process to ensure inter-annotator agreement.

First, an interface was developed to tag the citation sentences according to their class. Using this interface, the operators could quickly and practically complete the tagging process. The main tagging process was as follows;

- The operator logged into the system using their username and password. The main reasons for this check were to mark which label was being tagged by whom, and to prevent operators from seeing each other's' labels. Then, the operator selected the article using the dropdown menu.
- All sentences in the selected article were displayed on the right side of the interface. To the right of each sentence, there were sentence selection boxes to classify the cited sentences. As a citation sentence could be one sentence or more than one sentence, the term, "citation sentence," referred to a sentence or a group of sentences.
- The relevant citation sentence reference (or references) was chosen from the references in the dropdown menu on the left.
- The operator then tagged the citation sentences according to meaning, purpose, and shape and saved the citation sentence transaction.

One of the most important conditions when conducting natural language processing tasks is to ensure inter-annotator agreement. Studies have found that there can be several problems related to accuracy and objectivity in the tagging processes if there is no consensus (Artstein and Poesio 2008, p. 591; Landis and Koch 1977, p. 159). Therefore, each citation sentence was tagged by at least two operators and the tags that were most similar were analyzed in the natural language processing stage. Six expert operators (four Ph.Ds. and two under-graduate students) worked on the tagging process. At the end of tagging process, the first group of taggers had identified 14,259 citation sentences, and the second group of taggers had identified 14,840. Detailed statistical information on the subjects is presented in the findings section. From the tagging process and inter-annotator agreement, 13,866 citation sentences were determined for the meaning corpus, 10,437 for the purpose corpus, and 13,527 for the shape corpus.

Automatic categorization of citations

An automatic text categorization technique was used to classify the citations. The categorization algorithms were based on various classes depending on the basic text features (Blake 2013, p. 136). The tags selected by the data operators were compared and natural language processing tasks were applied to the most similar citation sentences using the Weka data mining tool, the details of which are shown in Fig. 3.

After the similarities between operators were identified, corpora were developed for each citation class, which were then converted to the arff file format used by the Weka tool, which is very similar to a csv structure. This file format was formed by combining two basic structures; the header and the data. After the required files were created, it was necessary to process the words before the application.

The word pre-processing method selected for the Turkish citations was the *n*-gram algorithm, which converted the texts into vectors. In this transformation, a weighted *n*-gram algorithm of 1–2 grams (bigram) and 1–3 gs (trigram) was preferred. The main aim of the *n*-gram was to determine the repetition rate in a given sequence; that is, to create a sequence of n consecutive numbers (Damashek 1995, p. 843). The word *n*-gram pre-processing was used to evaluate word frequency in terms of meaning and purpose, and the character *n*-gram was used to determine the shape of the citation sentences. The main reason for pre-processing method the citation sentences for shape was to account for the importance of characters such as apostrophes or parentheses.

Generally, in these kinds of analyses, stop words are excluded; however, the analyses revealed that many stop words in Turkish were important in the citation sentences. Therefore, stop words such as “but,” “thus,” “however,” “therefore” were not excluded before processing. The pre-process was followed by the application phase.

As the Naïve Bayes Multinomial, a statistical based algorithm, and the Random Forests algorithm, a decision tree algorithm, have been found to give the most successful results, only the performance results from these two algorithms were reported in this study. After reporting, the algorithmic performances were evaluated using methodological and quantitative techniques.

A tenfold cross validation has generally been preferred to verify algorithmic accuracy because as the analysis is repeatedly tested on the same dataset, *k*-fold cross validation has the longest validation method and can provide the most accurate results. In the cross validation method, the dataset was randomly divided into 10 equal parts, one of which was used for testing, with the remaining nine being used as the training set. This process was

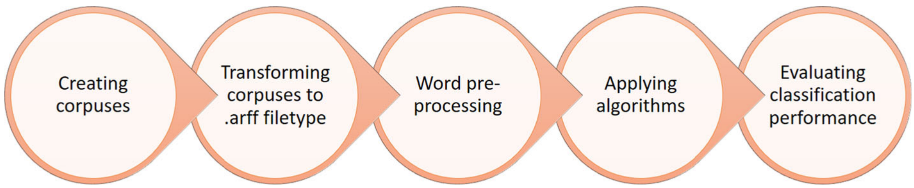

Fig. 3 Machine-learning process for the categorization of Turkish citations

repeated 10 times and the result was averaged to calculate the correctness (Kohavi 1995, p. 1138).

Algorithmic performance ratios (correctly classified citations/all citations in the database) and *f*-measure values were used to report the quantitative evaluations. Confusion matrices were also reported to indicate the success rates in each of the sub-classes.

Findings

The success of the machine-learning in each class was tested using the taxonomic citation classes. The success rates and other details are shown in Fig. 4.

Citation categorization for meaning

When categorizing the citations for meaning, the first data entry operator found 14,259 citations and the second identified 14,840 citation sentences, with the citation distribution in the sub-classes being similar for both operators. The tagged results were found to match, and a corpus for the positive, negative, and neutral citations was created using these matched results, in which 97.2% were identified as neutral, 2% were identified as positive, and 0.8% were identified as negative.

