



Yuchen Lin <yuchenlin1995@gmail.com>

Your ACL 2018 Submission (Number 638)

3 messages

Papers_acl2018 <papers_acl2018@softconf.com>
Reply-To: acl2018pc@ukp.informatik.tu-darmstadt.de
To: yuchenlin1995@gmail.com

Sat, Apr 21, 2018 at 4:02 AM

Dear Bill Y. Lin:

We are delighted to inform you that the following submission has been accepted to appear at ACL 2018:

Title: Mining Cross-Cultural Differences and Similarities in Social Media
Authors: Bill Y. Lin, Frank Xu, Kenny Zhu and Seung-won Hwang
Submission ID: 638

The Program Committee worked very hard to thoroughly review all the submitted papers. Please repay their efforts, by following their suggestions when you revise your paper.

Submit your final paper by *23:59 May 11 UTC-12* at the latest. Any delay will possibly cause an exclusion of your paper from the Proceedings. Upload the PDF file at the following site:

<https://www.softconf.com/acl2018/papers/>

You will be prompted to login to your START account. If you do not see your submission, you can access it with the following passcode:

638X-A2B6H2H7F2

Alternatively, you can click on the following URL, which will take you directly to a form to submit your final paper (after logging into your account):

<https://www.softconf.com/acl2018/papers/user/scmd.cgi?scmd=aLogin&passcode=638X-A2B6H2H7F2>

The reviews and comments are attached below. Again, try to follow their advice when you revise your paper.

Congratulations on your fine work. If you have any additional questions, please feel free to get in touch.

Best Regards,
Iryna Gurevych and Yusuke Miyao, Program Co-Chairs
ACL 2018

=====
ACL 2018 Reviews for Submission #638
=====

Title: Mining Cross-Cultural Differences and Similarities in Social Media
Authors: Bill Y. Lin, Frank Xu, Kenny Zhu and Seung-won Hwang
=====

REVIEWER #1

Reviewer's Scores

Appropriateness: Appropriate
Adhere to ACL 2018 Guidelines: Yes
Adhere to ACL Author Guidelines: Yes
Handling of Data / Resources: Yes
Handling of Human Participants: N/A

Summary and Contributions

The paper studies the problem of how to compute the similarity of terms between cultures and languages. E.g. the entity "Nagoya" has a positive, travel-related usage pattern in English but is used more critically in a political context in Chinese. Similarly, slang terms cannot be translated literally and so the task of finding the most similar slang term arises. The paper describes how such similarities between word or entities pairs in different languages/cultures can be computed using "pseudo-words" that make use of both translations and averages of uni-lingual word embeddings. This could be useful for translation applications and for studies in Computational Social Science.

Strengths

Overall, this is an interesting paper with a nice contribution to address an important problem. It's also nice to see two different types of evaluation: the cross-cultural entity similarity is arguably less interesting as the dimension ranges from "different" to "similar" without details on how or why the entities differ across languages. But the "find the most similar slang term in another language" is essentially a machine translation task which is more interesting.

The authors also make code and data available, which is another strength.

Weaknesses

The biggest issue is the write-up quality. Though generally understandable, concepts such as "social words" are not properly defined when first used and unusual word constructions such as "truth word sets" are used. Many sentences have (minor) English language issues which can make reading difficult in parts.

Some details on the training of the word embeddings are also missing. Please be sure to always give all parameters, sizes of data sets used and so on.

A key related paper is also not discussed:

Mining Cross-Cultural Differences of Named Entities: A Preliminary Study

http://kbcom.org/papers/KBCOM_2018_paper_12.pdf

This would have been another natural baseline to compare against.

Reviewer's Scores

NLP Tasks / Applications: Moderate contribution
Methods / Algorithms: Moderate contribution
Theoretical / Algorithmic Results: Marginal contribution
Empirical Results: Moderate contribution
Data / Resources: Moderate contribution
Software / Systems: Moderate contribution

Evaluation Methods / Metrics: Marginal contribution

Other Contributions: N/A

Originality (1-5): 3

Soundness/Correctness (1-5): 3

Substance (1-5): 3

Replicability (1-5): 5

Meaningful Comparison (1-5): 4

Readability (1-5): 2

Overall Score (1-6): 4

=====

REVIEWER #2

=====

Reviewer's Scores

Appropriateness: Appropriate

Adhere to ACL 2018 Guidelines: Yes

Adhere to ACL Author Guidelines: Yes

Handling of Data / Resources: Yes

Handling of Human Participants: N/A

Summary and Contributions

Summary: The paper presents a framework to detect cross-cultural differences and similarities in multilingual social media (Twitter and Weibo in this case). Authors propose a method to create a common vector space for two different languages (English and Chinese) from two different not-aligned corpora where to compute cultural similarities and differences. They also propose two novel tasks to evaluate their framework: mining cross-cultural differences of named entities in social media, and finding similar words for slang across languages.

Contribution 1: A (simple but effective) method to create a single vector space for two languages from two not-aligned corpora.

Contribution 2: The proposal of a task focused on mining cross-cultural differences in social media. I think it is really a very interesting task. It shows how NLP could be useful for cross-cultural studies.

Strengths

Strength argument 1: how this method is able to generate a single vector space from two corpora written in different languages I think it is a good contribution. I have some remarks (see below), but it is acceptable.

Strength argument 2: The cultural differences in the perception of named entities is an interesting topic. Authors apply the framework presented here to mining these differences in Twitter and Weibo. It is a complex task that might have some impact in NLP community.

Strength argument 3: How authors apply the framework to try to translate slang words is also remarkable.

Strength argument 4: The paper is clear and well written.

