Context Identification of Sentences in Related Work Sections using Conditional Random Fields: Towards Intelligent Digital Libraries

Angrosh M.A. Stephen Cranefield Nigel Stanger

Department of Information Sciences, University of Otago, Dunedin, New Zealand

{angrosh, scranefield, NStanger}@infoscience.otago.ac.nz

ABSTRACT

Identification of contexts associated with sentences is becoming increasingly necessary for developing intelligent information retrieval systems. This article describes a supervised learning mechanism employing conditional random fields (CRFs) for context identification and sentence classification. Specifically, we focus on sentences in related work sections in research articles. Based on the generic rhetorical pattern, a framework for modelling the sequential flow in these sections is proposed. Adopting a generalization strategy, each of these sentences is transformed into a set of features, which forms our dataset. Prominently, we distinguish between two kinds of features for each of these sentences viz., citation features and sentence features. While an overall accuracy of 96.51% is achieved by using a combination of both citation and sentence features, the use of sentence features alone yields an accuracy of 93.22%. The results also show F-Scores ranging from 0.99 to 0.90 for various classes indicating the robustness of our application.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Content Analysis and Indexing, Information Search and Retrieval, Digital Libraries

General Terms

Algorithms, Experimentation

Keywords

Sentence Classification, Citation Classification, Conditional Random Fields

1. INTRODUCTION

In recent times, the idea of using scientific knowledge to develop knowledge-based infrastructures for effective information dissemination has been heavily researched. The emergence of semantic and collaborative web technologies has further fuelled this process. These initiatives are primarily motivated by the increasing need for digital libraries capable of answering fundamental queries related to critical enquiry such as information about supporting and challenging documents for a given document; intellectual lineage of an idea; available data for supporting of an idea; available data for supporting a specific claim or an idea; information about others working on the same problem; and the applicability of similar approaches in other fields. These efforts are centred on filling a larger gap noted in the researcher's digital toolkit – the lack of tools that track contributions in a field and express, analyze, and contest their significance [1]. In essence, it involves establishing links between research articles and researchers and making use of the representation for answering the above queries.

Concomitantly, citation analysis has been extensively studied for establishing links between researchers and research articles. While bibliometric measures based on citations such as the Impact Factor are used to measure the impact of a researcher's work by how often they are cited [2, 3], these measures have been criticised for being purely quantitative and that many citations are done out of "politeness, policy and piety" [4]. At the same time, it is also noted that critical citations or citations in passing should not "count" as much as central citations in a paper or as those citations where a researcher's work is used as the starting point of somebody else's work [5]. In other words, citation analysis is expected to look beyond quantitative analysis and aim at providing finer representation of relations between researchers and articles. Such a representation could also effectively answer the above identified questions. However, the identification of the exact relationship between the citing paper and the cited work is a major challenge.

The finer representation of relations between articles could be used for developing intelligent digital libraries, offering various value-added information services. Among others, this could include search for identification of works used to progress or develop newer ideas, identifying interpretations of a given work by different authors in different contexts, improvement of Impact Factor calculations, identification of research gaps, identifying articles criticising a given work, and tools for tracing the intellectual lineage of a given idea.

Against this objective, the present study is taken up for identifying the context associated with sentences in research articles. To this end, we perceive context identification as a sequential classification problem and employ conditional random fields (CRFs) for classifying sentences into predefined categories. We show that the presence of citations is an important feature in modelling the sequential pattern in research writing. The paper is structured as follows: Section 2 describes the related work in the field of classifying citations. While Section 3 draws upon the essence of CRFs, Section 4 describes our approach for modelling sentences for classification using CRFs. Section 5 describes the experiments carried out and Section 6 puts forward the results achieved. We conclude this paper in Section 7.

2. RELATED WORK

Citations form an integral part of research writing. While laying the foundation of academia, citations provide for referencing other works, i.e., works that are already known, established or conceptual thought, for progressing one's ideas. These form the evidence and authors reason in their own way about this evidence in order to contribute to the domain of knowledge. Apart from providing valuable bibliometric measures, citations have been keenly studied with a focus on identifying the reasons for citations and more recently, experiments on the automated classification and identification of contexts associated with citations have been undertaken. As early as 1971, Weinstock put forward as many as 15 categories describing the different reasons for citations [6]. The first approach towards a formal classification of citations was proposed by Moravcsik and Murugesan [7]. This classification allowed for multiple attributes implying that for a single complex citation function, four values could be in use. The four categories were conceptual / operational; organic / perfunctory; evolutionary / juxtapositional and confirmative / negational. Nanba and Okumara summed up the various reasons for citations as identified by Weinstock into three fundamental categories for automatic generation of review articles [8]. The first category was type B, which included citations that point out other researchers' theories or methods for the theoretical basis. The second category of citations, type C, pointed out the problems or gaps in related work and the third category distinguished as type O referred to citations other than types B and C.

