SENTIMENT BEHIND CITATION

Tracking the need, Current work and Future directions of Citation Sentiment Analysis.

formation science and technology, 65(9), pp.1820-1833.

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

3. Athar, A. and Teufel, S., 2012, June. Context-enhanced citation sentiment detection. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 597-601). Association for Computational Linguistics.

4. Murray, J., 2015. Finding Implicit Citations In Scientific Publications: Improvements to Citation Context Detection Methods.

5. Athar Awais,"Sentiment Analysis of Citations using Sentence Structure-Based Features", Proceedings of the ACL-HLT 2011 Student Session, pages 81–87, Portland, OR, USA 19-24 June 2011.

6. MacRoberts, M.H. and MacRoberts, B.R., 2010. Problems of citation analysis: A study of uncited and seldom-cited influences. Journal of the American Society for Information Science and Technology, 61(1), pp.1-12.

7. Wilson, T., Wiebe, J. and Hoffmann, P., 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational linguistics, 35(3), pp.399-433.

8. Qazvinian, V. and Radev, D.R., 2010, July. Identifying non-explicit citing sentences for citation-based summarization. In Proceedings of the 48th annual meeting of the association for computational linguistics (pp. 555-564). Association for Computational Linguistics.

9. Yu, B., 2013. Automated citation sentiment analysis: What can we learn from biomedical researchers. Proceedings of the American Society for Information Science and Technology, 50(1), pp.1-9.

CITATION ANALYSIS

WHAT?

Citation Analysis is the study of citations in scholarly articles based on their frequency, quality and various other factors.

WHY?

Analyzing citations allow us to detect, analyze and predict
o   Impact of the papers, authors and institutions
o   Trace the interdisciplinary border
o   Access the impact of the research

HOW?

Simplest method of performing citation analysis involves counting the number of times an article has received citations and then measuring its impact.

SENTIMENT ANALYSIS

WHAT?

Sentiment Analysis also known as Opinion Mining is a process of computational linguistics that identifies and categories opinions present in a piece of text based on their polarity.

WHY?

Allows us to understand analyze and predict views of people towards any entity.
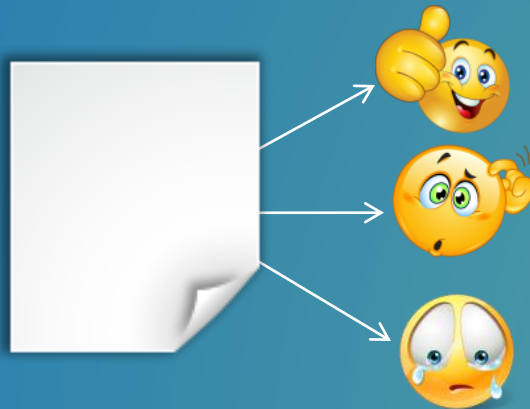
SENTIMENT ANALYSIS MEETS CITATION ANALYSIS

- Citation Sentiment Analysis the task of analyzing the sentiment behind citations.

- Following are some examples of negative, positive and neutral citations.

*Eg: "While SCL has been successfully applied to POS tagging and Sentiment Analysis (Blitzer et al., 2006), its effectiveness for parsing was rather **unexplored**."*	*The window length was motivated by recent research CIT (Qazvinian and Radev, 2010) which **shows the best score** for a four sentence boundary when detecting non-explicit citation.*	*Piao et al. CIT (2007) proposed a system to attach sentiment information to the citation links between biomedical papers by using existing semantic lexical resources.*
(A) Negative Citation	*(B) Positive Citation*	*(C) Neutral Citation*

WHY DO WE NEED CITATION SENTIMENT ANALYSIS?

o   All citations do not mean good news.

