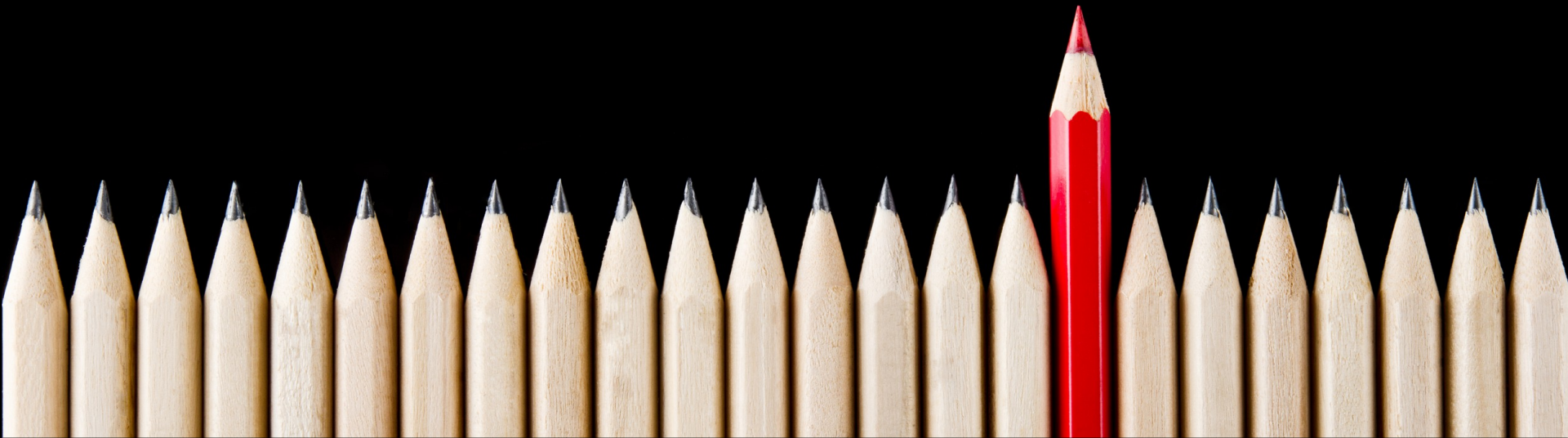


# Incidental or influential: Challenges in automatic detection of citation importance



David Pride and **Petr Knoth**  
Knowledge Media institute, The Open University, UK

**All citations are equal.  
But some are more equal than others...**

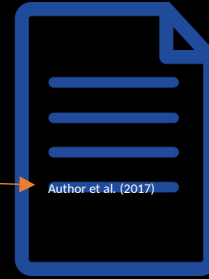


# Introduction

- Current quantitative research evaluation methods are largely based on citation counts.
  - : Journal Level – Journal Impact Factor (JIF)
  - : Author Level – *h-index*, *g-index*
- None of these metrics account for citation type or sentiment.
- Open Access means increased availability of full-text papers and articles for analysis.

# Citation Context Analysis

- Discover where the citation occurs in the full text of a document.



- Identifies the type, sentiment polarity or influence of the citation.



