



Wei Zhong <zhongwei@udel.edu>

ICTIR 2015: Notification

ICTIR 2015 PC Chairs <ictir2015@conftool.net>

Tue, Jul 14, 2015 at 7:54 AM

Reply-To: pcchairs@ictir2015.org

To: zhongwei@udel.edu

Cc: clock126@126.com, zhongwei@udel.edu, hfang@udel.edu

Dear Dear Wei Zhong,

We are sorry to inform you that your paper

ID: 35

Title: A Novel Similarity-Search method for Mathematical Content in LaTeX Markup

has not been accepted for the ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR 2015).

Competition for inclusion in the program was heavy; given a relatively high number of strong submissions, we have ended up with acceptance rates of 29 out of 57 long papers and 22 out of 43 short papers.

The results of the reviewing phase for your paper are included in the remainder of this email. Please read them carefully; they provide feedback and suggestions that should allow you to improve your paper. Based on the reviews, you may wish to revise your paper to address the concerns raised by the program committee members that assessed your paper, and consider submission to another venue. Depending on the topic of your paper and the level of support it received via the reviewer scores, you might consider submission to CHIIR (<http://sigir.org/chiir2016/>) or ECIR (<http://ecir2016.dei.unipd.it/>), both of which have submission deadlines in the early fall.

We would of course be glad to welcome you to the conference in September; helpful information for attending ICTIR 2015 is found at <http://ictir2015.org/attendees>.

Thank you for your submission to ICTIR,

Best regards,

Norbert, Yi, Cheng and Arjen

REVIEW RESULT OF THE PROGRAM COMMITTEE:

We are sorry to report that this contribution could not be accepted.

OVERVIEW OF REVIEWS

Review 1

=====

Contribution of the submission

This paper's main contribution is a set of techniques for defining the similarity between two LaTeX math expressions. The similarity measure is used to allow a system to rank candidate expressions against a query expression.

Evaluation of the contribution

Relevance (10%): 4
Reproducibility (10%): 4
Originality (10%): 3
Quality (10%): 2
Citations (10%): 4
Presentation (10%): 2
Impact (10%): 2
Resources (10%): 4
Overall Recommendation (20%): 2
Total points (out of 5) : 2.9

Comments for the authors

This paper outlines a novel approach to ranking mathematical expressions written in LaTeX against a query (also expressed as a LaTeX math expression). In other words, the paper's main contribution is a set of techniques for defining the similarity between two TeX expressions. The main claim of the paper is that the proposed method, based on a tree representation of each expression, allows important dimensions of similarity to be exposed for search.

The authors' ideas--decomposing expressions into leaf-root paths--is intuitively appealing and certainly seems plausible.

However, after reading the paper, I'm not totally clear about the authors' motivations. Surely there are simpler ways to utilize the proposed tree representations than what we see in Section 3. The expansiveness and turgidity of this section would be OK if we came away from it with a method that was provably optimal in some sense (or otherwise desirable). But in Section 3.4 it seems that ultimately, ranking decisions are made based on some very simple heuristics. Thus the paper's overarching contribution is hard to discern.

A second concern I have has to do with the experimental evaluation. Math IR is not my main area, and so I'm not deeply familiar with available test collections and experimental best-practice. But the authors note that NTCIR generated a collection for precisely this task; they couldn't use it because their method relies on the presence of TeX markup. This seems like a big disappointment. The "homegrown" collection that they do test on isn't compelling (e.g. it contains only 20 queries, and it's not clear how relevance assessment was done). Ultimately, this fine points of the collection don't matter tremendously, though, because the discussion of the experimental results is cursory.

In sum, this paper expresses an interesting idea. But I feel that the authors haven't done their idea justice, both rhetorically and experimentally speaking. I do concede, though, that this subject lies well outside my own area of expertise.

Review 2

=====

Contribution of the submission

The paper presents a method for measuring the similarity of two mathematical expressions based on (i) their structural similarity (i.e. similar operations performed over potentially different variables) and (ii) their symbolic similarity (i.e. the degree to which the variables overlap between the two expressions)

The approach is evaluated over a newly-assembled corpus from StackExchange and compared to a simplified version of itself. The authors claim satisfactory accuracy in their evaluation.

Evaluation of the contribution

Relevance	(10%): 4
Reproducibility	(10%): 3
Originality	(10%): 3
Quality	(10%): 2
Citations	(10%): 4
Presentation	(10%): 3
Impact	(10%): 2
Resources	(10%): 3
Overall Recommendation (20%):	2
Total points (out of 5)	: 2.8

Comments for the authors

The paper is reasonably well-written, but the ideas are not always easy to grasp from the text. There is a very good overview of related work, and the problem itself is well-explained and well-motivated. The ideas behind the authors' approach appear sensible, but I cannot judge whether it is novel.

The weakest part of the paper is evaluation (which is important, as this paper targets a specific application area, and is not a theoretical contribution). The authors choose to use their own dataset instead of the NTCIR benchmark collection. I understand the practical argument for wanting to parse latex, but see no reason why experiments could not have been performed over both datasets, allowing the accuracy of the proposed method to be compared to the previously-published results. Furthermore, the authors seem to compare their results to a simplified version of their own algorithm, rather than any state-of-the-art approach. Finally, there is insufficient description of how the queries for the dataset were developed, whether they were based on actual information needs, and how relevance judgements were collected (annotators, instructions, Kappa, adjudication, etc.)

