

A Corpus of Sentence-level Revisions:
Towards Understanding Statement Strength in
Communication

Chenhao Tan, Lillian Lee
Cornell University

Kunming Attack



The members of the Security Council (UN) condemned in the strongest terms the terrorist attack on March 1, 2014 in Kunming Train Station

[Chinese media] accused Western media of soft-pedaling the attack and failing to state clearly that it was an act of terrorism.” [The New York Times]

“Some Western media, including CNN, The Associated Press, The New York Times and The Washington Post, were mystifying, confusing, even to the point of sowing discord.”

‘Completely hypocritical and callous,’ [People’s daily]

In particular ...

..., the US embassy referred to this incident as the “terrible and senseless act of violence in Kunming”.

After Prodding, U.S. State Department Labels Kunming Attack ‘Terrorism’

By DIDI KIRSTEN TATLOW MARCH 4, 2014 1:05 AM 19 Comments

A weibo user: “If you say that the Kunming attack is a ‘terrible and senseless act of violence’, then the 9/11 attack can be called a ‘regrettable traffic incident’”

Understanding statement strength is important!



EDUCATIONFORUM

EDUCATION

Open Learning at a Distance: Lessons for Struggling MOOCs

Patrick McAndrew and Eileen Scanlon

Five education is changing how people think about learning online. The rise of Massive Open Online Courses (MOOCs) (1) shows that large numbers of learners can be reached. It also raises questions as to how effectively they support learning (2). There is a timeliness in the introduction of MOOCs, reflecting the right combination of online systems, interest from good teachers in reaching more learners, and banks of digital resources, predicted as a "perfect storm of innovation" (3). However, learning at scale, at a distance, is not a new phenomenon. Seeing MOOCs narrowly as a technology that expands access to in-classroom teaching can miss opportunities. Drawing on decades of lessons learned, we set out aims to help spur innovation in science education. Education based on gathering people together into a physical location is limited to those who can afford it and who make it past the filters that attenuate participation in higher levels of education. Those filters are inevitable on cost grounds to meet global needs "would require four major campus universities... to open every week" (4). The arrival of MOOC highlights that there are alternatives. With courses enrolling over 100,000 students, MOOCs can reach students who have breaks in study, change where they study, mix study with work, and take at least part of their study online. Such students are now the majority, forming more than 70% are now in U.S. post-secondary education (5).

Recommendations for Open Learning
We ought not behave as if learning at scale is unexplored territory and that there is no previous experience in being massive, open, or even online, upon which to build. Distance universities, such as The Open University



(OU) established in Britain more than 40 years ago, from their inception, ran courses for thousands of learners, accepted open entry, and led the move into online methods of teaching and learning. In each case, they provide lessons likely to apply in the new context of MOOCs.

Build on distance-learning pedagogy: Some of the steps taken toward "massive" classes simply follow the observation that a lecture presented to a few hundred students can be viewed by many more once put on the Web. But numbers of views and downloads of PowerPoint do not mean learners have engaged. Effective distance-learning pedagogies that lead the learner through tasks at scales that cannot be achieved in face-to-face classes. A classic challenge for distance learning is "would you teach surgery?" The University of Edinburgh now does just that with support from tutors and assessment, has enabled 1.6 million people (7) to complete university level courses without the need for most initial entry requirements. Teaching at a distance combines media to motivate and entice, including television programs broadcast through the BBC, experiment kits

Support for nontraditional students, team-based quality control, and assessment design are critical.

Laboratory builds a collection of tools to combine remote access, virtual experiments, and citizen science (8) into the curriculum. Advice: Interactions between student-teacher, student-student, and student-materials all can act to support learners (9). Paying attention to the content, and building materials that do the teaching (10), allows direct contact between teacher and learner to be reduced. Structured tasks guide the learner. Working online offers the chance to build in interactivity. Preparation to help learners who need support. Plans to help learners who need support. "Open" is not the same as "free." Openness means accepting those who want to learn as well as those who study to learn. Learning is challenging, so helping students is essential. Some people will manage on their own, but that is not enough for genuinely inclusive education. The self-paced, location-independent properties of online learning make it attractive to the marginalized and those with disabilities (11). Rapid fall-out identified in many MOOCs (12), where only 10% of those who register may complete the course, reflects common challenges. How we approach support for learners influences retention. Early contact with a tutor prevents drop-out, and student attitudes toward the tutor matter (13). For large-scale operation, tutors focus on effective and timely feedback to learners. Support is particularly important as activities start. Submission of the first assignment predicts eventual success with a course. Advice: A vital step in coping with access is to recognize the importance of sup-

motivations for participants, how to scale up to genuinely massive access to learning, and how best to assess learning. The opportunity for experimentation gives us the chance to learn more ourselves, as well as to educate others.

