

Detection of Implicit Citations for Sentiment Detection

Awais Athar

Simone Teufel

Computer Laboratory, University of Cambridge
15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
{awais.athar, simone.teufel}@cl.cam.ac.uk

Abstract

Sentiment analysis of citations in scientific papers is a new and interesting problem which can open up many exciting new applications in bibliometrics. Current research assumes that using just the citation sentence is enough for detecting sentiment. In this paper, we show that this approach misses much of the existing sentiment. We present a new corpus in which all mentions of a cited paper have been annotated. We explore methods to automatically identify these mentions and show that the inclusion of implicit citations in citation sentiment analysis improves the quality of the overall sentiment assignment.

1 Introduction

The idea of using citations as a source of information has been explored extensively in the field of bibliometrics, and more recently in the field of computational linguistics. State-of-the-art citations identification mechanisms focus either on detecting explicit citations i.e. those that consist of either the author names and the year of publication or bracketed numbers only, or include a small sentence window around the explicit citation as input text (Councill et al., 2008; Radev et al., 2009; Ritchie et al., 2008). The assumption behind this approach is that all related mentions of the paper would be concentrated in the immediate vicinity of the anchor text. However, this assumption does not generally hold true (Teufel, 2010; Sugiyama et al., 2010). The phenomenon of trying to determine a citations's *citation context* has a long tradition in library sciences

(O'Connor, 1982), and its connection with coreference has been duely noted (Kim et al., 2006; Kaplan et al., 2009). Consider Figure 1, which illustrates a typical case.

grams. In order to improve sentence-level evaluation performance, several metrics have been proposed, including ROUGE-W, ROUGE-S (Lin and Och, 2004) and METEOR (Banerjee and Lavie, 2005). ROUGE-W differs from BLEU and NIST

⋮

METEOR is essentially a unigram based metric, which prefers the monotonic word alignment between MT output and the references by penalizing crossing word alignments. There are two problems with METEOR. First, it doesn't consider gaps in the aligned words, which is an important feature for evaluating the sentence fluency; second, it cannot use multiple references simultaneously.¹ ROUGE and METEOR both use WordNet

Figure 1: Example of the use of anaphora

While the first sentence cites the target paper explicitly using the name of the primary author along with the year of publication of the paper, the remaining sentences mentioning the same paper appear after a gap and contain an indirect and implicit reference to that paper. These mentions occur two sentences after the formal citation in the form of anaphoric *it* and the lexical hook *METEOR*. Most current techniques, with the exception of Qazvinian and Radev (2010), are not able to detect linguistic mentions of citations in such forms. Ignoring such mentions and examining only the sentences contain-

ing an explicit citation results in loss of information about the cited paper. While this phenomenon is problematic for applications like scientific summarisation (Abu-Jbara and Radev, 2011), it has a particular relevance for citation sentiment detection (Athar, 2011).

Citation sentiment detection is an attractive task. Availability of citation polarity information can help researchers in understanding the evolution of a field on the basis of research papers and their critiques. It can also help expert researchers who are in the process of preparing opinion based summaries for survey papers by providing them with motivations behind as well as positive and negative comments about different approaches (Qazvinian and Radev, 2008).

Current work on citation sentiment detection works under the assumption that the sentiment present in the citation sentence represents the true sentiment of the author towards the cited paper (Athar, 2011; Piao et al., 2007; Pham and Hoffmann, 2004). This assumption is so dominant because current citation identification methods (Councill et al., 2008; Ritchie et al., 2008; Radev et al., 2009) can readily identify the citation sentence, whereas it is much harder to determine the relevant context. However, this assumption most certainly does not hold true when the citation context spans more than one sentence.

Concerning the sentiment aspect of the citation context from Figure 1, we see that the citation sentence does not contain any sentiment towards the cited paper, whereas the following sentences act as a critique and list its shortcomings. It is clear that criticism is the intended sentiment, but if the gold standard is defined by looking at the citation sentence in isolation, a significant amount of sentiment expressed in the text is lost. Given that overall most citations in a text are neutral with respect to sentiment (Spiegel-Rosing, 1977; Teufel et al., 2006), this makes it even more important to recover what explicit sentiment there is in the article, wherever it is to be found.