Teufel et al. developed an annotation scheme of seven categories for building automatic abstracting systems for scientific research articles [9]. The categories were based on rhetorical moves of argumentation within the paper. In principle, the seven categories included background, others, own, aim, textual, contrast and basis. Further, Teufel and Moens noted a common rhetorical pattern of scientific argumentation in the introduction section of research articles [10]. It was observed that there exists a background segment, which discusses the history and the importance of the task and is usually followed by a longer sequence of other sentences, which describe specific prior work in a neutral way. This discussion usually terminates in a criticism of the prior work, thus giving a motivation for the work presented in the paper. Garzone and Mercer presented a system for citation classification that relied on characteristic syntactic structure to determine citation category [11]. Mercer and Marco extended this idea to use fine-grained cue phrases within citation sentences as a stylistic basis for categorization [12]. More recently, Le et al employed Maximum Entropy Markov Models (MEMMs) for classifying citation sentences [13]. With a focus on detecting emerging trends, the authors proposed a classification scheme comprising six categories for classifying citation types. Based on findings from Swales [14] that scientific argument follows a general rhetorical structure, Teufel et al. introduced a citation annotation scheme involving 12 categories [15].

While most of the aforementioned studies limited their focus to classifying citation sentences, the present study, besides classifying citations, also focuses on classifying sentences adjacent to citations. While the final objective of this study is to classify citations appearing throughout the article and use the classified data for developing intelligent information systems, we present here the results obtained from our experiments specifically carried out on related work sections. The presence of a large number of citations in the related work section is the primary motivation for choosing this section. Furthermore, we view the classification task as a sequential classification problem and assume that authors follow a sequential rhetorical pattern while drawing upon related work. The sequential classification is achieved by using conditional random fields (CRFs) – a probabilistic framework for labelling sequential data based on supervised learning. Recently, CRFs have been successfully applied to various tasks such as Part of Speech (POS) tagging [16]; named entity recognition [17]; table extraction from government reports [18] and noun phrase segmentation [19].

3. CONDITIONAL RANDOM FIELDS

Conditional Random Fields (CRFs) [16] are undirected graphical models used for computing the conditional probability of values on designated output nodes given feature values assigned to other designated input nodes. CRFs offer several advantages over other sequential models such as Hidden Markov Models (HMMs). As discriminative models, CRFs do not model interdependence among observed data nor do they impose independence assumptions on the observations. The framework allows rich and unconstrained feature representation that could overlap or refer arbitrarily to the observation. Higher performance is achieved with CRFs as they are normalized over the full sequence, overcoming the “label-bias” problem, a weakness found in maximum entropy Markov models (MEMMs) [20] and other conditional Markov models based on directed graphical models.

Let x = $x_1....x_T$ be an input sequence and y = $y_1....y_T$ be a corresponding state (or label) sequence. Generally it is assumed that the dependencies of y, the state sequence, conditioned over x form a linear chain i.e., each state depends only on its predecessor. A linear chain CRF with parameters $\Lambda = \{\lambda_1,....\}$ defines a conditional probability for **y** given x to be

$$P_\Lambda(y|x) = \frac{1}{Zx} \exp\left(\sum_{t=1}^{T} \sum_k \lambda_k \ f_k\ (y_{t-1}, y_t, \mathrm{x}, \mathrm{t})\right) \quad (1)$$

where Z(x) is a normalisation factor that makes the probabilities of all label sequences sum to one, $f_k\ (y_{t-1}, y_t, \mathrm{x}, \mathrm{t})$ is a feature function and λ_k is a learned weight associated with feature f_k .

A feature function measures the state transition $y_{t-1} \rightarrow y_t$ and the observation sequence, x centered at the current time step *t*. For example, as in our case, a feature function may take value 1 if the current state y_t is RWD (sentence Describing Related Work) and the previous state y_{t-1} is RWD_CS (Citation Sentence describing Related Work) and the observation x_t, is a Description Term, or a value 0 otherwise.

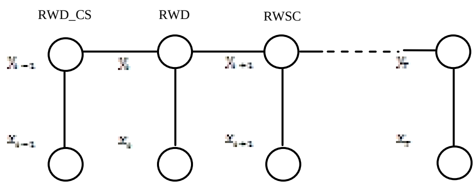

Figure 1: Graphical model for linear chain structure CRF

Inference in CRFs is done by finding the most probable label sequence y^*, for an input sequence x, given in model (1):

$$y^* = \operatorname{argmax} P_{\Lambda}(y|x)$$

This is calculated by dynamic programming using the Viterbi algorithm.

The learning task in a CRF is to choose values for the weights (also called parameters). Given a set of training examples, the goal is to choose parameter values $\Lambda = \{\lambda_1, ...\}$ that maximize the conditional probability of the training examples. Parameter estimation in CRF is carried out in an iterative procedure such as limited memory quasi-Newton method (L-BFGS), stochastic gradient method or voted perceptron.

4. MODELLING SENTENCES IN RELATED WORK SECTION FOR CRFs

4.1 Task and Approach

The related work section in a research article is primarily used by an author for drawing upon the related work in the research area and expressing his/her ideas in the context of the identified related work. A closer look at these sections reveals a common generic rhetorical pattern in these paragraphs. The section usually begins with a background sentence, which provides background information or an introduction to the related research area and continues to point out to a specific related work, which is of interest in the context of the author's idea. The latter, popularly known as citations, either simply describe the related work or refer to the outcome or strengths of a related work. The distinction between description and referring to an outcomes or strengths is purposely made for achieving a finer representation of relations between the cited work and the citing author.