Many studies have proven that the distribution of positive, negative, and neutral citations in scientific texts were generally similar to the results in this study. For instance, 2.4% of the 2309 citations examined in a literature study were found to be positive and 0.4% were negative (Spiegel-Rösing 1977, p. 105). In a study that sought to determine positive, negative, and neutral citations from scientific articles, 3% were identified as negative, 10% were positive, and the others were neutral (Athar 2011, p. 82). In Athar's other work (2014, p. 36), 3.2% of citations were found to be negative, 9.5% were positive, and 87.3% were neutral. Cano (1989, p. 286) found that 2% of a rarely seen class of citations were negative citations. In a study conducted on Supreme Court opinion citations, 33% were positive citations and 8% were negative (Johnson 1985, p. 513). In a study involving the semantic analysis of citations made in clinical trials, 17% were found to be positive and 7% were

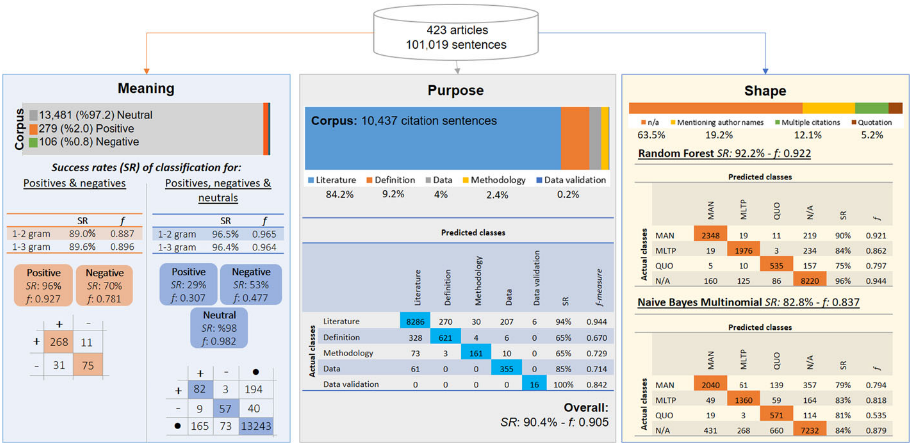

Fig. 4 Success rates for the citation categories

negative (Xu et al. 2015, p. 1338). When the results obtained from this paper's research were compared with the related literature, the semantic classes in the corpus were generally similar, but the number of negative citations was smaller. The main reason could have been that the authors were reluctant to give negative citations. While science develops from criticism, many authors prefer to only hint at negative approaches so as not to attract any negative reactions.

A closer examination of the positive and negative citations found that the authors generally chose certain words or word groups such as "attract attention," "a good example," "worth examining," "very important" in the positive citations and used the words "but," "however" and "although" in the negative citations, reinforcing the argument that negative citations are usually implicit. Athar (2011, p. 82) claimed that when authors make negative citations, they usually preferred to say something positive first, after which they expressed their criticisms using conjunctions and soft words such as "not right" rather than "it is mistaken." In another study, a similar topic was addressed and it was found that the positive and negative citation motives were generally the same and that there was a correlation between these two citation types (Brooks 1986, p. 34); that is, when an author cited a publication, they first emphasized the positive aspects of the paper and then the main criticism was made so as to reduce any negative reactions, which was in line with the results from the analysis in this paper.

After creating the corpus, it was converted to the arff format used by Weka for the analyses. As mentioned in the methodology section, the words had been converted into word vectors using the n-gram tokenizer. The 1–2 and 1–3 gram alternatives were tested for word pre-processing, from which it was found that both gave similar results.

After the word pre-processing step, the dataset in which only the positive and negative citations were included was analyzed to determine whether there were any meaningful differences between the language used for the positive and negative citation classes. As previously mentioned, the Naïve Bayes Multinomial and Random Forest algorithms were used to classifying texts by similarity. Although the Random Forest algorithm showed fairly accurate results ($f = 0.982$) for the two classes (positive and negative), only the results from the Naïve Bayes Multinomial algorithm were reported as the Random Forest performance rate was found to be less accurate when neutral citations were included. The analysis, which was tested using the tenfold cross validation, achieved an 89% performance for classifying the positive and negative citations, with 96% of the positives and 70% of the negatives being correctly detected by the algorithm. The confusion matrix is shown in Fig. 4.

A finding quite similar to this was also obtained in a study (Jha et al. 2016, p. 103) in which the positive and negative citations were classified using natural language processing techniques, with 93% for the positive citations and 78% for the negative citations being achieved. These results demonstrated that there was a significant difference between the positive and negative citations in terms of the language used; positive citations were found to have more determinative language than the negative citations. The main reason for the different success rates for the positive and negative citations was surmised to be because of the unequal distribution in the number of the citations in the categories in the dataset. To test this supposition, the positive and negative citations were equalized by subtracting the positive citations randomly and repeating the analysis, from which the findings for the negative citations increased significantly ($f = 0.871$). Therefore, it was concluded that the performance ratios would possibly increase if the number of negative citation counts in the corpus were greater. However, as negative citations were less common in the literature, this was difficult to do.

