The 2018 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies

NAACL HLT 2018

Author Response

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

Instructions

The author response period has begun. The reviews for your submission are displayed on this page. If you want to respond to the points raised in the reviews, you may do so in the boxes provided below.

Please structure your response according to the template provided in the text area.

Please note: *you are not obligated to respond to the reviews*.

In addition to your author response, please go back to your submission form and fill out the Reviews for Research and Review Quality questions there. Reviewers will not see these responses.

Review #1

Appropriateness (1-3): 3

Comments on Appropriateness

This submission is marginally appropriate / inappropriate for its track because:

NLP Tasks / Applications Description

Description of the task / application: The paper describes an approach to build a 3rd vector space (SocVec) to represent the cross-cultural differences between a pair of languages in social media. Two tasks are described to prove the validity of the approach: (1) cross-cultural differences towards named entities, and (2) computation of cross-cultural similarities between words.

Strengths: * Very important task, and well motivated. * Clever approach that builds a 3rd vector space.

Weaknesses: * It is claimed that the two introduced tasks are new, however it is also said that (Garimella, 2016) tackled a similar problem differently. It would be good to clarify if they worked on the same tasks. * The experiments are conducted on popular terms, which leaves one wondering if it will perform well for less popular terms.

Impact of NLP Tasks / Applications (1-4): 4

Description of Method

Description of the method: A novel approach is presented for computing cross-cultural representations of words.

Strengths: * All data and code has been released.

Weaknesses:

Impact of Methods (1-4): 4

Description of Theoretical / Algorithmic Results

What are the results?

Strengths of the method or results?

Weaknesses of the method or results?

Impact of Theoretical / Algorithmic Results (1-4): 1

Empirical Results: Hypotheses

Hypotheses?

Novelty of the hypotheses?

Substance of the hypotheses?

Empirical Results: Method

What is the method for testing the hypothesis/es? Solid comparison with respect to a number of baseline approach.

Strengths of the method and results?

Weaknesses of the method and results?

Impact of Empirical Results (1-4): 4

Description of Data / Resources

Description of the data / resource: All data and code are available.

Strengths:

Weaknesses:

Impact of Data / Resources (1-4): 4

Description of Software / Systems

Description of the software / system:

Strengths:

Weaknesses:

Impact of Software / System (1-4): 4

Description of Evaluation Methods / Metrics

Description of evaluation method / metric:

Strengths:

Weaknesses:

Impact of Evaluation Method / Metric (1-4): 3

Description of Other Contributions

Description of contribution:

Strengths:

Weaknesses:

Impact of Other Contribution (1-4): 1

Contributions Summary

Contribution 1: Novel approach for computing cross-cultural differences and similarities of words.

Contribution 2: Experimentation on two different tasks, showing that the approach performs well.

Contribution 3: Availability of data and code.

Originality (1-5): 5
Soundness/Correctness (1-5): 5
Substance (1-5): 4
Replicability (1-5): 5
Handling of Data / Resources: Yes
Handling of Human Participants: Yes

Discussion of the Handling of Data, Resources and Participants

Handling of data / resources:

Handling of human participants:

Meaningful Comparison (1-5): 5
Related Work: ACL Guidelines: Yes

Discussion of Related Work

Strengths:

Weaknesses:

Missing references:

Comments on adherence to the ACL guidelines for authors:

NAACL Guidelines:Yes

Discussion of Readability, Style and Format

Comments on structure:

Comments on clarity and writing:

Comments on the use of the NAACL style and format guidelines:

ACL Guidelines: Yes
Reviewer Confidence (1-5): 3
Presentation Format: Oral

Review #2

Appropriateness (1-3): 3

Comments on Appropriateness

This submission is marginally appropriate / inappropriate for its track because:

NLP Tasks / Applications Description

Description of the task / application:Cross-cultural comparison of terms

Strengths: The task is explored for mining cross-cultural differences of named entities and finding similar terms for slang.

Weaknesses: The experiments are restricted to English and Chinese.

Impact of NLP Tasks / Applications (1-4): 3

Description of Method

Description of the method:A method to compare terms across languages in a unique vector space.

Strengths: The use of word embeddings in a proper arrangment. Several ways to compare similarity are tested.

Weaknesses: Dependant on external resources.

Impact of Methods (1-4): 3

Description of Theoretical / Algorithmic Results

What are the results?