Eg: "While SCL has been successfully applied to POS tagging and Sentiment Analysis (Blitzer et al., 2006), its effectiveness for parsing was rather unexplored."

o   Sometimes authors cite out of pity, or courtesy.
o   Since negative citations are possible , citation count may give misleading results.
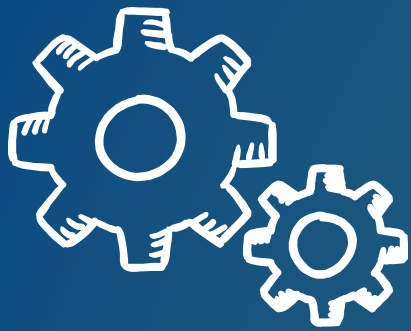
APPROACHES

For Citation Sentiment Analysis

APPROACHES

CITATION SENTIMENT

ANALYSIS

Sentence Level

Context Level

SENTENCE LEVEL CSA

This involves categorizing explicit citation sentences into positive, negative and neutral, only by looking at the sentences in which the citation has been made.

The following example explains it.

While SCL has been successfully applied to POS tagging and Sentiment Analysis (Blitzer et al., 2006), its effectiveness for parsing was rather unexplored."

For this purpose, existing work on co-reference resolution CIT (Lee et al., 2011) *may prove to be useful.*

Explicit Citation

Negative Sentiment

Positive Sentiment

SIGNIFICANT WORKS IN EXPLICIT CITATION SENTIMENT ANALYSIS

Explicit Citation Categorization Performed by Simone Teufel et.al (2012)

12 Categories Classification

1. Weak
2. CoCo-
3. Pmot
4. PUse
5. Pbas
6. PModi
7. Psim
8. PSup
9. CoCoGM
10. CoCoR0,
11. CoCoXY
12. Neut

Explicit Citation Analysis Performed By Athar Awais (2014)

STEP ONE: COLLECTION

Papers from ACL Anthology Network

Annotated Citation Dataset
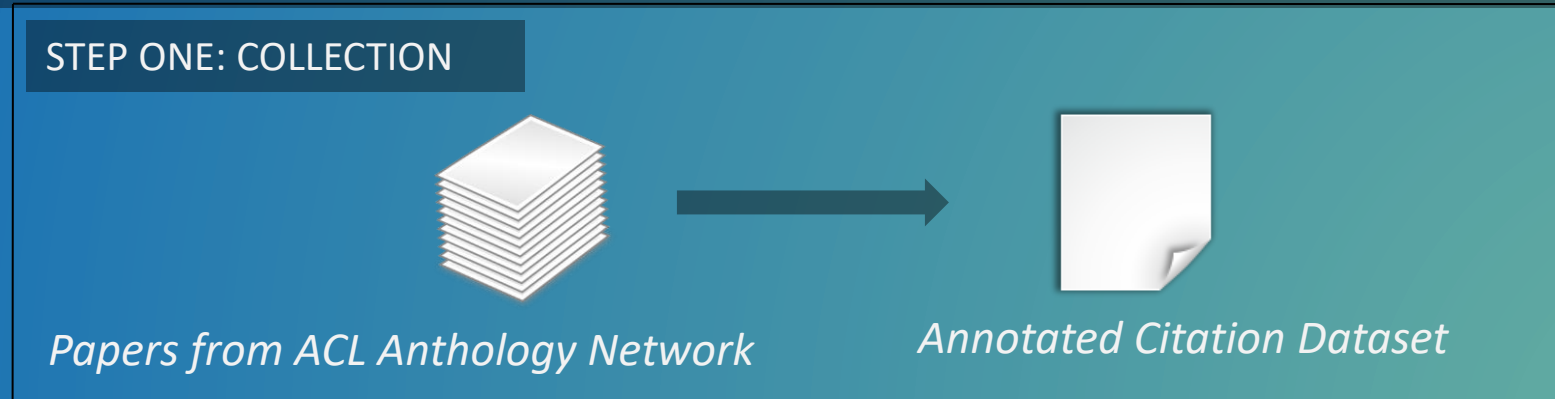
STEP TWO : CLASSIFICATION

