Subject: Your SemEval 2015 Submission (Number 26)
Date: Fri, 6 Mar 2015 13:01:41 -0800
From: preslav.nakov@gmail.com
To: albarron@gmail.com

Dear Alberto Barrón-Cedeño:

On behalf of the SemEval 2015 Program Committee, we are delighted
to inform you that the following submission has been accepted
to appear at the conference:

QCRI: Answer Selection for Community Question
Answering - Experiments for Arabic and English

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.

When you are finished, you can upload your final manuscript
at the following site:

https://www.softconf.com/naacl2015/semeval/

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:

26X-G6B6D4C3H8

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/naacl2015/semeval/user/scmd.cgi?scmd=aLogin&passcode=26X-G6B6D4C3H8

The deadline to submit your camera-ready copy is

March 30, 2015

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,
Preslav Nakov, Torsten Zesch, Daniel Cer, and David Jurgens
SemEval 2015

==========
SemEval 2015 Reviews for Submission #26
==========

Title: QCRI: Answer Selection for Community Question Answering - Experiments for Arabic and English

Authors: Massimo Nicosia, Simone Filice, Alberto Barrón-Cedeño, Iman Saleh, Hamdy Mubarak, Wei Gao, Preslav Nakov, Giovanni Da San Martino, Alessandro Moschitti, Kareem Darwish, Lluís Màrquez, Shafiq Joty and Walid Magdy
==========
REVIEWER #1

==========

Reviewer's Scores

Appropriateness: Appropriate (most submissions)
Clarity: 5
Soundness: 5
Replicability: 4
Novelty: Accept

Comments
---

1. Please consider correcting some English phrasing issues:
a. Abstract: “approach was best performing” -> “approach was the best performing”
b. Section 2.2 last paragraph: “request for information” -> “request information”; “invite to discussions” -> “invite discussions”
c. Section 3.1, English A: “manual inspection… show” -> “manual inspection… shows”
d. Section 3.1, English B: “a classifier as that for English task A” -> “a classifier as with English task A”; “differently to the rest” -> “differently from the rest”
e. Section 4.1: “shows to be the most important” -> rephrase
f. Section 4.3: “analyses pointed that” -> “analyses showed that”

2. Section 2.6: Do you have a notion of how much of train/test overlap there is between users? Is it common for the same user that was seen in the train set to post comments in the test set?

==========
REVIEWER #2
==========

Reviewer's Scores

Appropriateness: Appropriate (most submissions)
Clarity: 5
Soundness: 5
Replicability: 4
Overall Recommendation: Accept

Comments

This is an excellent paper which presents the taken approaches clearly and in detail. The overall impression is that the authors have thought about the task from various perspectives while trying to solve/describe the tasks in means of features. This is shown by the different proposed heuristics, which also contribute the most to the final F1.

Interesting is the focus on the social factor - user behaviour, although that these features do not contribute much to the final F1. As the authors mention the number of dialog comments from repeating users is too low to make good statistics about the user behaviour but applying this approach to larger data

could be benefitial.

There are few comments which could be addressed:
- 2.1. Similarity Measures - the last sentence I would like better if the types of similarity are listed in the end before describing each one of them.

- 4. from the post-submission experiments it seems that the similarity features contributed the least to the results and logically thinking these are the first meaningful features which come to mind (at least to mine). I would be curious to see some explanation if the authors have found such.

PS: I find the steps shown absolutely enough to replicate the study, I've marked the replicability of the results with a 4 only because the authors point to concepts and algorithms instead of implementations and various implementations could slightly affect the final results.

REVIEWER #3

Reviewer's Scores

Appropriateness	Appropriate (most submissions)
Clarity	5
Soundness	5
Replicability	5
Overall Recommendation	Accept

Comments

With the exception of the usage of n-grams, I found the explanations very clear and reasonable. The technical approach is well-founded. Due to the thorough presentation, I believe the results would be easy to replicate, with the slight exception of the bug mentioned, of course.

In terms of structure, presenting the classification methods used first would give the feature descriptions more context, making them a little easier to read.

REVIEWER #4

Reviewer's Scores

Appropriateness	Appropriate (most submissions)
Clarity	5
Soundness	4
Replicability	5
Overall Recommendation	Accept

Comments

In this paper the authors present their system for Answer Selection in a Community QA. The approach combines many features, both textual and meta textual, and a set of heuristics. While using meta textual features and more or less ad-hoc heuristics may be rewarded with a high accuracy for a certain corpus, in the long run this strategy is not a winner. This being said, the system offers a well defined solution for an important practical problem and solves it accurately, gaining the first position for arabic by a wide margin (circa 8%).

Please consider the following suggestions:

1) Please make a more general Introduction in which to sketch the principal characteristics of the system. Make a separate section for describing the data

2) Section 2 is a little bit fragmented. Section 2.1 and Section 2.3 seem to be better off if presented under the same section

3) The 2..1.1 method works fine when the idf is computed on a neutral, general corpus. It needs a little bit of adaptation for experiments in-domain only . Please clarify

4) section 2.5will benefit from an example, as such it is not clear how pol(c) functions. I would replace "discard in the range (-1,1)" with a more direct affirmation. The ones that remain are just -1, 1 respectively , right?

5) Section 3.2 must be presented with section 2.1.1 - see suggestion 2) above

6) the reading of "(F1=17.44) surpassing ... primary submission (F1=54.66)" is not clear. Is there a typo, F1=57.44?

--
SemEval 2015 - https://www.softconf.com/naacl2015/semeval