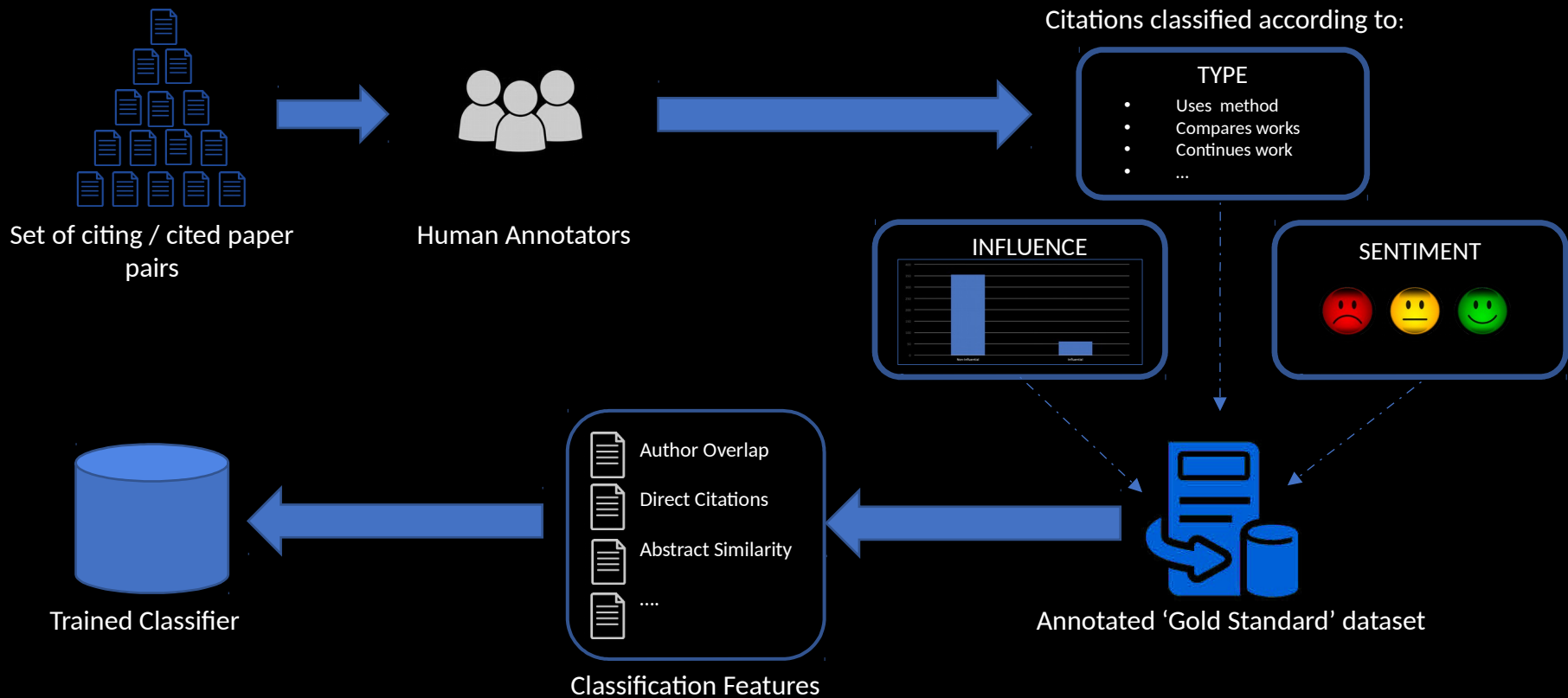
# Why do we cite something?

- Giving credit for related work
- Identifying methodology / equipment
- Providing background reading
- Correcting one's own work
- Correcting the work of others
- Criticizing previous work
- Substantiating claims
- Disputing priority claims of others
  - negative claims
- Providing leads to poorly disseminated, poorly indexed, or uncited work
- Authenticating data and classes of fact-physical constants.
- Identifying original publications in which an idea or concept was discussed.
- Identifying original publications or other work describing an eponymic concept or term
- Disclaiming work or ideas of others
  - negative homage

# Methodology

- Review of previous citation classification studies (Zhu, 2015; Valenzuela, 2015; Teufel, 2006).
- Comparative analysis of two of these studies (Zhu, 2015; Valenzuela, 2015)
- Goals:
  - Understand features and datasets used.
  - Identify which features perform best at identifying citation influence.
  - Investigate reproducibility of these studies.

# Training a Citation Classification Model

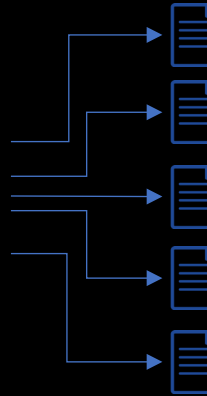


# Citation Classification Workflow

INPUT: Paper X



Citation Extraction



Citing / Cited Paper Pairs

[1] Knoth, P., Anastasiou, L., Charalampous, A., Cancellieri, M., Pearce, S., Pontika, N., Bayer, V.: Towards effective research recommender systems for repositories. In: Proceedings of Open Repositories 2017

[3] .....

[4] .....

[n] .....