Sentence Categorization using Naive Bayes and SVM using WEKA Toolkit

If

1. Direct mention of X's positive attributes.
   OR
2. X improves upon some paper

Positive

1. No positive or negative sentiment involved.

Neutral

1. Direct mention of X's shortcoming.
   OR
2. Some paper improves upon X.
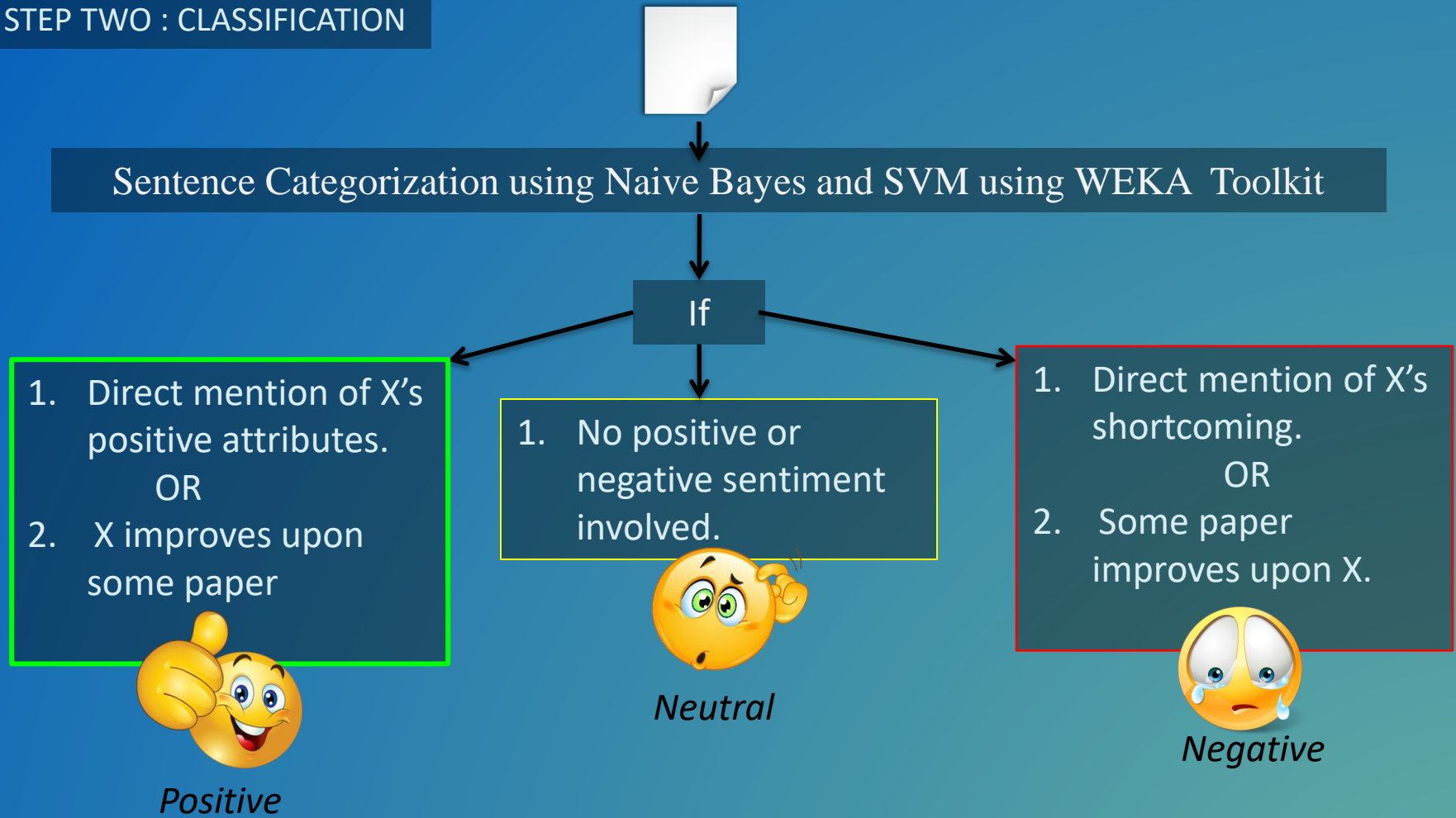
Negative