Following the testing of the success for the automatic detection of the positive and negative citations by language use, the same analysis was conducted on the entire corpus (positive, negative, and neutral). From this analysis, the overall performance of the classification algorithm was found to be 96% ($f = 0.965$); this high performance ratio was found to be because of the high number of neutral citations. Neutral citations influenced the performance rates as the success rate was 98% ($f = 0.982$), which indicated that few neutral citations were incorrectly identified as positive or negative by the algorithm. Unfortunately, it was difficult to interpret the positive and negative citations in the same way. The detection success rates for the positive citations was 29% ($f = 0.307$) and for the negative citations was 53% ($f = 0.477$), indicating that over the entire corpus, although there were fewer citations, the success rate for finding negative citations was higher than for finding positive citations. Some previous studies have had similar findings, while others have not. In a study that produced similar results and used similar algorithms, an 86% ($f = 0.883$) success was achieved for neutral citations, 68% ($f = 0.614$) for negative citations, and 61% ($f = 0.563$) for positive citations (Athar 2014, p. 79). In a study dissimilar to the present research (Xu et al. 2015, p. 1339), an *f*-measure was used which found 0.498 for negative citations, 0.719 for positive citations, and 0.924 for neutral citations using 1–2 g word pre-processing. When compared to these studies, the detection performance for neutral citations was higher in this study. Regardless, it was decided that the detection performance for the positive and negative citations needed to be improved. In the two papers cited above, linguistic additions (such as sentiment dictionary and parsing) were used to improve the machine's performance. For example, in Xu et al. (2015, p. 1339), the negative citation finding performance increased to $f = 0.551$ and the positive performance increased to $f = 0.723$ using sentiment dictionary and parsing techniques. In Athar (2014), the *f*-measure was used to detect the negative citations, increasing the success from 0.138 to 0.614 using various sentiment analysis additions.

Citation categorization for purpose

As mentioned in the methodology section, citation purpose was divided into 12 sub-classes; literature, definition, comparison, giving examples, proving, data, criticizing, mentioning pioneers, describing methods, using methods, generating ideas for the future, and validation. The overall detection performance rate for the citations based on purpose for these 12 sub-classes was 78% using the Naïve Bayes Multinomial algorithm. The most important detection ratio was for literature citations, with 92% of literature citations being detected ($f = 0.866$). However, as there were many sub-classes in this corpus, this reduced the performance of the algorithm; therefore, to increase this detection rate, it was necessary to examine the literature more deeply.

Various studies have been conducted on taxonomic classification literature citations since the 1960s. In a study evaluating citation classification schemes developed between 1965 and 1989 (Ding et al. 2014, p. 11,825), eight pioneering articles were grouped based on commonalities using content-based citation analyses; method citations were evaluated as a separate group, with conceptual framework, background, and previous research as the other three groups. Subgroups such as comparison, proofing, and validation citations were then classified under the main group of previous research. In more recent studies (Angrosh et al. 2010; Dong and Schäfer 2011; Teufel et al. 2006), positive, negative, and neutral citations were the main citation classification schemes, with literature studies, alternative approaches, comparisons, and methods and techniques being the other classes. Dong and Schäfer (2011, p. 624), using a similar classification to this paper, identified four basic citation groups using machine-learning algorithms such as the Naïve Bayes and Sequential Minimal Optimization; background, mentioning the originator of the idea, technical infrastructure, and comparison; with the performance ratios being between 0.510 and 0.670 (f-measure) from the various algorithms. In another study (Sendhilkumar et al. 2013, p. 417), the literature citations were detected at a 60% success rate, however, for other classes, it did not exceed 18%. From these studies, therefore, it could be concluded that the classification of citations into classes is more meaningful if main classes are first created. In this respect, five main classes were determined for this study; literature, definition, methodology, data, and data validation, after which the analysis was repeated for the new grouped sub-classes using a 1–2 g word pre-processing method and the Naïve Bayes Multinomial algorithm. The overall performance was 90.4% and $f = 0.905$, with all performances achieving greater than 65%. All data validation citations were correctly classified, suggesting that data validation citations have characteristic features. Although a lower success was found for the definition and method citations, it was found that many authors preferred to define their methods using other people's definitions. Therefore, the majority of the complexity came from the definitions when explaining the methods. In terms of the general framework, the application of the taxonomic classes categorized as the main classes had considerably higher success than when the analysis was applied over all classes.

Citation categorization for shape

When the citations in the Turkish library and information science literature were classified for shape, it was found that 63% of the citations had no determiner, and the most common types were citations that mentioned the authors' names. Previous studies have found that the most valuable citation types by shape mention author names and quotations (Bonzi 1982, p. 211; Zhu et al. 2015, p. 413). A study evaluating international LIS literature (Bonzi 1982, p. 212) confirmed that citations on author's names are the most common citation class, similar to our study.

Character n-gram pre-processing was performed to classify the citations by shape as letters and signs are important for this class (e.g., quotation marks, apostrophes), with the results of the analysis confirming this decision. While the performance rate of the analysis with 1–2 g word pre-processing was 69% (Naïve Bayes Multinomial), the success rate for the same algorithm increased to 83% when the 1–2 g character pre-processing was conducted, demonstrating that the character gram technique can give more accurate results when classifying citations according to shape.

The Random Forest algorithm, which was successful in classifying the positive and negative citations, was also able to achieve a high performance (92%, $f = 0.922$). With this algorithm, the rate of success for the citation classification and especially for author name citations increased significantly, with the f-measure values varying between 0.797 and 0.944, thereby confirming the algorithm's success. The lowest achieving algorithm performance class was for the quoted citations, possibly because the quoted citations in the texts were made by changing the text format rather than using quotation marks (such as starting from the inside of the paragraph, using smaller fonts, italics etc.). Further, as quotation marks can also be used for other purposes, they were expected to have a lower performance and lower f-measure values when determining the citations that contained quotation marks.