In this paper, we examine methods to extract all opinionated sentences from research papers which mention a given paper in as many forms as we can identify, not just as explicit citations. We present a new corpus in which all mentions of a cited paper

have been manually annotated, and show that our annotation treatment increases citation sentiment coverage, particularly for negative sentiment. We then explore methods to automatically identify all mentions of a paper in a supervised manner. In particular, we consider the recognition of named approaches and acronyms. Our overall system then classifies explicit and implicit mentions according to sentiment. The results support the claim that including implicit citations in citation sentiment analysis improves the quality of the overall sentiment assignment.

2 Corpus Construction

We use the dataset from Athar (2011) as our starting point, which consists of 8,736 citations in the ACL Anthology (Bird et al., 2008) that cite a target set of 310 ACL Anthology papers. The citation summary data from the ACL Anthology Network¹ (Radev et al., 2009) is used. This dataset is rather large, and since manual annotation of context for each citation is a time consuming task, a subset of 20 target papers (i.e., all citations to these) has been selected for annotation. These 20 papers correspond to approximately 20% of incoming citations in the original dataset. They contain a total of 1,555 citations from 854 citing papers.

2.1 Annotation

We use a four-class scheme for annotation. Every sentence which does not contain any direct or indirect mention of the citation is labelled as being excluded (x) from the context. The rest of the sentences are marked either positive (p), negative (n) or objective/neutral (o). To speed up the annotation process, we developed a customised annotation tool.

A total of 203,803 sentences have been annotated from 1,034 paper-reference pairs. Although this annotation been performed by the first author only, we know from previous work that similar styles of annotation can achieve acceptable inter-annotator agreement (Teufel et al., 2006). An example annotation is given in Figure 2, where the first column shows the line number and the second one shows the class label for the citation to *Smadja (1993)*. It should be noted that since annotation is always per-

¹<http://www.aclweb.org>

formed for a specific citation only, sentences such as the one at line 32, which carry sentiment but refer to a different citation, are marked as excluded from the context.

If there are multiple sentiments in the same sentence, the sentence has been labelled with the class of the last sentiment mentioned. In this way, a total of 3,760 citation sentences have been found in the whole corpus, i.e. sentences belonging to class o , n or p , and the rest have been labelled as x . Table 1 compares the number of sentences with only the explicit citations with all explicit and implicit mentions of those citations. We can see that including the citation context increases the subjective sentiment by almost 185%. The resulting negative sentiment also increases by more than 325%. This may be attributed to the strategic behaviour of the authors of ‘sweetening’ the criticism in order to soften its effects among their peers (Hornsey et al., 2008).

31	x	Church and Hanks (Church and Hanks 1990) employed mutual information to extract both adjacent and distant bi-grams that tend to co-occur within a fixed-size window.
32	x	But the method did not extend to extract n-grams.
33	o	Smadja (Smadja 1993) proposed a statistical model by measuring the spread of the distribution of cooccurring pairs of words with higher strength.
34	p	This method successfully extracted both adjacent and distant bi-grams and n-grams.
35	n	However, the method failed to extract bi-grams with lower frequency.

Figure 2: Example annotation of a citation context.

	Explicit mentions	All mentions
o	1,509	3,100
n	86	368
p	146	292

Table 1: Distribution of classes.

Another view of the annotated data is available in Figure 3a. This is in the form of interactive HTML where each HTML page represents all the incoming links to a paper. Each row represents the citing paper and each column square represents a sentence. The rows are sorted by increasing publication date.

Black squares are citations with the author name and year of publication mentioned in the text. The red, green and gray squares show negative, positive and neutral sentiment respectively. Pointing the mouse cursor at any square gives the text content of the corresponding sentence, as shown in the Figure 3a.

The ACL Id, paper title and authors’ names are also given at the top of the page. Similar data for the corresponding citing paper is made available when the mouse cursor is positioned on one of the orange squares at the start of each row, as shown in the Figure 3b. Clicking on the checkboxes at the top hides or shows the corresponding type of squares. There is also an option to hide/show a grid so that the squares are separated and rows are easier to trace. For example, Figure 3b shows the grid with the neutral or objective citations hidden.

In the next section, we describe the features set we use to detect implicit citations from this annotated corpus and discuss the results.