The citations are usually observed to follow a set of sentences which further describe or point out the outcome or strengths of the related work. Following such sentences, the author generally points out shortcomings in the related work. Besides identifying shortcomings, there could also be sentences referring to contrasting work for the cited work. Pointing out the shortcomings is necessary for progressing the author's argument and provides a link with the successive sentences referring to newer cited works. After identifying shortcomings, the author usually proceeds to usually point out alternate works which overcome the identified shortcomings. These could also be citations or regular sentences which refer to alternate approaches for a given study.

Finally, in relation to the description or reference to an outcome or strengths of a cited work, its shortcomings, and alternate approaches, the authors describe the outcomes of the current work. The authors also sometimes mention shortcomings in the current work. It needs to be noted that it is not necessary that the above identified pattern is rigidly followed in a water-tight manner in these sections across all papers. The style varies across papers in accordance with the writing style of authors. The proposed framework is a generic pattern obtained after studying a series of related work sections in different papers. The framework is developed with an objective of using the structure for obtaining semantic data for developing intelligent information retrieval systems. Further, it should also be noted that the CRF-based tool used in this study is capable of learning other rhetorical structures.

The generic rhetorical pattern observed in the related work section is diagrammatically shown in Figure 2.

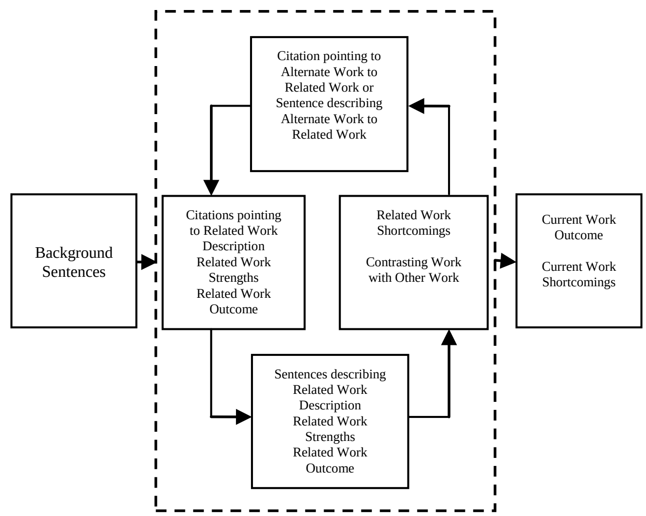

Figure 2: Rhetorical Pattern in Related Work Section

Based on the above framework, we define a classification scheme for classifying sentences in related work sections. Broadly, the classes can be divided into the following six categories:

1. Background – Sentences providing background in the research area or the introduction to an idea
2. Citation Sentences – Citations which either describe or refer to an outcome or strength of a cited work
3. Descriptive Sentences – Sentences that follow citations which further describe or refer to an outcome or strength of a cited work
4. Research Gap Sentences – Sentences that describe the shortcomings in a cited work
5. Alternate Approach Sentences – Sentences referring to alternate approaches for overcoming the shortcomings identified in the related work
6. Current Work Sentences – Sentences referring to outcomes of the current work or shortcomings of the current work.

Based on the above categorization, we define 13 different classes (labels) into which the sentences in the related work sections can be classified. **Table 1** lists the 13 different categories defined under the schema.

4.2 Feature Selection

The most important aspect of specifying the model is selecting a set of features that capture the relationship between each sentence and the label sequences. We distinguish between two kinds of features for each sentence viz., citation features and sentence features. While citation features denote the presence of a citation in sentences, sentence features are defined based on a generalization strategy adopted in the study. The generalization strategy categorizes certain kinds of words and these categories are used as features for each sentence. The following section details the adopted generalization method and the process of defining features based on it.

4.2.1 Generalization of Terms

In order to classify sentences into different contexts, we follow a generalization process, which identifies the presence of certain kinds of words which signify the context of a given sentence. This involves identifying the following nine categories for terms used by authors in research writing.

1. Inquiry Terms (IT Terms) – These are primarily the reporting verbs used in academic writing. The reporting verbs mainly define the degree of inquiry into a cited work. These verbs are also used to describe the kind of enquiry being made into a cited work. Examples of inquiry terms include *examine, propose, state*.

2. Outcome Terms (OCT Terms) - These terms also appear in the form of reporting verbs and are mainly used refer to the outcomes of the cited works. Examples of these terms include *show, develop*.

3. Strength Terms (STH Terms) – Again, these terms are reporting verbs but mainly describe the strengths of cited works. Examples of these terms include *improve, better performance, aids*.

4. Shortcoming Terms in Related Work (SCT) - These are the terms generally used by authors to denote shortcomings. The shortcomings can be identified in either the cited work or the current work. Examples of SCT terms include *Nevertheless, do not, however, but*.

5. Subjective Pronouns (SPN Terms) – Subjective pronouns are primarily terms that are used as the subject of a sentence. These terms are used by authors to refer to cited works. Prominently, these terms appear in the Descriptive Sentences following citations. These form an important feature and provides for connecting the descriptive sentences to citation sentences. The presence of these features indicates that the author is continuing to describe the related work. Examples of these terms include *They, The authors, In*

6. Words of Stress (WOS Terms) - These terms are generally used to stress the importance of a point made by the authors. WOS Terms is an important feature and aids in identifying specific types of sentences. For instance, use of words of stress by an author indicates emphasis on a specific work and is an indicator to classify the sentence as a 'Strength'. Examples of such terms include *Moreover, In addition, Therefore.*

7. Alternate Approach Terms (AAT Terms) - These are the terms used by the author for referring to alternate approaches for overcoming the shortcomings identified in a cited work. Examples of these terms include *may be*, *Another approach, alternative,*

8. Result Terms (RES Terms) - These are terms that mainly describe the results of the current work. Examples of these terms include *we have shown, our work shows, this paper*.