There are few studies that have classified citations by shape. In one important study, Athar (2014, pp. 84–86) calculated the f-measure as 0.446 for the determination of citations that contained author names using the Naïve Bayes and Support Vector Machines, which was quite low compared to the performance achieved in this study.

Categorization of citations for array

Two different analyses were conducted to classify citations for array; citation sections and the number of uses of the citations in the texts.

Citation sections

The results for the evaluation of citations by array are analyzed and visualized in Fig. 4. When the distribution of Turkish citations within the IMRAD structure was examined, it was found that the authors preferred to cite in the introduction section when the literature evaluations were also included; both operators tagged 84–85% citations in this section. The second most-cited section was research and the proportion of citations made in other sections showed a similar distribution at a high level.

As shown in Fig. 5, significant similarities were found between the operators. Therefore, it was possible to generalize findings for all LIS literature in Turkey. In the following, each is examined in detail:

- *Distribution of citations into sections by meaning* It was found that negative citations were mostly made in the research and discussion sections. The first operator did not tag any negative citations in the methods and other (footnote, acknowledgment, and appendix) sections. This gives some idea as to where negative citations can be found in scientific texts. Positive citations were mostly concentrated in the discussion section; however, they were also found in all sections.
- *Distribution of citations by purpose* The most prominent classes for citations by purpose were found in the methods, description, and data validation sections. Citations that did not differ between sections were found for the literature and data citations,

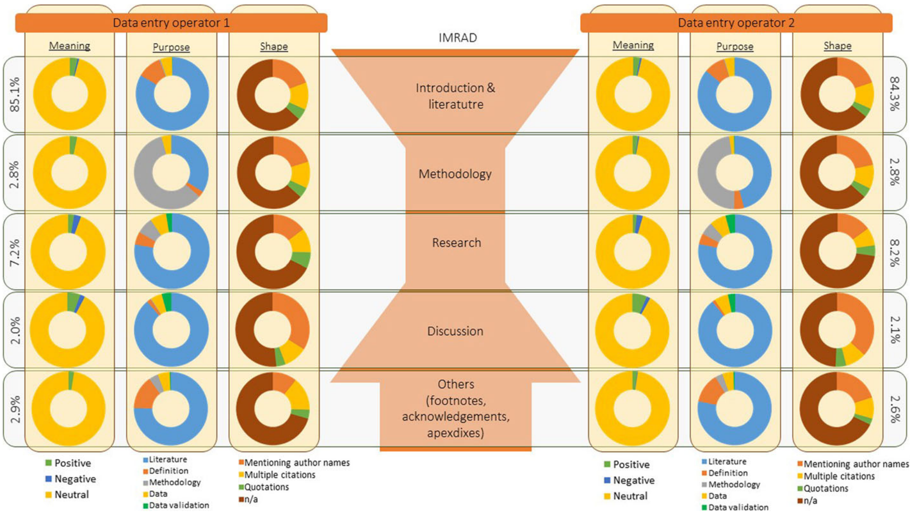

Fig. 5 Distribution of citations according to array in the IMRAD sections

which were found in almost all parts of the studies. While method citations were most often made in the methodology section as expected, definition citations were found in the introduction and in the footnotes. Data validation citations were found primarily in the research and discussion sections as these citations are generally a verification of past work to prove validity.

- *Distribution of citations by shape* Although there were no significant differences found for the distribution of citations by shape, author names were found to be more frequent in the research and discussion sections. Apart from these, there was no significant link found between the citation classes and the sections.

Various studies have been conducted on IMRAD sections and citations. A study on LIS literature (Ding et al. 2013, p. 583) found positive citations primarily in the introduction and literature sections. One of the earliest works that proved that not all citations are equal (Voos and Dagaev 1976) found that most citations were in the introduction section. Maricic et al. (1998, p. 539) claimed that citations in the methodology, research, and discussion sections were more meaningful than citations in the introduction and literature sections. Suppe (1998, p. 403) also argued that the methodology and research sections contributed the most as knowledge of new contributions are most often discussed in these sections. When examining the distribution of citations in terms of purpose and meaning in this study, a similar interpretation was made. Positive citations were found in almost all parts of the articles, while negative citations were most often found in the research and discussion sections.

Evaluation of the number of citations in the texts and the number of citations in the different sections

Significant results were achieved when the citation frequencies were evaluated.[3] While 67.5% of citations were mentioned only once in the text, surprisingly, 6.1% of the references were never mentioned. It was also observed that 1.1% of citations were not listed in references. The detailed frequencies are shown in Fig. 6, from which it can be seen that the probability of a reference being used more than once was about 30%.

From the results, 96% of citations were found in only one IMRAD section, 3% appeared in two sections and the remaining 1% in 3 or 4 sections. Only six citations were cited in all IMRAD sections.

Discussion and conclusion

Citations are used to provide a link between related articles. However, over the years, this purpose has changed and now citations are being used as a criterion for research/er evaluations, with one of the most important criteria for the granting of tenure, incentives, or rewards being the number of recent citations. As a result, many unethical practices associated with the use of citations have appeared. Therefore, as the content-based citation analysis in this study was developed based on the hypothesis that not all citations are equal, it is argued that a more accurate approach when evaluating citations is to consider the citation content.