3 Experiments and Results

For the task of detecting all mentions of a citation, we merge the class labels of sentences mentioning a citation in any form (o_n_p). To make sure that the easily detectable explicit citations do not influence the results, we change the class label of all those sentences to x which contain the first author’s name within a 4-word window of the year of publication.

Our dataset is skewed as there are many more objective sentences than subjective ones. In such scenarios, average *micro-F* scores tend to be slightly higher as they are a weighted measure. To avoid this bias, we also report the *macro-F* scores. Furthermore, to ensure there is enough data for training each class, we use 10-fold cross-validation (Lewis, 1991) in all our experiments.

We represent each citation as a feature set in a Support Vector Machine (SVM) (Cortes and Vapnik, 1995) framework. The corpus is processed using WEKA (Hall et al., 2008) and the Weka LibSVM library (EL-Manzalawy and Honavar, 2005; Chang and Lin, 2001). For each i^{th} sentence S_i , we use the following binary features.

- S_{i-1} contains the last name of the primary author, followed by the year of publication within a four-word window.

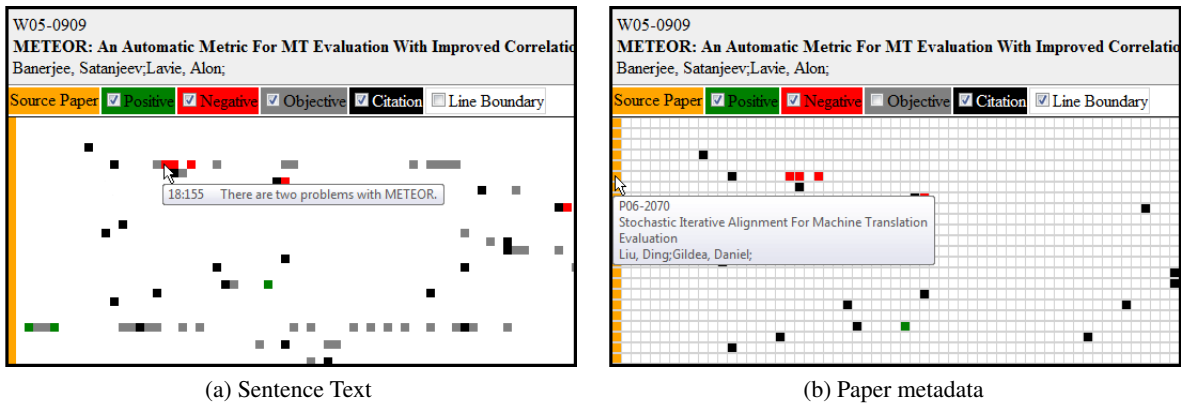


Figure 3: Different views of an annotated paper.

This feature is meant to capture the fact that the sentence immediately after an explicit citation is more likely to continue talking about the same work.

- S_i contains the last name of the primary author followed by the year of publication within a four-word window.

This feature should help in identifying sentences containing explicit citations. Since such sentences are easier to identify, including them in the evaluation metric would result in a false boost in the final score. We have thus excluded all such sentences in our annotation and this feature should indicate a negative instance to the classifier.

- S_i contains the last name of the primary author.

This feature captures sentences which contain a reference to tools and algorithms which have been named after their inventors, such as,

“One possible direction for future work is to compare the search-based approach of **Collins** and Roark with our DP-based approach.”

It should also capture the mentions of methods and techniques used in the cited paper e.g.,

“We show that our approach outperforms **Turney**’s approach.”

- S_i contains an acronym used in an explicit citation.

Acronyms are taken to be capitalised words which are extracted from the vicinity of the cited author’s last name using regular expressions. For example, *METEOR* in Figure 1 is an acronym which is used in place of a formal citation to refer to the original paper in the rest of the citing paper.

- S_i contains a determiner followed by a work noun.

We use the following determiners $D = \{the, this, that, those, these, his, her, their, such, previous, other\}$. The list of work nouns (*technique, method, etc.*) has been taken from Teufel (2010). This feature extracts a pattern which has been found to be useful for extracting citations in previous work (Qazvinian and Radev, 2010). Such phrases usually signal a continuation of the topics related to citations in earlier sentences. For example:

“Church *et al.*(1989), Wettler & Rapp (1989) and Church & Hanks (1990) describe algorithms which do this. However, the validity of **these algorithms** has not been tested by systematic comparisons with associations of human subjects.”