- Baseline – Trigrams
- Experiment with trigrams and a no. of different features
- Improvement in macro-F score =0.764 with (dependency triplet, window based negation with window length 15

CONTEXT LEVEL CITATION SENTIMENT ANALYSIS

- Context level analysis means analyzing the citation context i.e. the text surrounding the explicit citations.
- Involves two major tasks namely- Context detection and sentiment analysis of context.

STEP ONE:

CONTEXT DETECTION

- Detecting all the explicit and implicit references in the paper.
- Context is used as input for sentiment analysis.

STEP TWO:

SENTIMENT ANALYSIS OF CITATION CONTEXT.

- Classification of citation context into namely three classes- Positive (p) , Negative (n), Neutral or Objective (o).
- An 'x' class for sentences with no citations.

SIGNIFICANT WORKS IN CONTEXT DETECTION

DETECTION BY BUTT (2015)

Extract explicit citation sentences
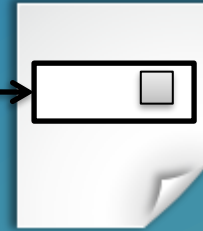
Extract context window of 5 sentences surrounding explicit citations.

DETECTION BY AWAIS ATHAR (2014)

○ are implicit citations

Extract Explicit citation sentences

Extract context window of 4 sentences surrounding explicit citation
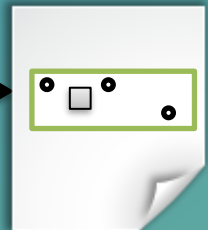
Implicit detected by word features

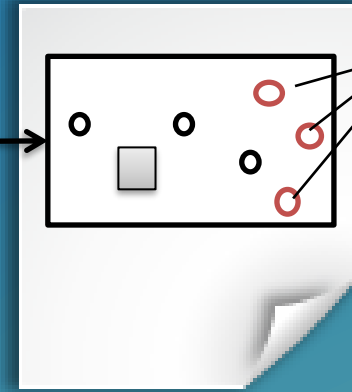# DETECTION BY JONATHAN MURRAY (2015)
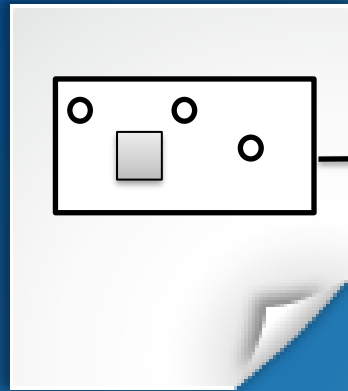
Use Athar's dataset of implicit sentences

By addition of more features, extract implicit sentences.
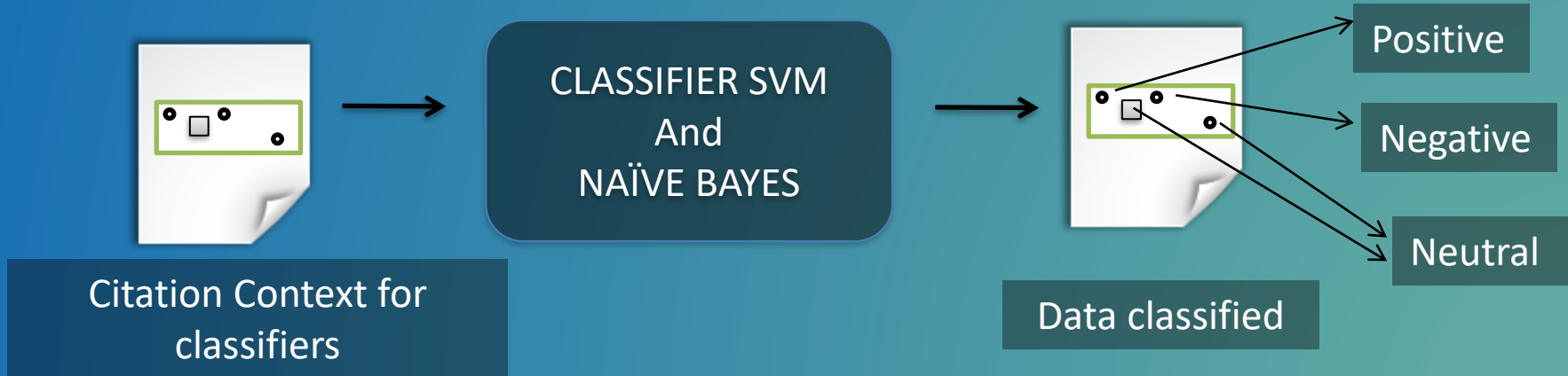
More Implicit Citations obtained