9. Contrasting Terms (CON Terms) – These are terms that are used to make a contrasting statement for a given work. For example, phrases such as *In Contrast*, our work *differs from that of...* are strong indicators of contrasting terms.

Table 2 sums up the different categories defined for classifying terms in sentences

4.2.2 Feature Definition

Based on the above categorization, we define 'sentence features' for each sentence in the related work section. Identification of terms belonging to different categories results in a corresponding feature for each sentence. We also define a feature for presence of a citation in a sentence. Table 3 lists the different features that could be defined for a sentence and a subset of these features would form the feature set for a given sentence in our experiment.

The process of term identification and feature definition for each sentence is automated through development of Python modules. We define certain conditions and regular expressions for identification of features in sentences. While most of these features are defined through matching of terms belonging to different categories (Table 2) using regular expressions, we lay down some conditions for defining certain features for terms belonging to the specific category of 'Inquiry Terms'. It was observed in our study that terms belonging to the category of 'Inquiry' appeared more commonly in citation sentences describing related work, sentences describing related work and background sentences. This led to a dual problem. While it was difficult to distinguish between the subject of inquiry terms and descriptive terms in these sentences, it also posed difficulties in maintaining the increasing list of background terms. In order to solve this problem, we defined the following condition for indentifying features in sentences having inquiry terms:

Condition	Feature
Sentence has citation and inquiry term	sentHasTerm=SOI
Sentence has no citation and has inquiry term	sentHasTerm=DES
Sentence has no citation and appears in the beginning of a paragraph	sentHasTerm=BGC

Sentence has no citation and previous two sentences does not have citation	sentHasTerm=BGC
Sentence has citation and inquiry term and subjective pronoun term	sentHasTerm=SOI sentHasTerm=SPN
Sentence has no citation and inquiry term and subjective pronoun term	sentHasTerm=DES sentHasTerm=SPN
Sentence has citation and inquiry term and shortcoming term	sentHasTerm=SOI sentHasTerm=SCT
Sentence has no citation and inquiry term and shortcoming term	sentHasTerm=DES sentHasTerm=SCT

The other features are defined based on the presence of terms belonging to different categories identified in the experiment. Table 4 lists the various conditions for defining features for each sentence.

Apart from 'sentence features', we also define 'citation features' which aim at identifying whether a given sentence is a citation or not. The 'citation features' primarily facilitate in modelling the sequential flow between sentences. To start with, we check whether a given sentence is a citation or not. To this end, we look for presence of terms such as *[x] [xx]*, *et, al.*, in a sentence, which are used for citation purposes[1]. Accordingly, a feature such as 'sentHasCitation' is added to a sentence. Further, we also add a feature for indicating whether the previous sentence is a citation or not. Once we have identified the status of a given sentence, with respect to its citation, we check for terms categorized in Table 2 and add corresponding features to the sentence.

Based on the categorization of Terms defined in Table 2 and features listed in Table 3, each of the sentences in the related work section is reduced to a set of features. Table 5 provides an example of feature sets obtained for sentences in each of the different classes defined in Table 1. These sentences come from the LNCS data set described in the next section.

5. EXPERIMENTS

5.1 Dataset

The dataset was developed from 50 research articles randomly selected from the Lecture Notes in Computer Science (LNCS) collection at springerlink.com [21]. The related work section in each of these articles was extracted for the purpose of this study. The training set of 50 research articles yielded 200 paragraphs which had a total of 1063 sentences. Each paragraph was represented as a sequence of sets of features with each sentence manually translated into a set of features based on the methodology described above.

Two different datasets were prepared from the training set. While the first dataset was developed using only sentence features, the second dataset was developed using both citation features and sentence features. The following illustrates the difference between two sets of features for a sample paragraph:

Sample Paragraph from Related Work Section

Several researchers have been studying the behavior of chains in MHWNs. Li et al. examine the performance of chains as the number of hops are increased and study the effect of cross-interference between chains [5]. They analyze the effect of MAC 802.11 behavior on the performance of multi-hop chains but do not categorize interference patterns that govern network performance in terms of throughput and bandwidth utilization. Ping et al. present a hop by hop analysis of a multi-hop chain, study the impact of hidden terminals on the throughput chains, and present a quantitative approach towards estimating this throughput [6]. They show that hidden terminals cause packet drops affecting chain throughput and causing route stability.

Source: Majeed et al. (2009) [22]

Sentence Features for the Sample Paragraph (one line per sentence)

```
sentHasTerm=BGT BG
sentHasTerm=SOI RWD_CS
sentHasTerm=SPN sentHasTerm=DES sentHasTerm=SCT
RWSC
sentHasTerm=SOI RWD_CS
sentHasTerm=SPN sentHasTerm=OCT RWO
```

Sentence and Citation Features for the Sample Paragraph (one line per sentence)

```
sentHasTerm=BGT BG
sentHasCitation sentHasTerm=SOI RWD_CS
prevSentHasCitation sentHasTerm=SPN
sentHasTerm=DES sentHasTerm=SCT RWSC
sentHasCitation sentHasTerm=SOI RWD_CS
prevSentHasCitation sentHasTerm=SPN
sentHasTerm=OCT RWO
```

A 10-fold cross validation was carried out on each dataset, for which the dataset was split to obtain 10 sets of individual test data and training data in the ratio of 9:1.