- S_i starts with a third person pronoun.

The feature also tries to capture the topic continuation after a citation. Sentences starting with a pronoun (e.g. *they, their, he, she, etc.*) are more likely to describe the subject citation of the previous sentence in detail. For example:

“Because Daume III (2007) views the adaptation as merely augmenting the feature space, each of his features has the same prior mean and variance, regardless of whether it is domain specific or independent. **He** could have set these parameters differently, but he did not.”

- S_i starts with a connector.

This feature also focuses on detecting the topic continuity. Connectors have been shown to be effective in other context related works as well (Hatzivassiloglou and McKeown, 1997; Polanyi and Zaenen, 2006). A list of 23 connectors (e.g. *however, although, moreover, etc.*) has been compiled by examining the high frequency connectors from a separate set of papers from the same domain. An example is:

“An additional consistent edge of a linear-chain conditional random field (CRF) explicitly models the dependencies between distant occurrences of similar words (Sutton and McCallum, 2004; Finkel et al. , 2005). **However**, this approach requires additional time complexity in inference/learning time and it is only suitable for representing constraints by enforcing label consistency.”

- S_i starts with a (sub)section heading.
- S_{i-1} starts with a (sub)section heading.
- S_{i+1} starts with a (sub)section heading.

The three features above are a consequence of missing information about the paragraph and section boundaries in the used corpus. Since the text extraction has been done automatically, the section headings are usually found to be merged with the text of the succeeding sentence. For example, the text below merges the heading of section 4.2 with the next sentence.

“4.2 METEOR vs. SIA SIA is designed to take the advantage of loose sequence-based metrics without losing word-level information.”

Start and end of such section boundaries can give us important information about the scope of a citation. In order to exploit this information, we use regular expressions to detect if the

sentences under review contains these merged section titles and headings.

- S_i contains a citation other than the one under review.

It is more probable for the context of a citation to end when other citations are mentioned in a sentence, which is the motivation behind using this feature, which might contribute to the discriminating power of the classifier in conjunction with the presence of a citation in the previous sentence. For example, in the extract below, the scope of the first citation is limited to the first sentence only.

“Blitzer et al.(2006) proposed a structural correspondence learning method for domain adaptation and applied it to part-of-speech tagging. **Daume III (2007)** proposed a simple feature augmentation method to achieve domain adaptation.”

- S_i contains a lexical hook.

The lexical hooks feature identifies lexical substitutes for the citations. We obtain these hooks by examining all explicit citation sentences to the cited paper and selecting the most frequent capitalized phrase in the vicinity of the author’s last name. The explicit citations come from all citing papers and not just the paper for which the features are being determined. For example, the sentences below have been taken from two different papers and cite the same target paper (Cutting et al., 1992). While the acronym *HMM* will be captured by the feature stated earlier, the word *Xerox* will be missed.

E95-1014: “This text was part-of-speech tagged using the **Xerox HMM** tagger (Cutting et al. , 1992).”

J97-3003: “The **Xerox** tagger (Cutting et al. 1992) comes with a set of rules that assign an unknown word a set of possible pos-tags (i.e. , POS-class) on the basis of its ending segment.”

This ‘domain level’ feature makes it possible to extract the commonly used name for a technique which may have been missed by the acronym feature due to long term dependencies. We also extrapolate the acronym for such

phrases, e.g., in the example below, *SCL* would also be checked along with *Structural Correspondence Learning*.

“The paper compares **Structural Correspondence Learning** (Blitzer et al., 2006) with (various instances of) self-training (Abney, 2007; McClosky et al., 2006) for the adaptation of a parse selection model to Wikipedia domains”

We also add n -grams of length 1 to 3 to this lexical feature set and compare the results obtained with an n -gram only baseline in Table 2. N -grams have been shown to perform consistently well in various NLP tasks (Bergsma et al., 2010).

Class	Baseline	Our System
x	0.995	0.996
o_n_p	0.358	0.513
<i>Avg.</i>	0.990	0.992
<i>Avg.(macro)</i>	0.677	0.754

Table 2: Comparison of F -scores for non-explicit citation detection.