Paper, Citation, Label

X, [1], incidental  
X, [2], incidental  
X, [3], influential  
X, [4], incidental  
X, [n], .....

Classifier

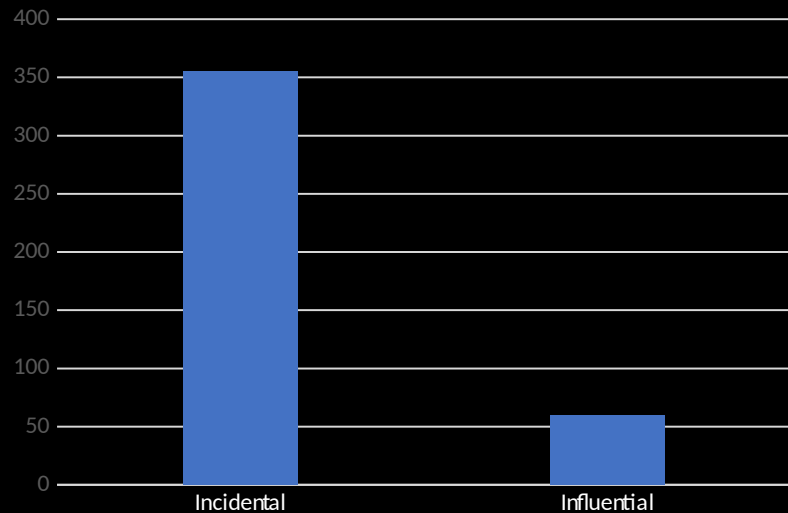
Author Overlap  
Direct Citations  
Abstract Similarity  
....

Feature Extraction



# Ground Truth Dataset

- 2 Annotators – binary influential / important judgements.
  - 465 Cited / Citing Pairs
  - ~15% of all citations are influential / important
  - ~4% of all citations are negative



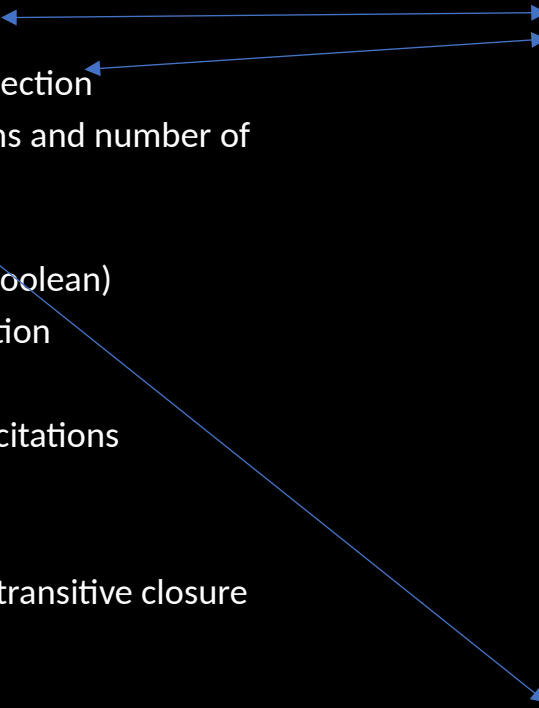
# Selection of classification features

## Valenzuela et al. Features

- F1 Total number of direct citations
- F2 Number of direct citations per section
- F3 Total number of indirect citations and number of indirect citations per section
- F4 Author overlap (Boolean)
- F5 Citation is considered helpful (Boolean)
- F6 Citation appears in table or caption
- F7 1 / Number of references
- F8 Number of paper citations / all citations
- F9 Similarity between abstracts
- F10 PageRank
- F11 Number of citing papers after transitive closure
- F12 Field of cited paper.

