Citation Context Sentiment Analysis for Structured Summarization of Research Papers

Niket Tandon[1,3], Ashish Jain[2,3]
ntandon@mpi-inf.mpg.de, and ashish.iiith@gmail.com

[1] Max Planck Institute for Informatics, Saarbrücken, Germany
[2] IIIT Hyderabad, India
[3] PQRS Research (pqrs-research.org)

Abstract. Structured tabular summarization tremendously helps humans understand a topic, e.g. Wikipedia infoboxes. However, few methods exist to generate summary of research papers although it is time taking and painstaking to read a paper and even more difficult to infer its merits and limitations. We propose a method to generate structured summary of research papers. We turn to opinion of citing papers, because they are shown to be more focused than abstracts and contain additional information. This paper is a first step towards structured summarization of research papers using citing papers.

1 Introduction

There is a plethora of research papers, making it hard for students and researchers to be abreast with the literature. Skimming through a research paper to get the broad ideas in the paper is an art. Judging positive and negative is best left to experts and requires time. Thus, the problem of quickly gaining insights about a research paper remains unaddressed. We propose to utilize opinion and summaries in citation contexts to address this problem. A citation context of a paper P, is the set of sentences about P in other articles which cite P. Citation context contains concise and precise analysis about a paper due to the space limitations in papers and due to the high quality of a paper in terms of correctness. This paper envisions to summarize these opinions and summaries from all citing papers and present in a table with five columns: summary, related work, strengths, limitations, and extensions. Such an example summary of an Information Extraction system, KnowItAll [4] is presented in Figure 1.

Problem statement Our problem can be divided into two sub-problems.

(i) Classifying citation context into one or more of the five classes. This is challenging because we have limited training data that classifies citation context. Secondly, very few techniques exist for sentiment analysis of research papers in more than two classes.

[4] cs.washington.edu/research/knowitall

(ii) Generating summary snippets and merging similar statements from the classified citation contexts. e.g. given a negative citation context: "We use the CPS transformation (citation) but our implementation is simplified by the fact that we start from a normalized direct style representation". We want a summary statement from this that says, "CPS transformation's implementation is not simple". We keep this as future work.

Summary	Strengths	limitations	Related work	Applications
-Extract structured information from the Web documents. - Input is a text document and output is tuples of the target relation. - employs the same generic patterns as Hearst (e.g., "NPs such as NP1, NP2, …"), to extract facts	- extract information from the web in a site-independent manner using simple yet effective natural language techniques - information extraction tools over web documents, which automatically construct corpora with large numbers of structured entities	- still lack the combined recall and precision necessary to allow for very robust queries	- DIPRE, DBLife, DbLife2, Rapier, Snowball, Freebase	- can be exploited for web-based question-answering - can extract hyponyms of a user-specified class... ...KnowItAll takes the description of a concept or class

Fig. 1. Our Vision: Structured summarization

Related Work. Although sentiment analysis is a well-studied topic, sentiment analysis of citation context got surprisingly less attention in the research community. Sentiments in citation contexts differ in both structure and language from standard use cases e.g. product reviews, thus standard sentiment analysis of citation context using lexical resources for instance, leads to poor coverage and accuracy [1].

Some attempts [2] have been made to utilize the sentiments of citation context by manually classifying sentiments as positive or negative. Other approaches [3] rely on manually defined phrase based rules for classifying sentiments. However, no large scale automated sentiment analysis over citation context exists. Further, these approaches have considered only classified citation context in two or three classes [3][1], although citation context can be leveraged in more than these two classes.

In one of the first attempts, [4] describe paper summarization as a classification task and classify each sentence in an article into aim, contrast and background. They do not leverage the citation context for summarization. A seminal approach in [5], [6] leverages citation context for summarization. Their approach is primarily geared towards multiple document summarization and less focused on single paper summarization. They consider the phrases in citation context as unstructured summarization i.e. consisting of uncategorized sentences, but do not consider the sentiments of the citation context. Unlike existing approaches that provide unstructured summarization of a research paper, our goal is to perform structured summarization of a research paper.