By adding the new features listed above, the performance of our system increases by almost 8% over the n -gram baseline for the task of detecting citation mentions. Using the pairwise Wilcoxon rank-sum test at 0.05 significance level, we found that the difference between the baseline and our system is statistically significant². While the *micro-F* score obtained is quite high, the individual class scores show that the task is hard and a better solution may require a deeper analysis of the context.

4 Impact on Citation Sentiment Detection

We explore the effect of this context on citation sentiment detection. For a baseline, we use features of the state-of-the-art system proposed in our earlier work (Athar, 2011). While there we used n -gram and dependency feature on sentences containing explicit citations only, our annotation is not restricted to such citations and we may have more than one

²While this test may not be adequate as the data is highly skewed, we are reporting the results since there is no obvious alternative for discrete skewed data. In future, we plan to use the continuous probability estimates produced by the classifier for testing significance.

sentiment per each explicit citation. For example, in Figure 2, our 2011 system will be restricted to analysing sentence 33 only. However, it is clear from our annotation that there is more sentiment present in the succeeding sentences which belongs to this explicit citation. While sentence 34 in Figure 2 is positive towards the cited paper, the next sentence criticises it. Thus for this explicit citation, there are three sentences with sentiment and all of them are related to the same explicit citation. Treating these sentences separately will result in an artificial increase in the amount of data because they participate in the same discourse. It would also make it impossible to compare the sentiment annotated in the previous work with our annotation.

To make sure the annotations are comparable, we mark the true citation sentiment to be the last sentiment mentioned in a 4-sentence context window, as this is pragmatically most likely to be the real intention (MacRoberts and MacRoberts, 1984). The window length is motivated by recent research (Qazvinian and Radev, 2010) which favours a four-sentence boundary for detecting non-explicit citations. Analysis of our data shows that more than 60% of the subjective citations lie in this window. We include the implicit citations predicted by the method described in the previous section in the context. The results of the single-sentence baseline system are compared with this context enhanced system in Table 3.

Class	Baseline	Our System
o	0.861	0.887
n	0.138	0.621
p	0.396	0.554
<i>Avg.</i>	0.689	0.807
<i>Avg.(macro)</i>	0.465	0.687

Table 3: F -scores for citation sentiment detection.

The results show that our system outperforms the baseline in all evaluation criteria. Performing the pairwise Wilcoxon rank-sum test at 0.05 significance level, we found that the improvement is statistically significant. The baseline system does not use any context and thus misses out on all the sentiment information contained within. While this window-based representation does not capture all the senti-

ment towards a citation perfectly, it is closer to the truth than a system based on single sentence analysis and is able to detect more sentiment.

5 Related Work

While different schemes have been proposed for annotating citations according to their function (Spiegel-Rosing, 1977; Nanba and Okumura, 1999; Garzone and Mercer, 2000), the only recent work on citation sentiment detection using a relatively large corpus is by Athar (2011). However, this work does not handle citation context. Other approaches to citation classification include work by Wilbur et al. (2006), who annotated a 101 sentence corpus on focus, polarity, certainty, evidence and directionality. Piao et al. (2007) proposed a system to attach sentiment information to the citation links between biomedical papers by using existing semantic lexical resources and NLP tools.

A common approach for sentiment detection is to use a labelled lexicon to score sentences (Hatzivassiloglou and McKeown, 1997; Turney, 2002; Yu and Hatzivassiloglou, 2003). However, such approaches have been found to be highly topic dependent (Engström, 2004; Gamon and Aue, 2005; Blitzer et al., 2007), which makes the creation of a general sentiment classifier a difficult task.

Teufel et al. (2006) worked on a 2,829 sentence citation corpus using a 12-class classification scheme. While the authors did make use of the context in their annotation, their focus was on the task of determining the author’s reason for citing a given paper. This task differs from citation sentiment detection, which is in a sense a “lower level” of analysis.

Some other recent work has focused on the problem of implicit citation extraction (Kaplan et al., 2009; Qazvinian and Radev, 2010). Kaplan et al. (2009) explore co-reference chains for citation extraction using a combination of co-reference resolution techniques (Soon et al., 2001; Ng and Cardie, 2002). However, the corpus that they use consists of only 94 citations to 4 papers and is likely to be too small to be representative.