## Zhu et al. Features

- 1.1 countsInPaper\_whole
- 1.2 countsInPaper\_secNum
- 1.3 countsInPaper\_related
- 1.4 countsInPaper\_intro
- 1.5 countsInPaper\_core
- 2.1 sim\_titleTitle
- 2.2 sim\_titleCore
- 2.3 sim\_titleIntro
- 2.4 sim\_titleConcl
- 2.5 sim\_titleAbstr
- 2.6 sim\_contextTitle
- 2.7 sim\_contextIntro
- 2.8 sim\_contextConcl
- 2.9 sim\_contextAbstr
- 3.1 contextMeta\_authorMentioned
- 3.2 contextMeta\_appearAlone
- 3.3 contextMeta\_appearFirst
- 3.4 contextLex\_relevant
- 3.5 contextLex\_recent
- 3.6 contextLex\_extreme
- 3.7 contextLex\_comparative
- 3.8 contextLexOsg\_wnPotency
- ....
- 5.1 aux\_citeCount
- 5.2 aux\_selfCite
- 5.3 aux\_yearDiff



## 2. Selection of classification features

Valenzuela et al. Features

- F1 Total number of direct citations

**Fewer than half of these features performed better than the baseline.**

(Valenzuela et al. 2015)

- F8 Number of paper citations / all citations
- F9 Similarity between abstracts
- F10 PageRank
- F11 Number of citing papers after transitive closure
- F12 Field of cited paper.

Zhu et al. Features

- 1.1 countsInPaper\_whole
- 1.2 countsInPaper\_secNum
- 1.3 countsInPaper\_related
- 1.4 countsInPaper\_intro
- 1.5 countsInPaper\_core
- 2.1 sim\_titleTitle
- 2.2 sim\_titleCore
- 2.3 sim\_titleIntro
- 2.4 sim\_titleConcl
- 2.5 sim\_titleAbstr
- 2.6 sim\_contextTitle
- 2.7 sim\_contextIntro
- 2.8 sim\_contextConcl
- 2.9 sim\_contextAbstr
- 3.1 contextMeta\_authorMentioned
- 3.2 contextMeta\_appearAlone
- 3.3 contextMeta\_appearFirst
- 3.4 contextLex\_relevant
- 3.5 contextLex\_recent
- 3.6 contextLex\_extreme
- 3.7 contextLex\_comparative
- 3.8 contextLexOsg\_wnPotency
- ....
- 5.1 aux\_citeCount
- 5.2 aux\_selfCite
- 5.3 aux\_yearDiff

# Selection of classification features

Valenzuela et al. Features

Zhu et al. Features

- F1 Total number of direct citations

- ▶ 1.1 countsInPaper\_whole
- ▶ 1.2 countsInPaper\_secNum

**Fewer than half of these features performed better than the baseline.**

(Valenzuela et al. 2015)

**Of 40 features, a combination of just FOUR features provided the best performance.**

(Zhu et al. 2015)

- F8 Number of paper citations / all citations
- F9 Similarity between abstracts
- F10 PageRank
- F11 Number of citing papers after transitive closure
- F12 Field of cited paper.

- 3.1 contextMeta\_authorMentioned
- 3.2 contextMeta\_appearAlone
- 3.3 contextMeta\_appearFirst
- 3.4 contextLex\_relevant
- 3.5 contextLex\_recent
- 3.6 contextLex\_extreme
- 3.7 contextLex\_comparative
- 3.8 contextLexOsg\_wnPotency
- ....
- ▶ 5.1 aux\_citeCount
- ▶ 5.2 aux\_selfCite
- ▶ 5.3 aux\_yearDiff

# Irreproducible features

F5 - Citation is considered helpful (Boolean)

How is 'considered helpful' defined? No cue phrases provided.

F10 - PageRank

Based on what corpora - again, details not provided.

F12 - Field of cited paper.

This feature is not complete.