Strengths of the method or results?

Weaknesses of the method or results?

Impact of Theoretical / Algorithmic Results (1-4): 1

Empirical Results: Hypotheses

Hypotheses? There is no hypothesis clearly stated. The empirical results are given to validate the tasks and methods proposed.

Novelty of the hypotheses?

Substance of the hypotheses?

Empirical Results: Method

What is the method for testing the hypothesis/es?

Strengths of the method and results?

Weaknesses of the method and results?

Impact of Empirical Results (1-4): 3

Description of Data / Resources

Description of the data / resource: The authors will make their data available to the community.

Strengths: The data could be useful for researchers interested in cross-cultural studies in Chinese and English.

Weaknesses: New external data will be necessary for experiments in other languages.

Impact of Data / Resources (1-4): 3

Description of Software / Systems

Description of the software / system: The authors will make their code available to the community.

Strengths: The code could be useful for researchers interested in cross-cultural studies in Chinese and English, or other languages.

Weaknesses: New external data will be necessary for experiments in other languages.There is no datails about the implementation.

Impact of Software / System (1-4): 3

Description of Evaluation Methods / Metrics

Description of evaluation method / metric:

Strengths:

Weaknesses:

Impact of Evaluation Method / Metric (1-4): 1

Description of Other Contributions

Description of contribution:

Strengths:

Weaknesses:

Impact of Other Contribution (1-4): 1

Contributions Summary

Contribution 1: Cross-cultural comparison of terms

Contribution 2: A method to compare terms across languages in a unique vector space.

Contribution 3: Experimental evaluation in mining cross-cultural differences of named entities and finding similar terms for slang.

Originality (1-5): 4
Soundness/Correctness (1-5): 4
Substance (1-5): 4
Replicability (1-5): 3
Handling of Data / Resources: N/A
Handling of Human Participants: N/A

Discussion of the Handling of Data, Resources and Participants

Handling of data / resources:

Handling of human participants:

Meaningful Comparison (1-5): 4
Related Work: ACL Guidelines: Yes

Discussion of Related Work

Strengths:

Weaknesses:

Missing references:

Comments on adherence to the ACL guidelines for authors:

Readability (1-5): 4
NAACL Guidelines: Yes

Discussion of Readability, Style and Format

Comments on structure:

Comments on clarity and writing: Table 5 is discussed after Table 6. Possibly "them" is not necessary at the end of first paragraph of Section 2.1 Minor typos: "Emapth" (Sec 4.3), "translatation" (Sec 5.3)

Comments on the use of the NAACL style and format guidelines:

ACL Guidelines: Yes
Reviewer Confidence (1-5): 3
Presentation Format: Oral

Review #3

Appropriateness (1-3): 3

Comments on Appropriateness

This submission is marginally appropriate / inappropriate for its track because:

NLP Tasks / Applications Description

Description of the task / application: The authors address a problem of finding similarity/dissimilarity between words in two different languages/cultures. The proposed system is also extended to explain slang words.

Strengths: Simple and intuitive method to map words from two or more languages into a common third space.

Weaknesses: The application requires external data sources that may not be available for certain languages.

Impact of NLP Tasks / Applications (1-4): 3

Description of Method

Description of the method: Projecting words from separate languages into a common space defined by a set of pivots. In this case pivots are defined by BSL.

Strengths: Simple and intuitive method to map words from two or more languages into a common third space.

Weaknesses: The application requires external data sources that may not be available for certain languages.

Impact of Methods (1-4): 3

Description of Theoretical / Algorithmic Results

What are the results? n/a

Strengths of the method or results? n/a

Weaknesses of the method or results? n/a

Impact of Theoretical / Algorithmic Results (1-4): 1

Empirical Results: Hypotheses

Hypotheses? The authors tested hypothesis that their approach outperforms a number of competing methods on word matching and slang matching tasks.

Novelty of the hypotheses? No novelty

Substance of the hypotheses? No issues with empirical evaluation of hypothesis.

Empirical Results: Method

What is the method for testing the hypothesis/es? The authors had two separate sets of experiments. First is inferring matching between word meanings in two languages (such as difference in meaning of "Nagoya" in Chinese and English. Second is matching slang between two languages.

Strengths of the method and results? The results are interesting and seem to show that the proposed method is outperforming competition.