For citation extraction, the most relevant work is by Qazvinian and Radev (2010) who proposed a framework of Markov Random Fields to extract only the non-explicit citations for a given paper. They

model each sentence as a node in a graph and experiment with various window boundaries to create edges between neighbouring nodes weighted by lexical similarity between nodes. However, their dataset consists of only 569 citations from 10 papers and their annotation scheme deals with neither acronyms nor sentiment.

6 Discussion

What is the role of citation contexts in the overall structure of scientific context? We assume a hierarchical, rhetorical structure not unlike RST (Mann and Thompson, 1987), but much flatter, where the atomic units are textual blocks which carry a certain functional role in the overall scientific argument for publication (Teufel, 2010; Hyland, 2000). Under such a general model, citation blocks are certainly a functional unit, and their recognition is a rewarding task in their own right. If citation blocks can be recognised along with their sentiment, this is even more useful, as it restricts the possibilities for which rhetorical function the segment plays. For instance, in the motivation section of a paper, before the paper contribution is introduced, we often find negative sentiment assigned to citations, as any indication can serve as a justification for the current paper. In contrast, positive sentiment is more likely to be restricted to the description of an approach which the authors include in their solution, or further develop.

Another aspect concerns which features might help in detecting coherent citation blocks. We have here addressed coherence of citation contexts via certain referring expressions, lexical hooks and also coherence-indicating conjunctions (amongst others). The reintroduction of citation contexts was addressed via lexical hooks. Much more could be done to explore this very interesting question. A more fine-grained model of coherence might include proper anaphora resolution (Lee et al., 2011), which is still an unsolved task for scientific texts, and also include models of lexical coherence such as lexical chains (Barzilay and Elhadad, 1997) and entity coherence (Barzilay and Lapata, 2008).

7 Conclusion

In this paper, we focus on automatic detection of citation sentiment using citation context. We annotate a new large corpus and show that ignoring the citation context would result in loss of a lot of sentiment, specially criticism. We also report the results of the state-of-the-art citation sentiment detection systems on this corpus and when using this context-enhanced gold standard definition.

References

- A. Abu-Jbara and D. Radev. 2011. Coherent citation-based summarization of scientific papers. In *Proc. of ACL*.
- A. Athar. 2011. Sentiment analysis of citations using sentence structure-based features. In *Proc of ACL*, page 81.
- Regina Barzilay and Michael Elhadad. 1997. Using lexical chains for text summarization. In Inderjeet Mani and Mark T. Maybury, editors, *Proceedings of the ACL/EACL-97 Workshop on Intelligent Scalable Text Summarization*.
- Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, (1):1–34.
- Shane Bergsma, Emily Pitler, and Dekang Lin. 2010. Creating robust supervised classifiers via web-scale n-gram data. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 865–874, Uppsala, Sweden, July. Association for Computational Linguistics.
- S. Bird, R. Dale, B.J. Dorr, B. Gibson, M.T. Joseph, M.Y. Kan, D. Lee, B. Powley, D.R. Radev, and Y.F. Tan. 2008. The acl anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In *Proc. of LREC*.
- J. Blitzer, M. Dredze, and F. Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proc. of ACL*, number 1.
- C.C. Chang and C.J. Lin. 2001. LIBSVM: a library for support vector machines, 2001. *Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>*.
- C. Cortes and V. Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.
- I.G. Councill, C.L. Giles, and M.Y. Kan. 2008. Parscit: An open-source crf reference string parsing package. In *Proc. of LREC*, volume 2008. Citeseer.
- Y. EL-Manzalawy and V. Honavar, 2005. *WLSVM: Integrating LibSVM into Weka Environment*. Software available at <http://www.cs.iastate.edu/~yasser/wlsvm>.
- C. Engström. 2004. Topic dependence in sentiment classification. *Unpublished MPhil Dissertation. University of Cambridge*.
- M. Gamon and A. Aue. 2005. Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms. In *Proc. of the ACL*, pages 57–64.
- M. Garzone and R. Mercer. 2000. Towards an automated citation classifier. *Advances in Artificial Intelligence*.
- D. Hall, D. Jurafsky, and C.D. Manning. 2008. Studying the history of ideas using topic models. In *EMNLP*, pages 363–371.
- V. Hatzivassiloglou and K.R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proc. of ACL*, page 181.
- M.J. Hornsey, E. Robson, J. Smith, S. Esposito, and R.M. Sutton. 2008. Sugaring the pill: Assessing rhetorical strategies designed to minimize defensive reactions to group criticism. *Human Communication Research*, 34(1):70–98.
- Ken Hyland. 2000. *Disciplinary Discourses; Social Interaction in Academic Writing*. Longman, Harlow.
- D. Kaplan, R. Iida, and T. Tokunaga. 2009. Automatic extraction of citation contexts for research paper summarization: A coreference-chain based approach. In *Proc. of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*.
- D. Kim, P. Webber, et al. 2006. Implicit references to citations: A study of astronomy papers.
- H. Lee, Y. Peirsman, A. Chang, N. Chambers, M. Surdeanu, and D. Jurafsky. 2011. Stanford’s multi-pass sieve coreference resolution system at the conll-2011 shared task. *ACL HLT 2011*.
- D.D. Lewis. 1991. Evaluating text categorization. In *Proc. of Speech and Natural Language Workshop*, pages 312–318.
- M.H. MacRoberts and B.R. MacRoberts. 1984. The negational reference: Or the art of dissembling. *Social Studies of Science*, 14(1):91–94.
- William C. Mann and Sandra A. Thompson. 1987. Rhetorical Structure Theory: A theory of text organisation. ISI/RS-87-190. Technical report, Information Sciences Institute, University of Southern California, Marina del Rey, CA.
- H. Nanba and M. Okumura. 1999. Towards multi-paper summarization using reference information. In *IJCAI*, volume 16, pages 926–931. Citeseer.
- V. Ng and C. Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proc. of ACL*, pages 104–111.