Contribution. This paper aims at filling the gap between sentiment analysis and citation context summarization by proposing a structured summarization approach. An example that depicts our goal is shown in Figure 1.

– Unlike standard sentiment analysis for items like product review, we propose an automated approach directed towards research papers.
– Unlike standard summarization approach that is unstructured, we provide structured summarization of research papers that is more desired.

2 Methodology

Structured summarization of a research paper can be viewed as classification of a citation context into one or more of the following classes: summary, related work, strengths, limitations, and extensions. The classification problem here is multilabel because a citation context can belong to one or more classes. Consider the following citation context that summarizes the paper as well as describes an application of the work, "The (KnowItAll system) employs the same generic patterns as Hearst (e.g., NPs such as NP1, NP2,), and more besides, to extract a whole range of facts that can be exploited for web-based question-answering". Multilabel classification can lead to an intermediate summarization as presented in Figure 2.

We use a Language Model(LM) approach for classification of citation context. In brief, language models are constructed for each of the five classes. Subsequently, the most likely language models that would have generated a citation context are estimated. Our LM based classification approach is similar to a sentiment classification approach used in [7].

LM construction. Given a collection of citation contexts D, we manually annotate them into the five classes : summary, related work, strengths, limitations, and extensions. We identify the "opinion vocabulary", consisting of two kinds of terms: phrases denoting the context, and opinion terms describing opinions on the cited paper. In the opinion vocabulary, bigrams are taken as context while unigram verbs, adjectives, and adverbs are assumed to be opinion related.

An LM M_{ci} of a particular class c_i is estimated as the interpolation of a bigram phrase denoting context B and a unigram opinion term U over all phrases and opinion terms in the collection. Such an interpolated LM benefits from two LMs.

$$P_{M\,ci}(t_i|D) = (1-\alpha)P_B(t_i|D) + \alpha P_U(t_i|D)$$

where $P_B(t_i|D)$: LM of D over binary terms, $P_U(t_i|D)$:LM of D over unary terms, t_i: a term and α: interpolation parameter (estimated by minimizing perplexity). The unigram and bigram models are obtained using the general form: $P(t_i|D) = \frac{c(t_i,D)}{\sum_{t_j \in D} c(t_j,D)}$ where $c(t_j, D)$ denotes the frequency of term t_j in the collection D. Further, Good Turing smoothing is applied because several out of vocabulary(OOV) words could exist.

Classifying the citation context. The citation context is modeled as a query Q_{CT} by extracting the binary and unigram patterns from the citation context, $Q_{CT} = \{B \cup U\}$ Similarly to LM usage in information retrieval, we estimate the query likelihood of a query given the LM of each class c_i.

$$P(Q_{CT}|M_{ci}) = \prod_{t_i \in Q_{CT}} P(t_i|M_{ci})$$

In case of single label classification, the model that has the highest likelihood of generating the query is selected. However, we consider a multilabel classification that requires an additional step. Our hypothesis is that if two or more LMs have query likelihood in the neighborhood δ of the best LM, then the LMs should also be accepted since the problem is multilabel classification. On the other hand, LMs whose query likelihood is futher off from the best LM should not be considered. We empirically estimate δ.