KEY POINTS

(1) In Athar's work, amongst all features he took, lexical hooks and acronyms gave the highest contribution in detecting implicit citations. While closest citations (previous and following) showed an average contribution for the same.

(2) Murray added the features like distance between citations, similarities in titles, content and determiners etc. The similarity in text was the most helpful among these.

SENTIMENT ANALYSIS OF DETECTED CONTEXT

SENTIMENT CLASSIFICATION BY BUTT (2015)

LEXICON

CLASSIFIER NAÏVE BAYES

P

N

Use Sentiment Lexicon to assign polarity to the citation

Classifier determines positively and negatively cited papers.
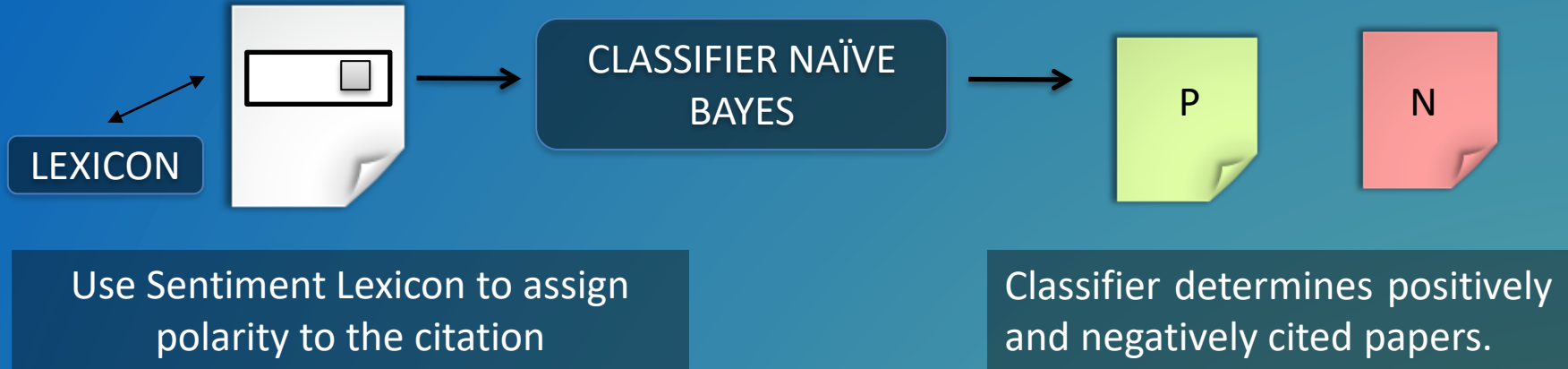
SENTIMENT CLASSIFICATION BY AWAIS ATHAR (2014)

CLASSIFIER SVM
And
NAÏVE BAYES

Positive

Negative

Neutral

Citation Context for classifiers

Data classified

CHALLENGES

- Classifiers and algorithms **not performing so well on citation context.** Results vary for a large corpus which needs immediate work if Citation Sentiment Analysis has to be implemented in a wide research domain as stated by (Athar, A., 2014).
- Recent studies have **not placed enough importance** on the researcher's viewpoint (Yu, B., 2013).
- Most of the inspiring research works are not cited or are rarely cited (MacRoberts, M.H. and MacRoberts, B.R., 2010).
- Even after detecting implicit citations, the problem of analyzing and **identifying the true sentiments is yet to be accurately addressed** (Athar, A. and Teufel, S., 2012).