Weaknesses

Weakness argument 1: The method to generate a single vector space from two corpus written in two different languages is effective only for domain specific corpora; that is to say, it needs a fixed amount of words, a fixed vocabulary. In this case they are working only with social vocabulary. With this fixed vocabulary they can construct a single vector space for two languages. The problem is that it is not possible (or it is very hard) to extend the

method for general purpose systems.

Weakness argument 2: The method depends on a bilingual lexicon. In some way it is a classical dictionary-based method.

Weakness argument 3: The method proposed outperforms slightly other methods. I don't know if this improvement is statistically significant.

Reviewer's Scores

NLP Tasks / Applications: Strong contribution
 Methods / Algorithms: Moderate contribution
 Theoretical / Algorithmic Results: Moderate contribution
 Empirical Results: Moderate contribution
 Data / Resources: Marginal contribution
 Software / Systems: Moderate contribution
 Evaluation Methods / Metrics: N/A
 Other Contributions: N/A
 Originality (1-5): 4
 Soundness/Correctness (1-5): 5
 Substance (1-5): 4
 Replicability (1-5): 5
 Meaningful Comparison (1-5): 3
 Readability (1-5): 5
 Overall Score (1-6): 4

REVIEWER #3

Reviewer's Scores

Appropriateness: Appropriate
 Adhere to ACL 2018 Guidelines: Yes
 Adhere to ACL Author Guidelines: Yes
 Handling of Data / Resources: Yes
 Handling of Human Participants: Yes

Summary and Contributions

Summary:

The authors propose a framework for extracting cross-cultural differences and similarities from social media, specifically cross-cultural differences of named entities and similar terms for slang across languages (tested here for English-Chinese). The model outperforms a number of baselines on both tasks and is potentially useful in a range of social science applications.

Contribution 1:

- a model to compute cross-cultural differences of terms from different languages;

Contribution 2:

- release a prototype tool for building cross-cultural term similarities/differences and two datasets on the above tasks as well as a bilingual social lexicon;

Contribution 3:

Strengths

Strength argument 1:

- interesting and important problem with numerous applications of social media;

Strength argument 2:

Strength argument 3:

Strength argument 4:

Strength argument 5:

Weaknesses

Weakness argument 1:

- there are some aspects that are not clearly described in the paper (see Additional comments for details);
- what is meant by "social word vocabulary" (especially why this definition);
- lack of details on the low inter-annotator agreement;

Weakness argument 2:

Weakness argument 3:

Weakness argument 4:

Weakness argument 5:

Questions to Authors (Optional)

Question 1: How do the authors know that the social words thus translated still preserve their social meaning?

Question 2: What is the rationale for using semantic classes such as human perception and psychological processes as well as opinion and sentiment to define/identify social words?

Question 3: Why was the inter-annotator agreement (kappa) so low?

Reviewer's Scores

NLP Tasks / Applications: Moderate contribution

Methods / Algorithms: Moderate contribution

Theoretical / Algorithmic Results: Moderate contribution

Empirical Results: N/A

Data / Resources: Moderate contribution

Software / Systems: Moderate contribution

Evaluation Methods / Metrics: N/A

Other Contributions: N/A

Originality (1-5): 3

Soundness/Correctness (1-5): 3

Substance (1-5): 3

Replicability (1-5): 3

Meaningful Comparison (1-5): 4

Readability (1-5): 4

Overall Score (1-6): 4

Additional Comments (Optional)

More detailed comments/questions:

1) This example question is not clear - and it is important especially since it is used in the introduction:

"1. In 2012, did western people hold different views about a Japan's city, Nagoya, from Chinese people?"

2) "We then use BL to translate each social word in the ESV to a list of Chinese words and then filter out all the words that are not in the CSV. Afterwards, we have a set of Chinese social words for each English social word, which we denoted as a "translation set"."

How do the authors know that the social words thus translated still preserve their social meaning?

3) What is really meant by "social word vocabulary" ? This is a bit confusing since later in subsection 4.1 the authors say that they come from Empath and OpinionFinder for English, and TextMind for Chinese. However, these contain simple words that belong to some semantic categories - of interest being human perception and psychological processes as well as opinion and sentiment. What is the rationale for this approach?

4) Why was the inter-annotator agreement (kappa) so low? A short explanation is necessary in the body of the paper and not as a footnote.

--
ACL 2018 - <https://www.softconf.com/acl2018/papers>

Yuchen Lin (Bill) <i@yuchenlin.xyz>

Sat, Apr 21, 2018 at 4:04 AM

To: Kenny Zhu <kzhu@cs.sjtu.edu.cn>, Frank Xu <frankxu2004@gmail.com>, Seungwon Hwang <seungwon.hwang@gmail.com>

Congrats!!!

[Quoted text hidden]

Seung-won Hwang <seungwon.hwang@gmail.com>

Sat, Apr 21, 2018 at 4:29 AM

To: "Yuchen Lin (Bill)" <i@yuchenlin.xyz>, Kenny Zhu <kzhu@cs.sjtu.edu.cn>, Frank Xu <frankxu2004@gmail.com>

Congrats to all :)

보낸 사람: [Yuchen Lin \(Bill\)](#)

보낸 날짜: 2018-04-21 오전 5:04

받는 사람: [Kenny Zhu](#); [Frank Xu](#); [Seungwon Hwang](#)

제목: Fwd: Your ACL 2018 Submission (Number 638)

Congrats!!!

[Quoted text hidden]

Weakness argument 3: The method proposed outperforms slightly other methods I don't know if this improvement is statistically significant.

[Quoted text hidden]

[원본 메시지 일부만 포함되었습니다.]