Summary	The KnowItAll system of \Etzioni et al. (2004)\ employs the same generic patterns as Hearst (e.g., "NPs such as NP1, NP2, …"), and more besides, to extract a whole range of facts that can be exploited for web-based question-answering	To extract the structured information from text documents, we can use an information extraction system, such as Snowball [3], Proteus [21], MinorThird [12], or KnowItAll \[16]\, which take as input a text document and produce tuples of the target relation...	Recently, a number of research projects [4\, 2, 6] have identified the exciting opportunity to collect structured data from the Web...
Strengths	The KnowItAll \[30]\ and the TextRunner work at the University of Washington also aim to extract information from the web in a site-independent manner using simple yet effective NLP.	Another example is the use of information extraction tools over web documents, which automatically construct corpora with large numbers of structured entities [2, \6\].	
Limitations	\[4]\[10]), but still lack the combined recall and precision necessary to allow for very robust queries...		
Related work	and structured integrated corpora such as Freebase and KnowItAll \[30]\, and community driven mass collaboration sites such as DBLife [25, 48].	A large family of existing solutions [33, \16\, 3, 29, 31] focus on improving the extraction accuracy by directly manipulating the information extraction system for a given task...	...Examples of reallife extraction systems include Avatar1, DBLife2, DIPRE [3], KnowItAll \[10]\,
Applications	The KnowItAll system of \Etzioni et al. (2004)\ employs the same generic patterns as Hearst (e.g., "NPs such as NP1, NP2, …"), and more besides, to extract a whole range of facts that can be exploited for web-based question-answering	For example, the KnowItAll \[5]\ system can extract hyponyms of a user-specified class... ...KnowItAll \[5]\ takes the description of a concept or class	

Fig. 2. Classification of citation contexts

3 Experiments

Experimental Setup. We use a standard multilabel classification metric, Average Precision, which computes accuracy(precision) of each class and averages them over all the classes. Citation contexts for research papers are available online on Microsoft Academic search engine [5]. There is no annotated dataset for our purpose, so we create an annotated set of 30 research papers, totaling an annotation of 500 citation contexts.

[5] `academic.research.microsoft.com`

Baseline: As a baseline for multilabel classification, we use Random k-Labelsets with Naive Bayes algorithm as the basis [8]. As features for the baseline, we consider the combinations of the following: (i) Adjectives in each class, (ii) Verbs, (iii) n-grams.

Experimental results. The baseline is trained and language models are constructed on a total of 500 labeled citation contexts in the collection D. A combination of adjectives, verbs and bigrams achieves 68.54% average precision, see Table 1, marginally beating the LM. We postulate that the LM accuracy could be further improved by increasing and cleaning the collection e.g. "limitation" class has only 47 instances annotated out of 500. Learning difficult underlying patterns with small dataset leads to low precision for that class as well as reduces the overall average precision of the experiment.

Classifier	Features	Average Precision(%)
Baseline	Adj	65.54
Baseline	Verb	66.30
Baseline	Adj+Verb	67.48
Baseline	**Adj+Verb+Bigram**	68.54
LM	**Bigram terms B + Unigram terms U**	67.00

Table 1. Average precision of MultiLabel Classifier

4 Conclusion

We introduced a new framework based on citation sentiments for structured summarization of a research paper. Our results are encouraging given the simplicity of our model i.e. multilabel classification. In the future, we would enhance our approach by employing more sophisticated algorithms like LDA and address snippet generation.

References

1. Piao, S., Ananiadou, S., Tsuruoka, Y., Sasaki, Y., McNaught, J.: Mining opinion polarity relations of citations. In: International Workshop on Computational Semantics (IWCS 2007)
2. Stamou, S., Mpouloumpasis, N., Kozanidis, L.: Deriving the impact of scientific publications by mining citation opinion terms. IJDIM 2009 **7**(5) (2009)
3. Nanba, H., Kando, N., Okumura, M., et al.: Classification of research papers using citation links and citation types: Towards automatic review article generation. (2000)
4. Teufel, S.: Argumentative zoning for improved citation indexing. Computing Attitude and Affect in Text: Theory and Applications (2006) 159–169
5. Qazvinian, V., Radev, D.: Scientific paper summarization using citation summary networks. In: COLING 2008. (2008) 689–696
6. Elkiss, A., Shen, S., Fader, A., Erkan, G., Radev, D., et al.: Blind men and elephants: What do citation summaries tell us about a research article? JASIST **59**(1) (2008) 51–62
7. Awadallah, R., Ramanath, M., Weikum, G.: Language-model-based pro/con classification of political text. In: Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval. (2010) 747–748
8. Tsoumakas, G., Vlahavas, I.: Random k-labelsets: An ensemble method for multilabel classification. Machine Learning: ECML 2007 (2007) 406–417