Distance universities, from their inception, ran courses for thousands of learners, accepted open entry, and led the move into online methods of teaching and learning.

References and Notes
1. J. B. Aronson, *Science*, 2012, 346, 1210-1211.
2. C. A. Lounsbury, *J. Educ. Res.*, 2012, 114, 10-15.
3. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
4. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
5. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
6. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
7. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
8. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
9. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
10. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
11. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
12. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.
13. S. J. Liebowitz, *J. Econ. Surv.*, 2012, 26, 1-15.

We regret to inform you that your paper has been rejected

A first step to understand statement strength is to distinguish strong and weak statements.

Statement strength is inherently relative.



Cornell University
Library

arXiv.org

Authors post latex source for different versions of the same paper

arXiv.org > math > arXiv:1109.4363

Mathematics > Probability

The Segregated Lambda-coalescent

Nic Freeman

(Submitted on 20 Sep 2011 (v1), last revised 2 Nov 2013 (this version, v3))

Submission history

From: Nic Freeman [[view email](#)]

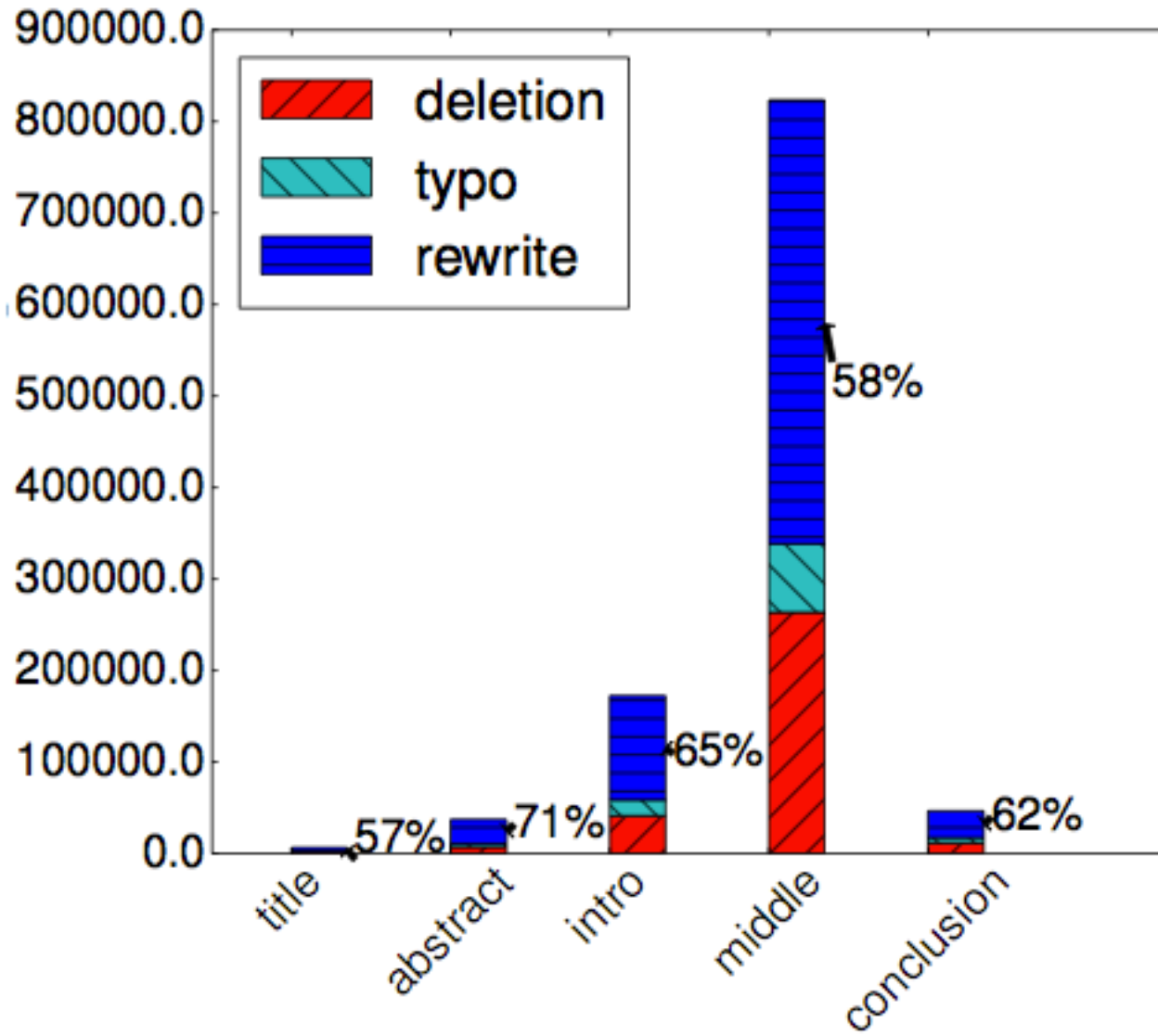
[v1] Tue, 20 Sep 2011 17:13:33 GMT (125kb,D)