Weaknesses of the method and results? The metrics are not well introduced and explained.

Impact of Empirical Results (1-4): 3

Description of Data / Resources

Description of the data / resource: For the most part authors are using existing external data sets. However, they also introduce a new data set that was manually labeled, where labelers had to indicate level of similarity between lists of words in English and Chinese. In addition, they hand-generated a small data set of matching slang words in Chinese and English.

Strengths: no issues that I see

Weaknesses: no issues that I see

Impact of Data / Resources (1-4): 3

Description of Software / Systems

Description of the software / system: n/a

Strengths: n/a

Weaknesses: n/a

Impact of Software / System (1-4): 1

Description of Evaluation Methods / Metrics

Description of evaluation method / metric: n/a

Strengths: n/a

Weaknesses: n/a

Impact of Evaluation Method / Metric (1-4): 1

Description of Other Contributions

Description of contribution: n/a

Strengths: n/a

Weaknesses: n/a

Impact of Other Contribution (1-4): 1

Contributions Summary

Contribution 1: New method based of pivot words (chosen to be words from social word vocabulary) for embedding words from two separate languages in a third common embedding space.

Contribution 2: New manually curated data sets that the authors use to evaluate methods on word and slang matching tasks.

Contribution 3:

Originality (1-5): 3
Soundness/Correctness (1-5): 4
Substance (1-5): 4
Replicability (1-5): 4
Handling of Data / Resources: Yes
Handling of Human Participants: N/A

Discussion of the Handling of Data, Resources and Participants

Handling of data / resources:

Handling of human participants:

Meaningful Comparison (1-5): 4
Related Work: ACL Guidelines: Yes

Discussion of Related Work

Strengths:

Weaknesses:

Missing references:

Comments on adherence to the ACL guidelines for authors:

Readability (1-5): 3
NAACL Guidelines: Yes

Discussion of Readability, Style and Format

Comments on structure: OK

Comments on clarity and writing: Some parts of the paper could have been better explained, especially those related to empirical evaluation and comparison to baselines.

Comments on the use of the NAACL style and format guidelines: OK

ACL Guidelines: Yes
Reviewer Confidence (1-5): 3
Presentation Format: Poster

Questions for Authors

The authors present an interesting work on matching words and concepts from two languages (they used Chinese and English). The idea is to use social lexicon to get matching pivot words in both languages, which will be used to generate word embeddings of words from both languages in a common space. First, vectors for matching social words are found, and then for any word in either language its original embedding (e.g., learned by glove or word2vec) is compared to embeddings of social words which then gives a new representation of an original words (e.g., if there are 10 pivots, new representation will be of length 10). The authors use the new embeddings to measure matching between the two languages, as well as to explain slang words.

The papers is relatively nicely written, and reader can understand and follow the ideas easily. However, some parts of the paper should be better explained, especially those related to empirical evaluation. Detailed comments are given below. - The authors should write the paper so that non-Chinese speaking readers can follow completely. E.g., comment in lines 623-626 is not clear for non-speakers. - "confidence" is mentioned in line 268 and 270, but not explained. - Notation could be improved. C_x is already used, when authors introduce C_i as well. Very similar notation meaning different things. - Typos: "Enlish", "similarities them", "entity linking a" - Authors use notation "L" before being introduced. - In footnote 10, it is mentioned that slang words are removed. How? - Explanations in evaluation section could be improved. E.g., it was not super clear to me how EM-JS works, how many top words were found, and similar. Similar holds for other methods. - Moreover, how were numbers in Table 2 found? Correlation between what was computed, how? - What was the mentioned ranking and classification problem? - Why not check slang result manually, there were only 200 of them? - Equation in line 686 is not well explained. What is A and what is B? This needs to be explained much better. - Line 729, "filtering out ordinary words", how was that done?

All in all, interesting work with interesting and simple idea that seems to work. The paper could be written better, some critical parts were skipped and left reader scratching their head.

Submit Response to Reviewers

Use the following boxes to enter your response to the reviews.

Response to Review #1:

Response to Review #2:

Response to Review #3:

General Response to Reviewers:

- Response to reviewers’ comments on methods (theoretical/algorithmic/empirical):

- Response to reviewers’ comments on task and application:

- Response to reviewers’ comments on data and resources:

- Response to reviewers’ comments on software / systems:

- Response to reviewers’ comments on evaluation methods:

Submit