5.2 Training CRFs

In the present study, for training the CRF model, we used Mallet [23], a Java-based package that provides an implementation of linear chain CRF algorithms for working with sequential data.

6 RESULTS

In Table 7, we present the confusion matrix for evaluating the performance of the classifier on the dataset. We also report for each classification label, the precision, recall and the F-score. The F-score is computed as follows:

$$P = \frac{|TP|}{|TP| + |FP|};\ R = \frac{|TP|}{|TP| + |FN|};\ F = \frac{2\,PR}{P+R}$$

where P represents precision, R represents recall, TP is the set of true positive, TN is the set of true negatives, and FP is the set of false negatives. We also report the accuracy, defined as the percentage of correctly labelled sentences.

The results showing precision, recall and F-score for each of the classification labels are tabulated in Table 6. As may be seen from Table 6, the classifier achieves a higher accuracy of 96.51% when trained with the combination of both citation features and sentence features as against a lower accuracy of 93.22% obtained when trained with sentence features alone. This implies that the use of citation features for modelling sequential flow plays an

[1] The dataset used in this study was developed from the LNCS collection, in which these are the standard forms of citation. However, our regular expressions are defined to include other forms of citation.

important role in achieving higher accuracy. The use of combined features also helps in achieving a high F-score for various classes. While the classes of BG, RWD_CS, RWSC, RWD, CWO and RWS obtained a high F-score of 0.99, 0.98, 0.98, 0.97, 0.96 and 0.90 respectively through the use of combined features, the same classes achieved a lower F-score of 0.93, 0.97, 0.94, 0.92, 0.94 and 0.59 respectively by using sentence features alone. It is also observed that fewer classes obtained a lower F-score ranging from 0.72 to 0.24. This could be attributed to the fact that there are fewer occurrences of instances in these classes and could be considerably improved by developing additional training data. However, it needs to be noted that the use of both the features helped in achieving higher F-scores for these low-performing classes as against the use of sentence features alone.

The confusion matrix for the classifier is shown in Table 7. As may be seen from the confusion matrix, a large number of sentences fell in the categories of BG, RWD_CS, RWSC, RWD and CWO. 30% of the sentences fell in the category of RWD_CS, making it the highest class and RWSC and RWD followed RWD_CS with 23% and 20% respectively. Thus, it can be inferred that about 30% of the related work section generally comprises citation sentences describing the cited work, approximately 20% of sentences are non-citation sentences further describing the related work and 20 % of the sentences point out shortcomings in the cited works. The remaining 30 % is made up of other types of sentences as identified above.

It may also be noted from the matrix that the recall measure of certain classes varies from as low as 0.20 to 0.66. This is particularly evident for classes RWO_CS, RWS_CS, AWRW_CS, RWO and CWSC which achieved a recall measure of 0.20, 0.20, 025, 0.44 and 0.66 respectively. While this could be attributed to the occurrence of fewer instances in these classes, it should also be noticed that the classifier classifies these instances in closely related neighboring class. For example, while instances of RWO_CS is classified as RWO and RWS_CS as either RWS or RWD_CS, instances of AWRW_CS are classified as instances of AWRW and instances of class RWO are largely classified as RWD. The instances of CWSC are mainly classified as RWSC. The presence of features identifying the class characteristics forms a prominent reason for this classification. For example, the presence of features which identify outcome terms and alternate approach terms facilitates classification of instances in RWO and AWRW classes, instead of RWO_CS and AWRW_CS respectively. Similarly, features which represent shortcomings help classifying instances in RWSC instead of CWSC. Thus, the classifier selects closest neighbor in the classification and not abruptly choose a non-related class. This problem could be solved by appropriately increasing the number of training instances in classes which suffer from lower F-scores.

7 CONCLUSIONS AND FUTURE WORK

This paper presented our work on identifying contexts attached to sentences in related work sections of research articles and classifying them according to a sequential framework developed on the basis of a generic rhetorical pattern observed in these sections. We showed that the linear chain CRF fitted well with our application and facilitated in learning and applying the sequential pattern successfully. The study distinguished between two kinds of features: ‘citation features’ and ‘sentence features’. While an accuracy of 96.51% was achieved by using a combination of both citation and sentence features, the use of sentence features alone yielded an accuracy of 93.22%. We also obtain high F-scores ranging from 0.99 to 0.90 for various classes. Our future work includes the following. Taking note of the lower F-Score achieved for some classes, we intend to develop a more balanced data set, mainly focusing on adding more instances belonging to these classes. We also aim to use the classification data for developing semantic web based information retrieval systems capable of answering the fundamental questions of critical inquiry identified in the beginning of this paper.