FUTURE SCOPE

- **Better annotations** will result in better results (Teufel, S., Siddharthan, A. and Tidhar, D., 2006)

- For efficient detection of context, **more features** can be added such that they perform equally well on any corpus domain.

- Need for detecting **citations outside the context window**.

- More work can be done on **assignment of weights** on citations which is also done by Wan, Xiaojun, and Fang Liu(2014).

- **Automatic summarization** it is another promising avenue for further research.

CONCLUSION

o From the work performed till now, it can easily be concluded that **studying of context is important** for Citation Sentiment Analysis.

o Murray's, Butt's and Awais's attempt to **identify implicit citations** also shows the importance of citation context and implicit citations.

o We can see that different approaches have been used yet, there is **no method robust enough.**

o **Citation sentiment analysis is a budding field which needs to be explored thoroughly to fully utilize its power in the research arena**.

REFERENCES

- Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X. and Zhai, C., 2014. Content-based citation analysis: The next generation of citation analysis. *Journal of the association for information science and technology*, *65*(9), pp.1820-1833.
- Athar, A., 2014. Sentiment analysis of scientific citations. Technical Report, University of Cambridge, Computer Laboratory,(UCAM-CL-TR-856).
- Athar, A. and Teufel, S., 2012, June. Context-enhanced citation sentiment detection. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 597-601). Association for Computational Linguistics.
- Murray, J., 2015. Finding Implicit Citations in Scientific Publications: Improvements to Citation Context Detection Methods.
- Athar Awais,"Sentiment Analysis of Citations using Sentence Structure-Based Features", Proceedings of the ACL-HLT 2011 Student Session, pages 81–87, Portland, OR, USA 19-24 June 2011.

REFERENCES

- MacRoberts, M.H. and MacRoberts, B.R., 2010. Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, *61*(1), pp.1-12.
- Wilson, T., Wiebe, J. and Hoffmann, P., 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, *35*(3), pp.399-433.
- Qazvinian, V. and Radev, D.R., 2010, July. Identifying non-explicit citing sentences for citation-based summarization. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 555-564). Association for Computational Linguistics.
- Yu, B., 2013. Automated citation sentiment analysis: what can we learn from biomedical researchers. *Proceedings of the American Society for Information Science and Technology*, *50*(1), pp.1-9.
- Wan, Xiaojun, and Fang Liu. "Are all literature citations equally important? Automatic citation strength estimation and its applications." Journal of the Association for Information Science and Technology 65.9 (2014): 1929-1938.

REFERENCES

- Kazi, P., Patwardhan, M. and Joglekar, P., 2016. Towards a new perspective on context based citation index of research articles. *Scientometrics*, *107*(1), pp.103-121.
- Kaplan, D., Tokunaga, T. and Teufel, S., 2016. Citation Block Determination using Textual Coherence. *Journal of Information Processing*, *24*(3), pp.540-553.
- Butt, B.H., Rafi, M., Jamal, A., Rehman, R.S.U., Alam, S.M.Z. and Alam, M.B., 2015. Classification of Research Citations (CRC). *arXiv preprint arXiv:1506.08966*.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
- Teufel, S., Siddharthan, A. and Tidhar, D., 2006, July. Automatic classification of citation function. In Proceedings of the 2006 conference on empirical methods in natural language processing (pp. 103-110). Association for Computational Linguistics.
- Abu-Jbara, A., Ezra, J. and Radev, D.R., 2013. Purpose and Polarity of Citation: Towards NLP-based Bibliometrics. In HLT-NAACL (pp. 596-606).
- Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M., 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. ICWSM, 10, pp.178-185.