# Reproducible features

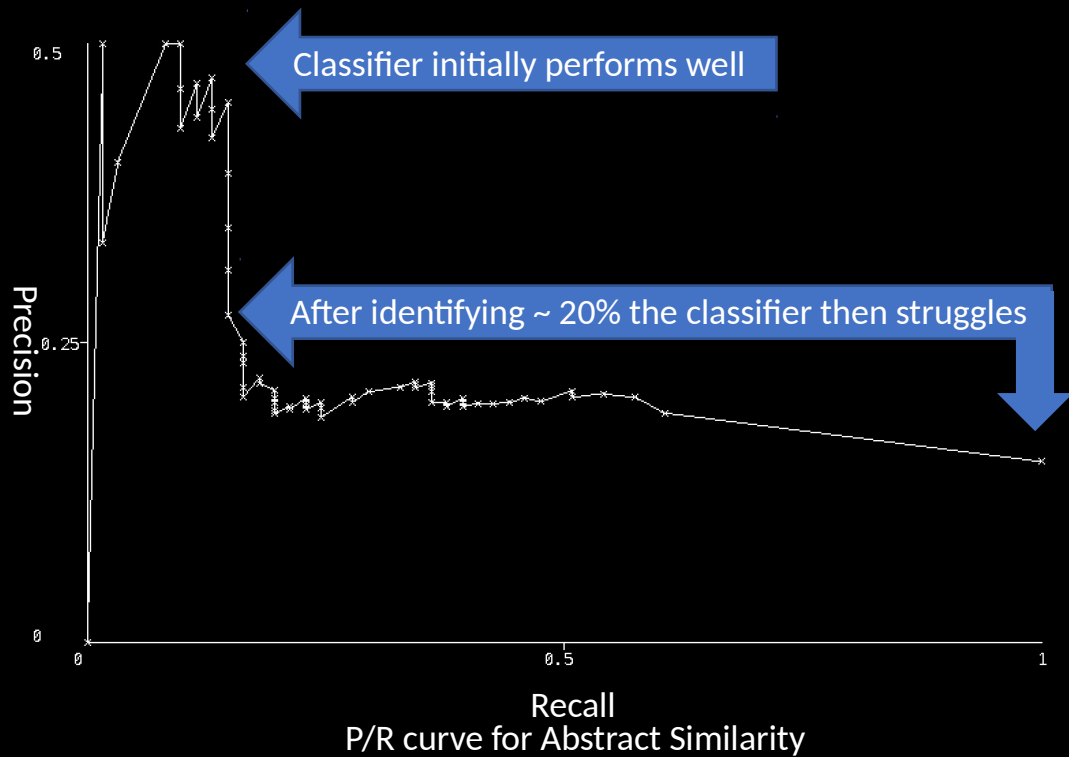
F1 – Number of Direct Citations / 'countsinPaper\_Whole'

F4 – Author Overlap / auxSelfCite

F10 – Abstract Similarity

# Evaluation

- Valenzuela measures Precision @ R 0.90
- Masks some predictive ability of features.
- Zhu measure in terms of Pearson  $r$  correlation.
- Our study shows results in both formats.
- Random Forest Classifier = best results



# Results of experiments

Feature	Precision@Recall=0.9		Pearson $r$	
	Valenzuela et al. [1]	Our results	Zhu et al. [4]	Our results
Direct Citations	<b>0.30</b>	<b>0.21</b>	<b>0.330</b>	0.281
Abstract Similarity	0.14	0.14	N/A	<b>0.373</b>
Author Overlap	0.22	0.16	0.020	0.132

- Features tested using Valenzuela dataset
- Results measured in terms of P/R and Pearson  $r$
- Difference in Author Overlap – different datasets
- Abstract Similarity shows highest  $r$  value of tested features



# The challenges

- Lack of large 'ground truth' dataset for training classifiers.
- Complex or irreproducible features.
- PDF Extraction issues.

# Conclusions

- Lack of massive scale gold-standard dataset.
- Raises questions regarding publication of datasets as well as results.
- Abstract Similarity shown to be better predictor of citation influence than demonstrated by earlier studies.
- Serious concerns with reproducibility of previously tested features.
- Significant variances in quality of PDF extraction tools.

# Thank you for listening



For full details of the work being done by CORE and KMi visit:

<http://www.core.ac.uk>

<http://www.kmi.open.ac.uk>

[petr.knoth@open.ac.uk](mailto:petr.knoth@open.ac.uk)

[david.pride@open.ac.uk](mailto:david.pride@open.ac.uk)

# Citation Classification Schemes

Teufel, S., Siddharthan, A., & 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.

Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). *Measuring academic influence*. *Journal of the Association for Information Science and Technology*, 66(2), 408-427.

Valenzuela, M., Ha, V., & Etzioni, O. (2015, April). *Identifying Meaningful Citations*. In *AAAI Workshop: Scholarly Big Data*.