Table 1: Labels for Sentences in Related Work Section

Sl. No.	Class	Label	Description
Background / Introduction			
1.	Background	BG	Background sentence describing background in the research area
Citation Sentences			
2.	Related Work Description – Citation Sentence	RWD_CS	Citation sentence describing the related work
3.	Related Work Outcome – Citation Sentence	RWO_CS	Citation sentence pointing out an outcome of the related work
4.	Related Work Strengths – Citation Sentence	RWS_CS	Citation sentence describing the strengths of the related work
Descriptive Sentences			
5.	Related Work Description	RWD	Sentence that describes the related work
6.	Related Work Outcome	RWO	Sentence that mention the outcomes of a related work
7.	Related Work Strength	RWS	Sentence that describe the strength of a related work
Research Gap Sentences			
8.	Related Work Shortcoming	RWSC	Sentence noting the shortcomings in the related work
9.	Contrasting Work for a Related Work	CWRW	Sentence describing contrasting work for a related work
Alternate Approaches			
10.	Alternate Work for a Related Work –	AWRW_CS	Citation sentence pointing out an alternate work for a related work

	Citation Sentence		
11.	Alternative Approach to a Related Work	AARW	Sentence that points out alternate approaches for a related work
Current Work			
12.	Current Work Outcome	CWO	Sentence describing the outcome of the current work
13.	Current Work Shortcoming	CWSC	Sentence describing the shortcomings in the current work

Table 2: Categories of Terms defined for Generalization of Sentences

Sl. No.	Category	Examples of Terms	Description
1.	Inquiry Terms (IT)	examine, propose, demonstrate	Reporting verbs which describe something
2.	Outcome (OCT)	develop, show	Reporting verbs that refer to an outcome
3.	Strength (STH)	Improve	Reporting verbs that refer to a strength
4.	Shortcomings (SCT)	Notwithstanding, does not, Despite this improvement	Terms that denote gaps in related work
5.	Subjective Pronouns (SPN)	They, The authors, In	Subjective terms that refer to related work
6.	Words of Stress (WOS)	Hence, Furthermore, Therefore, In addition	Terms used to stress a point of view
7.	Alternate Approaches (AAT)	may be, different, instead	Terms used to refer to alternate approaches
8.	Result Terms (RES)	We show, our work shows	Terms that refer to the outcome of the current work
9.	Contrast Terms (CON)	In Contrast, differs from that of	Terms used to make contrasting statements for a given work

Table 3: Various features defined for representing terms in sentences

Sl. No.	Feature	Description
Citation Features		
1.	sentHasCitation	Sentence has citation
2.	prevSentHasCitation	Previous sentence has Citation
Sentence Features		
3.	sentHasTerm=BGC	Sentence has a background context term
4.	sentHasTerm=SOI	Sentence has a subject of enquiry term
5.	sentHasTerm=SPN	Sentence has a subjective pronoun
6.	sentHasTerm=SCT	Sentence has a shortcoming term
7.	sentHasTerm=DES	Sentence has a descriptive term
8.	sentHasTerm=OCT	Sentence has an outcome term
9.	sentHasTerm=STH	Sentence has a strength term
10.	sentHasTerm=RES	Sentence has a term referring to current work outcome
11.	sentHasTerm=WOS	Sentence has a word of stress term
12.	sentHasTerm=AAT	Sentence has an alternate approach term
13.	sentHasTerm=CON	Sentence has a contrasting term

Table 4: Conditions for Adding Features for Sentences

Sl. No.	Term	Conditions for Adding Features			
		Sentence has Citation	**¬ Sentence Has Citation**	**¬ Sentence has citation and sentence is in the beginning of paragraph**	**¬ Sentence has citation and previous two sentences does not have citation**
1.	Inquiry Term	sentHasTerm=SOI	sentHasTerm=DES	sentHasTerm=BGC	sentHasTerm=BGC
2.	Inquiry Term and Subjective Pronoun Term	sentHasTerm=SPN sentHasTerm=SOI	sentHasTerm=SPN sentHasTerm=DES	-	-
3.	Inquiry Term and Shortcoming Term	sentHasTerm=SOI sentHasTerm=SCT	sentHasTerm=DES sentHasTerm=SCT	-	-
4.	Outcome Term	sentHasTerm=OCT	sentHasTerm=OCT	-	-
5.	Outcome Term	sentHasTerm=SPN	sentHasTerm=SPN	-	-

	and Subjective Pronoun Term	sentHasTerm=OCT	sentHasTerm=OCT		
6.	Outcome Term and Shortcoming Term	sentHasTerm=OCT sentHasTerm=SCT	sentHasTerm=OCT sentHasTerm=SCT		-
7.	Strength Term	sentHasTerm=STH	sentHasTerm=STH	-	-
8.	Strength Term and Subjective Pronoun Term	sentHasTerm=SPN sentHasTerm=STH	sentHasTerm=SPN sentHasTerm=STH	-	-
9.	Strength Term and Shortcoming Term	sentHasTerm=STH sentHasTerm=SCT	sentHasTerm=STH sentHasTerm=SCT	-	-
10.	Shortcoming Term	sentHasTerm=SCT	sentHasTerm=SCT	-	-
11.	Alternate Approach Term	sentHasTerm=AAT	sentHasTerm=AAT	-	-
12.	Alternate Approach Term and Shortcoming Term	sentHasTerm=AAT sentHasTerm=SCT	sentHasTerm=AAT sentHasTerm=SCT	-	-
13.	Result Term	sentHasTerm=RES	sentHasTerm=RES	-	-
14.	Result Term and Shortcoming Term	sentHasTerm=RES sentHasTerm=SCT	sentHasTerm=RES sentHasTerm=SCT	-	-
15.	Result Term and Contrast Term	sentHasTerm=RES sentHasTerm=CON	sentHasTerm=RES sentHasTerm=CON	-	-