[3] Evaluations based on the tags made by data entry operator 1.

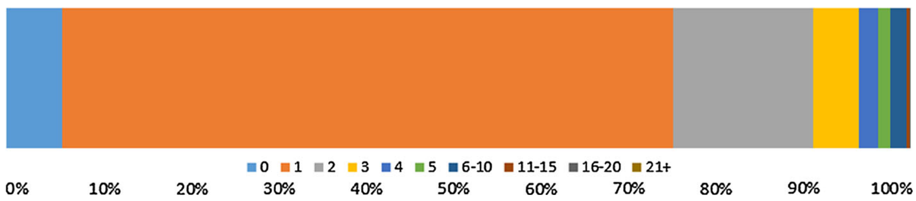

Fig. 6 Distribution of the number of use of the single references

This study proved that it was possible to develop a content-based citation analysis approach using the semantic and syntactic features in high-volume texts, along with the significant machine-learning achievements that can be obtained from text categorization algorithms. Therefore, a classification scheme was created for Turkish LIS citations, the taxonomic classification for which was quite similar to classifications reported in international studies. This study proved that citation motivations in Turkey are similar to those in international literature, which makes it easier to adapt this model to international literature with English-language corpora.

In this study, four basic taxonomic citation categories were identified for Turkish citations in the LIS field, and subgroups for these citation classes were established. In this context, it was thought that positive and negative citations were the most important types of citations as they carried the author intentions toward the cited articles. This study was an important step in evaluating citations by meaning.

It was also found in this study that another important class in content-based citation analysis was the identification of citations by purpose. The results demonstrated that the success rates for the machine-learning algorithms effectively classified these types of citations accurately, and could be easily adapted to other fields and languages.

The proposed content-based citation analysis approach is believed to be capable of improving the applications of the four roles involved in scholarly communication processes (see Fig. 7). These roles are those of the researchers, editors, database providers and policymakers. The researcher role includes not only professionals who work in scientific production but also those who are looking for tenures or awards. All these roles may benefit from content-based analysis for citations.

During the article writing process, researchers may benefit from content-based citation analyses by tracing references. As the number of publications increases on a daily basis, difficulty arises in accessing the most accurate publications and searching the literature that will serve as the research framework. For this reason, a well-structured literature search is needed. Content-based citation analysis enables access to the resources that are needed quickly. For instance, a researcher who searches for publications related to a specific method can easily access the required information by following method citations. In addition, researchers who search for tenures or awards can evaluate the “real” impact of their papers on academia and content-based citations provides feedback to them.

Citation database providers can benefit from this approach to enhance their services. A new generation citation analysis model can be developed for citation indexes. A sample model for applying this approach to citation indexes is shown in Fig. 8. In addition, this model presents an effective evaluation tool. Citation manipulations can be detected easily by analyzing distributions of citations to the classes. The performance evaluation processes of journals indexed on the Web of Science can be managed effectively by the model.

Fig. 7 Four roles that can benefit content-based citation analysis

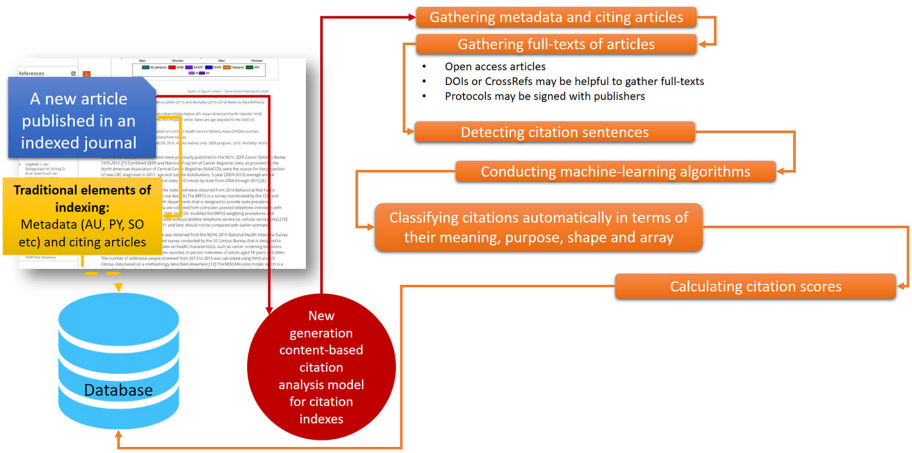

Fig. 8 Sample content-based analysis model for database providers

The most important control mechanism before the publication of a scientific journal is editorial control. For this reason, the editors have serious duties. Automated systems for citations can make facilitate editorial processes and make them manageable. Journals may adapt the content-based citation approach to their management systems and eliminate articles without citation quality with less human effort.

The most important issue that policymakers and managers must consider is that the citations are not just numbers. A criterion used to evaluate the performance of researchers is citations, but this criterion does not make any sense when used alone. One concern is that the processes can be employed in unethical practices. Thus, determining basic principles and policies is vital for objective and accurate evaluation. The content-based approach to publications and citations may enhance the quality of research outputs.