[v2] Wed, 9 Nov 2011 15:03:40 GMT (137kb,D)

[v3] Sat, 2 Nov 2013 22:15:38 GMT (108kb,D)

Is it only typos?

A lot of rewrites are made between different versions



Align different versions of the same paper to find sentence pairs

[Barzilay and Elhadad 2003]

Phase transitions in a spatial coalescent

Nic Freeman

(Submitted on 20 Sep 2011 (this version), latest version 2 Nov 2013 (v3))

We construct a natural extension of the Lambda-coalescent to a spatial continuum, and analyse its behaviour.

Like the Lambda-coalescent, at any time $t > 0$ the individuals in our model can be separated into (i) a dust component and (ii) large blocks of coalesced individuals.

We identify a five phase system, where our phases are defined according to changes in the qualitative behaviour of the dust and blocks. We completely classify the phase behaviour, and obtain necessary and sufficient conditions for the model to come down from infinity.

The Segregated Lambda-coalescent

Nic Freeman

(Submitted on 20 Sep 2011 (v1), last revised 2 Nov 2013 (this version, v3))

We construct an extension of the Lambda-coalescent to a spatial continuum and analyse its behaviour. Like the Lambda-coalescent, the individuals in our model can be separated into (i) a dust component and (ii) large blocks of coalesced individuals. We identify a five phase system, where our phases are defined according to changes in the qualitative behaviour of the dust and large blocks. We completely classify the phase behaviour, including necessary and sufficient conditions for the model to come down from infinity.

Examples of potential strength changes

The algorithm is
studied in this paper .

... circadian pattern
and burstiness in
human communication
activity .

The algorithm is
proposed in this paper .