Review 3

=====

Contribution of the submission

TASK ORIGINALITY

Retrieval based on an equation is a relatively new task, previously studied in the context of NTCIR 10 and 11 Math Tasks.

ALGORITHMIC CONTRIBUTIONS

The authors present novel equation similarity metrics which, although based on latex, offer insights into how to perform matching of structured content in general.

EXPERIMENTAL CONTRIBUTIONS

There is comparison to a naive baseline and, given the newness of the domain, reimplementing of other NTCIR runs may have been difficult. However, I would have liked to see experiments on NTCIR 11 data.

Evaluation of the contribution

Relevance (10%): 5
Reproducibility (10%): 4
Originality (10%): 4
Quality (10%): 4
Citations (10%): 3
Presentation (10%): 4
Impact (10%): 3
Resources (10%): 3
Overall Recommendation (20%): 3
Total points (out of 5) : 3.6

Comments for the authors

SUMMARY

The authors present a method for ranking documents in response to query equations. Experiments are run on queries from the NTCIR 10 Math Task on a new corpus of documents including (latex) math markup.

TASK ORIGINALITY

Retrieval based on an equation is a relatively new task, previously studied in the context of NTCIR 10 and 11 Math Tasks.

ALGORITHMIC CONTRIBUTIONS

The authors present novel equation similarity metrics which, although based on latex, offer insights into how to perform matching of structured content in general.

EXPERIMENTAL CONTRIBUTIONS

There is comparison to a naive baseline and, given the newness of the domain, reimplementing of other NTCIR runs may have been difficult. However, I would have liked to see experiments on NTCIR 11 data.

DETAILS

This is the first Math Task paper I've read and I was impressed by the thoughtfulness of the design process. But I was a little underwhelmed by the experimental results. More below.

Algorithmically, the authors motivate their design as domain experts (i.e. individuals who know about math). In terms of full text retrieval, this is

akin to using linguistics to derive matching functions. I would have liked more of a motivation from the user perspective. Answering the question "why is a searcher querying by equation?" may better help answer "what is the appropriate similarity metric design?". That said, I think, since this is such a new domain, the "task-independent" motivation by the authors is important.

As I mentioned earlier, the experiments leave something to be desired. If this were a brand new task, I would be slightly more forgiving, especially given the theoretical nature of the conference. However, although the authors convince me that NTCIR 10 is inappropriate for their task, it's unclear why NTCIR 11 Math Task is inappropriate. I'm left wondering if there is some issue on that collection. Nevertheless, the authors do construct a new corpus and offer to distribute it publicly. I also would have liked to have seen more experimental conditions testing the design decisions made in Section 3. The current experiments only test all algorithmic properties at once. Granted, the size of the evaluation set may not support this but I would still like discussion of this, perhaps in the future. On that note, because of the small data size, the experiments require significance testing to be used as supporting evidence.

In summary, I think this submission contains several nice ideas on a new domain and passes the bar for presentation. The contributions are mainly theoretical/technical and, if empirical claims are made, the authors should improve the experimentation.

Meta-Review 4

=====

Contribution of the submission

The paper proposes a new algorithm for matching mathematical expressions based on their structural and symbolic similarity.

Evaluation of the contribution

Relevance	(10%): 4
Reproducibility	(10%): 4
Originality	(10%): 3
Quality	(10%): 2
Citations	(10%): 3
Presentation	(10%): 3
Impact	(10%): 3
Resources	(10%): 3
Overall Recommendation (20%):	2
Total points (out of 5)	: 2.9

Comments for the authors

This is the meta-review for the paper.

The consensus amongst the reviewers is that the paper is not ready for publication at ICTIR at this time.

The reviewers agreed that the paper is well written and addresses an

interesting and novel problem.

The main issue with the paper is the evaluation. As one of the reviewers rightfully points out, the ideas in the paper are fairly specific to the application of computing similarity between mathematical expressions. Thus, in this case, a robust and meaningful evaluation and error analysis is key.

The evaluation is limited in several ways:

1. The authors created a new test collection, but omitted important details about the test collection construction process. For example, what were the instructions given to the judges about assessing the level of structural and symbolic similarity between expressions? What was the level of agreement? Without knowing these details, it's difficult to determine the reliability of the results.
2. The evaluation was conducted on a small number of queries (i.e., 20) and significance tests were not performed to determine whether the results suggest an actual trend.
3. The evaluation used a weak baseline (a simplified version of the proposed approach) instead of re-implementing a more competitive baseline (possibly from NCTIR-10 or NCTIR-11).
4. There is no error analysis, so it's impossible to know which components of the proposed method actually made a difference. For example, what was it about queries 3, 7, and 9 that made the proposed method outperform the baseline?

The reviewers all agreed that the underlying problem is interesting and important and encourage to authors to consider their points in future revisions of the paper.