If all roles in the scholarly communication process are conscious of the differences in the citations, it is possible to see citations again as frozen footprints on the path to scientific knowledge. Along with developments in computational linguistics, many of the topics mentioned above can be easily resolved using automated methods. For this reason, it is of utmost importance that all actors in the scholarly communication process be aware of these techniques and, if necessary, they should collaborate with experts working in this field. These tasks can only be achieved through the collaboration of information scientists, linguists, and computer scientists.

Acknowledgements This article is based on Taşkın's (2017) Ph.D. dissertation and was supported in part by a research grant from the Turkish Scientific and Technological Research Center (115K440).

References

Akademik Teşvik Ödeneği Yönetmeliği [Academic Incentive Regulation], T.C. Resmi Gazete [Official Gazette]. (13271644, 27.12.2016). http://goo.gl/GTtGkZ.

Al, U., & Soydal, İ. (2012). Dergi kendine atıfının etkisi: Energy Education Science and Technology örneği [The impact of journal self-citation: The case of Energy Education Science and Technology]. *Türk Kütüphaneciliği [Turkish Librarianship], 26*(4), 699–714.

Al, U., & Soydal, İ. (2014). Akademinin atıf dizinleri ile savaşı [The war of academia with citation indexes]. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi [Hacettepe University Journal of Faculty of Letters], 31*(1), 23–42.

Al, U., & Soydal, İ. (2015). Bilimsel iletişimin farklı bir yüzü: Geri çekilen makaleler [The other face of scholarly communication: Retracted articles]. In U. Al & Z. Taşkın (Eds.), *Prof. Dr. İrfan Çakın'a Armağan* (pp. 22–37). Ankara: Hacettepe University.

Aljaber, B., Stokes, N., Bailey, J., & Pei, J. (2010). Document clustering of scientific texts using citation contexts. *Information Retrieval, 13*(2), 101–131.

Angrosh, M.A., Cranefield, S., & Stanger, N. (2010). Context identification of sentences in related work sections using a conditional random field: Towards intelligent digital libraries. In *Proceedings of the ACM, JCDL'10* (pp. 293–302). Queensland: ACM.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics, 34*(4), 555–596.

Arunachalam, S., & Manorama, K. (1988). Are citation-based quantitative techniques adequate for measuring science on the periphery? *Scientometrics, 15*(5–6), 393–408.

Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In *HLT-SS '11 Proceedings of the ACL 2011 student session* (pp. 81–87). Stroudsburg: Association for Computational Linguistics.

Athar, A. (2014). *Sentiment analysis of scientific citations (Technical report, UCAM-CL-TR-856)*. Cambridge: University of Cambridge Computer Laboratory.

Bertin, M. (2008). Categorizations and annotations of citation in research evaluation. 13. Natural language processing; 13.1 discourse. http://aaaipres.org/Papers/FLAIRS/2008/FLAIRS08-108.pdf.

Blake, C. (2013). Text mining. *Annual Review of Information Science and Technology, 45*(1), 121–125.

Bonzi, S. (1982). Characteristics of a literature as predictors of relatedness between cited and citing works. *Journal of the American Society for Information Science, 33*(4), 208–216.

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation, 64*(1), 45–80.

Brooks, A. T. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science, 37*(1), 34–36.

Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science, 40*(4), 284–290.