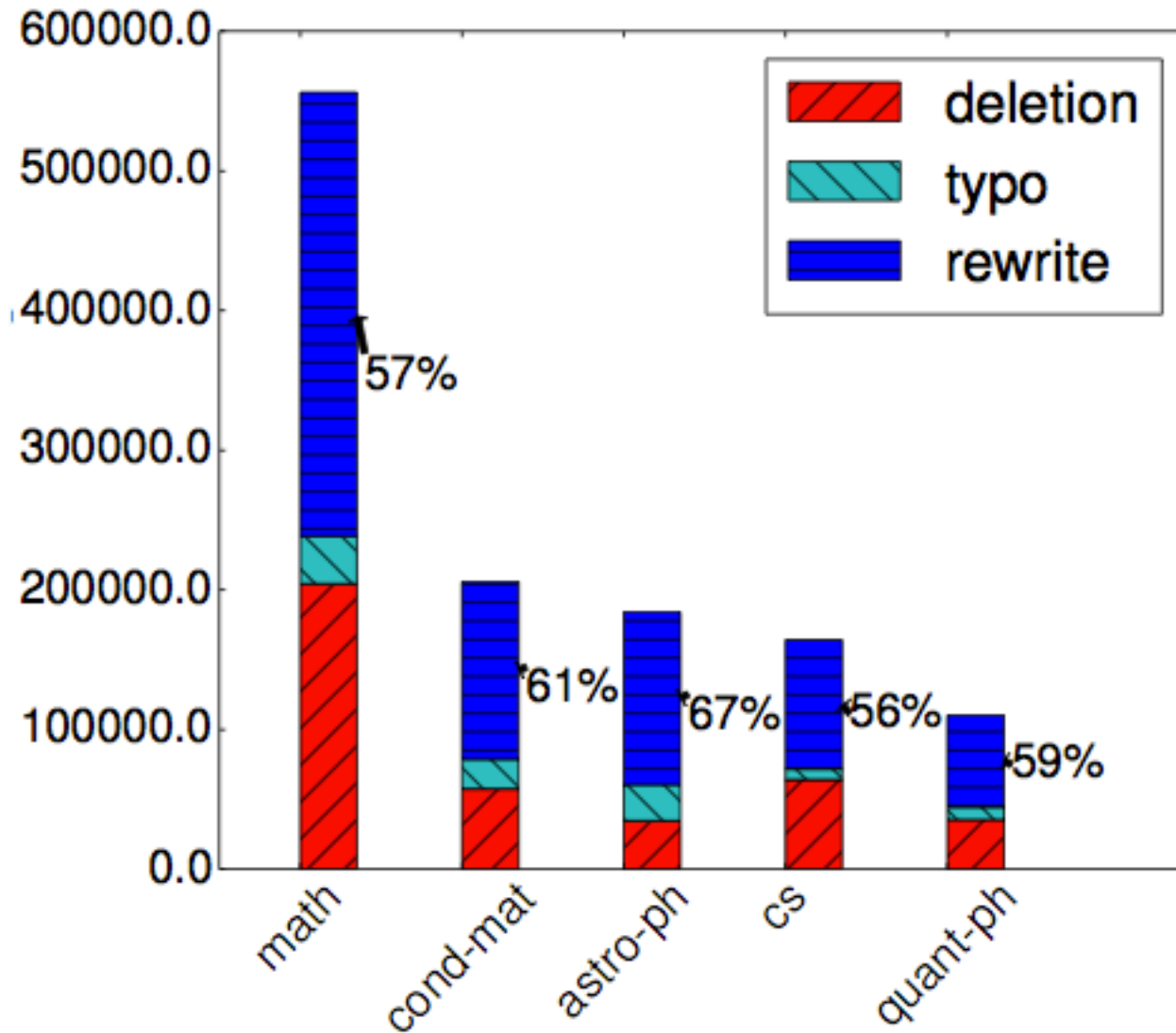
... circadian pattern and
burstiness in *mobile*
phone communication .

Examples of potential strength changes

they maximize the expected revenue of the seller but *induce efficiency loss* .

they maximize the expected revenue of the seller but *are inefficient* .

Top categories in making changes



A corpus of sentence-level revisions focusing on potential strength changes

- 108K pairs from abstracts or introductions
 - similarity score for the pair was larger than 0.5
- Final labeling instructions:
stronger, weaker, no strength change, I can't tell
- Labeled 500 pairs on Amazon Mechanical Turk
 - 9 labels and *COMMENTS* each

Overall labeling results

- Among the 500 pairs, Fleiss' Kappa was 0.242, which indicates fair agreement
- 386 pairs have an absolute-majority label
Fleiss' Kappa is 0.322, and 74.4% of pairs were strength changes
(93 weaker, 194 stronger, 99 no change)
- Most labels agree with our intuitions, but there are also some differences

Participants are swayed by specificity

S1: ... using data from numerics and experiments .

S2: ... using data sets from numerics in the point particle limit and one experimental data set .

S2 is stronger: “S2 is more specific in its description which seems stronger.”

S2 is weaker: “‘one experimental data set’ weakens the sentence”

Similar findings in courts [Bell and Loftus (1989)]

Participants interpret constraints/conditions not in strictly logical ways

S1: we also proved that if
[MATH] is sufficiently
homogeneous then ...