Table 5: Example of Features for Different Sentences

Sl. No.	Sentence	Corresponding Set of Features	Class / Label
1.	Several researchers **have been studying** the behavior of chains in MHWNs.	sentHasTerm=BGC	BG – Background
2.	Li et al. **examine** the performance of chains as the number of hops are increased and study the effect of cross-interference between chains [5].	sentHasCitation sentHasTerm=SOI	RWD_CS – Related Work Description – Citation Sentence
3.	Several similar techniques are **reported** to incorporate the synonyms **[9]**, hypernyms [12], hyponyms [13], meronyms and holonyms [16] of words found in the training documents for classifier training.	sentHasCitation prevSentHasCitation sentHasTerm=OCT	RWO_CS – Related Work Outcome – Citation Sentence
4.	In his following work [7], he refined the system and **improved** the performance by considering the temporal and comment information and enhancing the quality measure according to the query dependent context.	sentHasCitation prevSentHasCitation sentHasTerm=STH	RWS_CS - Related Work Strength - Citation Sentence
5.	**They** further extend their work to **analyze** chains of n hops.	sentHasTerm=SPN sentHasTerm=DES	RWD - Related Work Description
6.	**They show** that hidden terminals cause packet drops affecting chain throughput and causing route stability.	prevSentHasCitation sentHasTerm=SPN sentHasTerm=OCT	RWO - Related Work Outcome
7.	The experimental results show it can **improve** the ranking result.	sentHasTerm=STH	RWS - Related Work Strength
8.	**They analyze** the effect of MAC 802.11 behavior on the performance of multi-hop chains but **do not** categorize interference patterns that govern network performance in terms of throughput and bandwidth utilization.	sentHasTerm=SPN sentHasTerm=DES sentHasTerm=SCT	RWSC - Related Work Shortcoming
9.	**Different from the previous work**, **our** approach assesses the relevant documents by employing the click-through data of statistically significant users in real Web search settings.	sentHasTerm=RES sentHasTerm=CON	CWRW - Contrasting Work with Related Work
10.	Another novel and important approach in feature space transformation for unsupervised learning is spectral clustering [10].	sentHasCitation sentHasTerm=AAT	AWRS_CS - Alternate Work for Related Work Citation Sentence

11.	Two **alternatives** are studied: either sampling documents or using a reference corpus independent of the target retrieval collection.	prevSentHasCitation sentHasTerm=AAT	AWRW - Alternate Work for Related Work
12.	As **we will show** in this paper, the types of interactions within chain have a substantial effect on the performance of a network under TCP traffic.	sentHasTerm=RES	CWO - Current Work Outcome
13.	**The paper does not consider** how to actually determine (collect) and distribute this information.	prevSentHasCitation sentHasTerm=SCT sentHasTerm=RES	CWSC - Current Work Shortcomings

Table 6: Classification results of the Classifier

Sl. No.	Label	Accuracy = 96.51			Accuracy = 93.22		
		with Citation and Sentence Features			**with Sentence Features Only**		
		Precision	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
1.	BG	1.00	0.99	0.99	0.96	0.92	0.93
2.	RWD_CS	0.98	0.99	0.98	0.98	0.97	0.97
3.	RWSC	0.98	0.99	0.98	0.92	0.98	0.94
4.	RWD	0.98	0.98	0.97	0.91	0.94	0.92
5.	CWO	0.94	1.0	0.96	0.92	0.98	0.94
6.	RWS	0.90	0.90	0.90	0.66	0.54	0.59
7.	CWOW	1.00	0.81	0.88	1.00	0.72	0.83
8.	ASRW	0.83	0.83	0.82	0.75	0.50	0.60
9.	CWSC	0.80	0.66	0.72	0.60	0.50	0.54
10.	RWO	0.57	0.44	0.49	0.57	0.44	0.49
11.	AWRW_CS	1.00	0.25	0.40	0.50	0.25	0.33
12.	RWS_CS	1.00	0.20	0.33	-	-	-
13.	RWO_CS	0.33	0.20	0.24	0.25	0.20	0.22

Table 7: Confusion Matrix for the Classifier

Label	BG		RWD_CS		RWSC		RWD		RWO		RWO_CS		CWO		RWS		RWS_CS		ASRW		AWRW_CS		CWSC		CWOW		Total
	D1*	D2**	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2	
BG	102	95	0	0	0	0	0	8	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	103
RWD_CS	0	1	330	323	0	2	0	2	0	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0	331
RWSC	0	0	0	0	201	199	1	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	203
RWD	0	2	0	0	2	8	225	217	0	0	0	0	2	4	0	0	0	0	0	0	0	0	0	0	0	0	229
RWO	0	0	0	0	0	0	2	1	4	4	2	3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	9
RWO_CS	0	0	0	0	0	0	1	1	3	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
CWO	0	0	0	0	0	1	0	1	0	0	0	0	140	138	0	0	0	0	0	0	0	0	0	0	0	0	140
RWS	0	0	0	1	0	0	0	2	0	0	0	0	0	0	10	6	1	2	0	0	0	0	0	0	0	0	11
RWS_CS	0	0	3	1	0	2	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0	0	5
ASRW	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	5	3	0	1	0	0	0	0	6
AWRW_CS	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	4
CWSC	0	0	0	0	2	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4	3	0	0	6
CWOW	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	0	0	0	0	0	0	0	0	9	8	11
																										Total	1063