Carter, G.M. (1974). Peer review, citations, and biomedical research policy: NIH grants to Medical School Faculty. *Rand Report, R-1583*. Santa Monica: Rand. http://www.rand.org/content/dam/rand/pubs/reports/2008/R1583.pdf.
Cavalcanti, D.C., Prudêncio, R.B.C., Pradhan, S.S., Shah, J.Y. & Pietrobon, R.S. (2011). Good to be bad? Distinguishing between positive and negative citations in scientific impact. In *23rd IEEE international conference on tools with artificial intelligence (ICTAI)* (pp. 156–162). Boca Raton: IEEE.
Chubin, D. E. (1980). Letter to editor: Is citation analysis a legitimate evaluation tool? *Scientometrics, 2*(1), 91–94.
Cole, J. R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge a festschrift in honor of eugene garfield* (pp. 281–300). New Jersey: Information Today.
Cole, J. R., & Cole, S. (1971). Measuring the quality of sociological research: Problems in the use of the Science Citation Index. *The American Sociologist, 6,* 23–29.
Cole, J. R., & Cole, S. (1972). The ortega hypothesis: Citation analysis suggests that only a few scientists contribute to scientific progress. *Science, 178*(4059), 368–375.
COPE. (2012). *Citation manipulation*. http://publicationethics.org/files/u7141/Forum%20discussion%20topic_final.pdf.
Cozzens, S. E. (1985). Comparing the sciences: Citation context analysis of paper from neuropharmacology and the sociology of science. *Social Studies of Science, 15,* 127–153.
Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation, 37*(1), 16–24.
Damashek, M. (1995). Gauging similarity with n-grams language independent categorization of text. *Science, 267*(5199), 843–848.
Davis, P. (2017). Citation cartel or editor gone rogue? [BlogPost]. Scholarly Kitchen. http://scholarlykitchen.sspnet.org/2017/03/09/citation-cartel-or-editor-gone-rogue/.
Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics, 7,* 583–592.
Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology, 65*(9), 1820–1833.
Dong, C., & Schäfer, U. (2011). Ensemble-style self-training on citation classification. In *5th international joint conference on natural language processing, IJCNLP 2011* (pp. 623–631). Chiang Mai: AFNLP.
Elkiss, E., Shen, S., Fader, A., Erkan, G., States, D., & Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology, 59*(1), 51–62.
Fu, L. D., & Aliferis, C. F. (2010). Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. *Scientometrics, 85,* 257–270.
Garfield, E. (1970). Can citation indexing be automated? *Essays of an Information Scientists, 1,* 84–90.
Garfield, E. (1973). Citation frequency as a measure of research activity and performance. *Essay of an Information Scientist, 1,* 406–408.
Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359–375.
Goudsmith, S. A. (1974). Citation analysis. *Science, 183*(4120), 28.
Hacettepe Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi Uygulama Esasları ve Araştırmacı Bilgilendirme Kılavuzu [Implementation Guideline of Hacettepe University Scientific Research Projects Coordination Unit and Information for Researchers]. (2015). http://apsis.hacettepe.edu.tr/GetDoc.ashx?ID=19.
Halevi, G., & Bar-Ilan, J. (2016). Post retraction citations in context. In G. Cabanac, M. K. Chandrasekaran, I. Frommholz, K. Jaidka, M. Y. Kan, P. Mayr, & D. Wolfram (Eds.), *BIRNDL 2016 bibliometric-enhanced information retrieval and natural language processing for digital libraries* (pp. 23–29). Newark: CEUR.
Herlach, G. (1978). Can retrieval of information from citation indexes be simplified?: Multiple mention of a reference as a characteristic of the link between cited and citing article. *Journal of the American Society for Information Science, 29*(6), 308–310.
Jha, R., Jbara, A.-A., Qazvinian, V., & Radev, D. R. (2016). NLP-Driven citation analysis for scientometrics. *Natural Language Engineering, 23*(1), 93–130.
Johnson, C. A. (1985). Citations to authority in supreme court opinions. *Law and Policy, 7*(4), 509–523.
Kaplan, A. (2013). Üniversitelerde bilimsel yayın çalışmaları [Studies on scientific publications in universities] [Presentation]. Bilimsel Dergilerimiz ve Uluslararası İndekslerdeki Yeri Çalıştayı [Workshop on Our Scientific Journals and Their Roles on International Indexes]. http://webftp.gazi.edu.tr/bidb/dergilercalistayi/sunular/ULAKBIM_Dergi_sunu.pptx.

Kaplan, P. (2014). Akademisyenlerin atıf çetesi [Citation gang of academics]. HaberTürk [news]. http://www.haberturk.com/yazarlar/pervin-kaplan/974474-akademisyenlerin-atif-cetesi.
Kochen, M. (1974). *Principles of information retrieval*. Los Angeles: Melville.
Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI'95 proceedings of the 14th international joint conference on artificial intelligence* (pp. 1137–1143). Montreal: ACM.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
Lerner, J., & Wulf, J. (2007). Innovation and incentives: Evidence from corporate R&D. *The Review of Economics and Statistics, 89*(4), 634–644.
Liu, S., Chen, C., Ding, K., Wang, B., Xu, K., & Lin, Y. (2014). Literature retrieval based on citation context. *Scientometrics, 101*(2), 1293–1307.
Liu, Y., Yan, R. & Yan, H. (2013). Guess what you will cite: Personalized citation recommendation based on users' preference. In R. E. Banchs, F. Silvestri, T.-Y. Liu, M. Zhang, S. Gao, & J. Lang (Eds.) *Information retrieval technology: 9th Asia information retrieval societies conference, AIRS 2013 Singapore, December 2013 Proceedings* (pp. 428–239). Heidelberg: Springer.
MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435–444.
Maričić, S., Spaventi, J., Pavičić, L., & Pifat-Mrzljak, G. (1998). Citation context versus the frequency counts of citation histories. *Journal of the American Society for Information Science, 49*(6), 530–540.
Markey, K. & Cochrane, P.A. (1981). Online training and practice manual for ERIC database searchers. New York: ERIC Clearing House on Information Sciences. http://files.eric.ed.gov/fulltext/ED212296.pdf.
Matthew 25:29. (2004). http://biblehub.com/tur/matthew/25.htm.
Merton, R. K. (1968). The Matthew Effect in science. *Science, 159,* 56–63.
Miller, J. C., Coble, K. H., & Lusk, J. L. (2013). Evaluating top faculty researchers and the incentives that motivate them. *Scientometrics, 97,* 519–533.
Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science, 5,* 86–92.
Öğretim Üyeliğine Yükseltilme ve Atanma Yönetmeliği [Regulation on Promotion and Appointment to Instructional Membership]. (1982, 28 January). Resmi Gazete [Official Gazette] (Number: 17588). http://www.yok.gov.tr/web/guest/icerik/-/journal_content/56_INSTANCE_rEHF8BIsfYRx/10279/17641.
Oppenheim, C. (1996). Do citations count? Citation indexing and the research assessment exercise (RAE). *Serials, 9*(2), 155–161.
Oransky, I. (2017). Citation-boosting episode leads to editors' resignations, university investigation [BlogPost]. Retraction Watch. http://retractionwatch.com/2017/03/03/citation-boosting-episode-leads-editors-resignations-university-investigation/.
Öztürk, K. (2012). Şişme dergiler ve etik ihlalleri [Bloated journals and ethical violations] [BlogPost]. http://mkoz.wordpres.com/2012/06/23/sisme-dergiler-ve-etik-ihlalleri/.
Öztürk, K. (2013). Şişme dergiler, yeniden [Bloated journals, again] [Blogpost]. http://mkoz.wordpres.com/2013/01/25/sisme-dergiler-yeniden/.
Price, D. J. D. S. (1986). *Little science, big science… and beyond*. New York: Columbia University Press.
Ritchie, A. (2008). Citation context analysis for information retrieval. Ph.D. Dissertation, University of Cambridge.
Schneider, J.W. & Borlund, P. (2005). A bibliometric-based semi-automatic approach to identification of candidate thesaurus terms: Parsing and filtering of noun phrases from citation contexts. In F. Crestani & I. Ruthven (Eds.) *Context: Nature, impact and role, 5th international conference on conceptions of library and information sciences, CoLIS 2005 Glasgow, UK, June 2005, Proceedings* (pp. 226–237). Heidelberg: Springer.
Sendhilkumar, S., Elakkiya, E., & Mahalakshmi, G.S. (2013). Citation semantic based approaches to identify article quality. In D. C. Wyld (Ed.), *Computer science and information technology (CS & IT)* (pp. 411–420). Delhi: ICCSEA.
Shum, S. B. (1998). Evolving the web for scientific knowledge: First step towards an "HCI knowledge web". *Interfaces, British HCI Group Magazine, 39,* 16–21.
Silva, J. A. T., & Dobránszki, J. (2017). Highly cited retracted papers. *Scientometrics, 110*(3), 1653–1661.
Simkin, M. V., & Roychowdhury, V. P. (2003). Read before you cite! *Complex Systems, 14,* 269–274.
Simkin, M. V., & Roychowdhury, V. P. (2006). Do you sincerely want to be cited? Or: Read before you cite. *Significance, 3*(4), 179–181.
Smith, L. C. (1981). Citation analysis. *Library Trends, 30,* 83–106.