S2: we also proved that if
[MATH] is *not totally
disconnected* and sufficiently
homogeneous then ...

(stronger) We have more detail/proof in S2

(stronger) the words "not totally disconnected" made the sentence sound more impressive.

Participants can have a different understanding of domain-specific terms

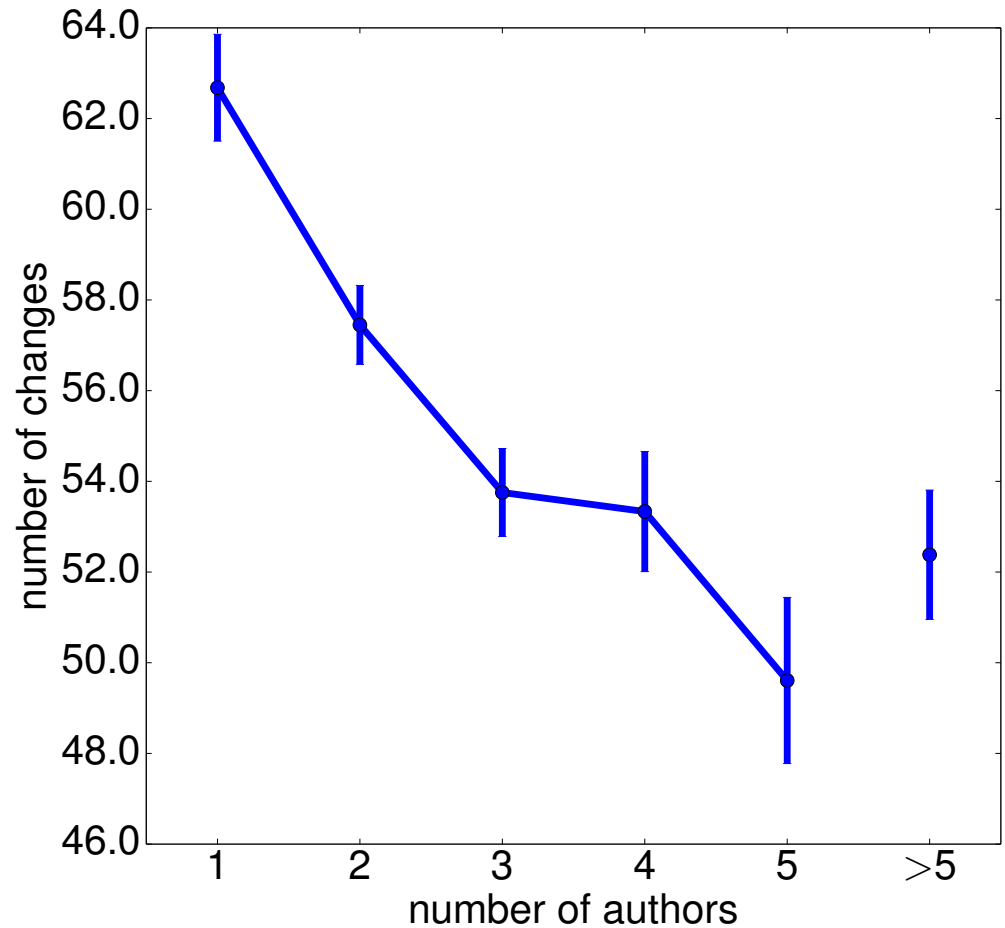
S1: in the current paper we *discover* several variants of qd algorithms for quasiseparable matrices .

S2: in the current paper we *adapt* several variants of qd algorithms to quasiseparable matrices .

S2 is stronger: “in S2 Adapt is stronger than just the word discover. adapt implies more of a proactive measure.”

This type of corpus can enable other interesting studies

The more authors,
the fewer changes!



Conclusion

- A corpus of sentence-level revisions to study statement strength is available at <http://chenhaot.com/pages/statement-strength.html>
- The labels and *comments* we collected can hopefully provide insights into better ways to define and approach this problem.
- The ultimate goal of our study is to understand the effects of statement strength on the public, which can lead to various applications in public communication.