- J. O'Connor. 1982. Citing statements: Computer recognition and use to improve retrieval. *Information Processing & Management*, 18(3):125–131.
- S.B. Pham and A. Hoffmann. 2004. Extracting positive attributions from scientific papers. In *Discovery Science*, pages 39–45. Springer.
- S. Piao, S. Ananiadou, Y. Tsuruoka, Y. Sasaki, and J. McNaught. 2007. Mining opinion polarity relations of citations. In *International Workshop on Computational Semantics (IWCS)*. Citeseer.
- L. Polanyi and A. Zaenen. 2006. Contextual valence shifters. *Computing attitude and affect in text: Theory and applications*, pages 1–10.
- V. Qazvinian and D.R. Radev. 2008. Scientific paper summarization using citation summary networks. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 689–696. Association for Computational Linguistics.
- V. Qazvinian and D.R. Radev. 2010. Identifying non-explicit citing sentences for citation-based summarization. In *Proc. of ACL*.
- D.R. Radev, M.T. Joseph, B. Gibson, and P. Muthukrishnan. 2009. A Bibliometric and Network Analysis of the field of Computational Linguistics. *Journal of the American Soc. for Info. Sci. and Tech.*
- A. Ritchie, S. Robertson, and S. Teufel. 2008. Comparing citation contexts for information retrieval. In *Proc. of ACM conference on Information and knowledge management*, pages 213–222. ACM.
- W.M. Soon, H.T. Ng, and D.C.Y. Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Comp. Ling.*, 27(4):521–544.
- I. Spiegel-Rosing. 1977. Science studies: Bibliometric and content analysis. *Social Studies of Science*.
- K. Sugiyama, T. Kumar, M.Y. Kan, and R.C. Tripathi. 2010. Identifying citing sentences in research papers using supervised learning. In *Information Retrieval & Knowledge Management, (CAMP), 2010 International Conference on*, pages 67–72. IEEE.
- S. Teufel, A. Siddharthan, and D. Tidhar. 2006. Automatic classification of citation function. In *Proc. of EMNLP*, pages 103–110.
- Simone Teufel. 2010. *The Structure of Scientific Articles: Applications to Citation Indexing and Summarization*. Stanford: CSLI Publications.
- P.D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL*.
- W.J. Wilbur, A. Rzhetsky, and H. Shatkay. 2006. New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC bioinformatics*, 7(1):356.
- H. Yu and V. Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proc. of EMNLP*, page 136.