* D1 – Dataset 1 – With both Citation Features and Sentence Features; ** D2 – Dataset 2 – With Sentence Features alone

- Expected Label

Total Number of Sentences – 1063; Number of sentences classified correctly with citation features and sentences features – 1026

Number of sentences classified correctly with sentence features alone - 991

8 REFERENCES

[1] Buckingham Shum, S. J., Uren, V., Li, G., Sereno, B. and Mancini, C. 2007. Modeling naturalistic argumentation in research literatures: Representation and interaction design issues. *International Journal of Intelligent Systems*, 22 (1), 17-48.

[2] Borgman, C. L. and Furner, J. 1990. *Scholarly communication and bibliometrics*. Sage Newbury Park, CA.

[3] Luukkonen, T. 1992. Is scientists' publishing behaviour reward seeking? *Scientometrics*, 24 (2), 297-319.

[4] Ziman, J. M. 1968. *Public knowledge: an essay concerning the social dimension of science*. Cambridge University Press, Cambridge.

[5] Bonzi, S. 2007. Characteristics of a literature as predictors of relatedness between cited and citing works. *Journal of the American Society for Information Science and Technology*, 33 (4), 208-216.

[6] Weinstock, M. 1971. Citation indexes. In *Encyclopedia of library and information science*, 5, Dekker, New York, 16–40.

[7] Moravcsik, M. J. and Murugesan, P. 1975. Some results on the function and quality of citations. *Social Studies of Science*, 5 (1), 86-92.

[8] Nanba, H. and Okumura, M. 1999. Towards multi-paper summarization using reference information. In *16th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers, 926-931.

[9] Teufel, S., Carletta, J. and Moens, M. 1999. An annotation scheme for discourse-level argumentation in research articles. In *9th European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, 110-117.

[10] Teufel, S. and Moens, M. 2000. What's yours and what's mine: determining intellectual attribution in scientific text. In. *2000 Joint SIGDAT Conference on Empirical Methods in NLP, 2000,* Association for Computational Linguistics, 9-17.

[11] Garzone, M. and Mercer, R. E. 2000. Towards an automated citation classifier. In *Advances in Artificial Intelligence: Lecture Notes in Computer Science*, 1822 (2000), Springer Berlin, Heidelberg, 337-346.

[12] Mercer, R. E. and Di Marco, C. 2000. The importance of fine-grained cue phrases in scientific citations. In *Advances in Artificial Intelligence: Lecture Notes in Computer Science*, 2671 (2003), Springer Berlin, Heidelberg, 550-556.

[13] Le, M., Ho, T. B. and Nakamori, Y. 2006. Detecting citation types using finite-state machines. In *Advances in Knowledge Discovery and Data Mining: Lecture Notes in Computer Science*, 3918 (2006), Springer Berlin, Heidelberg, 265-274.

[14] Swales, J. 1986. Citation analysis and discourse analysis. *Applied Linguistics*, 7 (1), 39-56.

[15] Teufel, S., Siddharthan, A. and Tidhar, D. 2006. An annotation scheme for citation function. In *7th SIGdial Workshop on Discourse and Dialogue*, Sydney, Australia, 80-87. http://acl.ldc.upenn.edu/W/W06/W06-1312.pdf

[16] Lafferty, J., McCallum, A. and Pereira, F. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*. 2001. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.9849&rep=rep1&type=pdf

[17] McCallum, A. and Li, W. 2003. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proceedings of 7th Conference on Natural Language Learning*. 2003. http://acl.ldc.upenn.edu/W/W03/W03-0430.pdf

[18] Pinto, D., McCallum, A., Wei, X. and Croft, W. B. 2003. Table extraction using conditional random fields. In *Proceedings of the 2003 Annual National Conference on Digital Government Research*, Digital Government Society of North America, 1-4.

[19] Sha, F. and Pereira, F. 2003. Shallow parsing with conditional random fields. In *Proceedings of HLT-NAACL 2003*, Edmonton, May-June 2003, 134-141. http://acl.ldc.upenn.edu/N/N03/N03-1028.pdf

[20] McCallum, A., Freitag, D. and Pereira, F. 2000. Maximum entropy Markov models for information extraction and segmentation. In *17th International Conference on Machine Learning*, Morgan Kaufmann Publishers, 591-598.

[21] Springerlink.com http://www.springerlink.com/home/main.mpx

[22] Majeed, A., Razak, S., Abu-Ghazaleh, N.B. and Harras, H.A. 2009. TCP over Multi-Hop Wireless Networks: The impact of MAC level interactions. In *Ad-Hoc Mobile and Wireless Networks: Lecture Notes in Computer Science*, 5793 (2009), Springer Berlin, Heidelberg, 1-15.

[23] McCallum, A.K. 2002. *Mallet: A Machine Learning for Language Toolkit*. http://mallet.cs.umass.edu.