Spiegel-Rösing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science, 7*(1), 97–113.
Stigler, S. M. (1980). Stigler's law of eponymy. *Transactions. New York Academy of Sciences, 39*(1), 147–157.
Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science, 65*(3), 381–405.
Tandon, N. & Jain, A. (2012). Citation context sentiment analysis for structured summarization of research papers. http://people.mpi-inf.mpg.de/~ntandon/papers/ki2012-tandon.pdf.
Taşkın, Z. (2017). *İçerik tabanlı atıf analizi modeli tasarımı: Türkçe atıflar için metin kategorizasyonuna dayalı bir uygulama (Designing a model for content-based citation analysis: an application for Turkish citations based on text categorization)*. Unpublished Ph.D. Dissertation, Hacettepe University.
Testa, J. (2008). Regional content expansion update: Web of Science 5.0. http://wokinfo.com/media/pdf/RExEssay.pdf.
Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific text.* Unpublished Ph.D. Dissertation, University of Edinburg.
Teufel, S., Siddharthan, A. & Tidhar, D. (2006). Automatic classification of citation function. http://www.cl.cam.ac.uk/~sht25/papers/emnlp06.pdf.
Title Suppressions. (2016). http://ipscience-help.thomsonreuters.com/incitesLiveJCR/JCRGroup/titleSuppressions.html.
Tonta, Y. (2014). *Akademik performans, öğretim üyeliğine yükseltme ve yayın destekleme ölçütleriyle ilgili bir değerlendirme [An evaluation of criteria on academic performance, tenure and publication support].* http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tonta-yukseltme-kriterleri-hakkinda-degerlendirme-11-Temmuz-2014.pdf.
Van Raan, A.F.J. (2004). Measuring science: Capita selecta of current main issues. In H. F. Moed, W. Glänzel, & U. Schmoch, (Eds.) *Handbook of quantitative science and technology research* (pp. 15–50). Dordrecht: Kluwer Academic.
Van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), 133–143.
Vinkler, P. (1994). Words and indicators: As scientometrics stands. *Scientometrics, 30*(2), 495–504.
Voos, H., & Dagaev, K. S. (1976). Are all citations equal? Or did we Op. Cit. your Idem? *The Journal of Academic Librarianship, 1*(6), 19–21.
Wetterer, J. K. (2006). Quotation error, citation copying, and ant extinctions in Madeira. *Scientometrics, 67*(3), 351–372.
Woolgar, S. (1991). Beyond the citation debate: Towards a sociology of measurement technologies and their use in science policy. *Science and Public Policy, 18*(5), 319–326.
Xu, J., Zhang, Y., Wu, Y., Wang, J., Dong, X., & Xu, H. (2015). Citation sentiment analysis in clinical trial papers. *AMIA Annual Symposium Proceedings, 2015,* 1334–1341.
Yu, B. (2013). Automated citation sentiment analysis: What can we learn from biomedical researchers? ASIS&T 2013 Annual Meeting Montréal, Quebec, Canada, November 1–5, 2013. http://www.asis.org/asist2013/proceedings/submissions/papers/105paper.pdf.
Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology, 66*(2), 408–427.
Ziman, J. M. (1968). *Public knowledge: An essay concerning the social dimension of science*. Cambridge: Cambridge University Press.
Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison Wesley Pres.