Gmail by Google

Eduardo Gade Gusmão <eggduzao@gmail.com>

Your BSB 2012 paper 98289

1 mensagem

BSB 2012 <mcps@cin.ufpe.br> 26 de abril de 2012 23:32
Para: egg@cin.ufpe.br

Dear Mr. Eduardo Gusmão:

Congratulations - your paper "Issues on Sampling Negative Examples for Predicting Prokaryotic Promoters" has been accepted for oral presentation at the BSB 2012 and for publication in the conference proceedings published by Springer Verlag in the series LNCS/LNBI.

It is essential that you take into account the valuable remarks made by the reviewers, which are at the end of this message or can be found at https://submissoes.sbc.org.br/PaperShow.cgi?m=98289.

Remember that spelling and grammar mistakes have a negative impact on the value of a paper and of the publication as a whole, so we request that you do a thorough English review of the text, preferably using the help of an expert.

As the paper will be published by Springer Verlag, in the series LNCS/LNBI; we need you to become familiar with the style and format details of LNCS/ LNBI, and carefully review your paper to comply with them.

** IMPORTANT ** Below we provide you with some information needed to prepare the final version of your paper and submit it for inclusion in the proceedings. *** Deadline May 14, 2012 ***

To produce the BSB proceedings we will need to receive the sources files of your paper. Latex users must send: a single TEX file with embedded bibliographic references, additional style files, special math fonts, and figure files. Word users must send a single DOC file. Please, make sure that there are no academic titles in the names, and that they are in the order in which they should appear in the paper. We also require a PDF file of the paper.

The paper will be readjusted to **LNCS style**, so make sure that **no modifications ** were made to fonts, margins and other style parameters and the **author guide has been followed** (http://www.springer.com/computer/lncs?SGWID=0-164-6-793341-0). The **limitation on the number of pages remains the same**. If your paper does not conform to the style, it will be sent back to you. As we have a tight schedule, if the issues are not solved on time your paper will not be included in the proceedings.

The files regarding the final version of your paper should be sent by e-mail to gpt@ic.unicamp.br (Guilherme Telles) ***no later than May 14, 2012 ***, together with the name and e-mail contact of a corresponding author. Please, make sure that this contact person is aware of the publication format details and will be available during the following eight weeks to fix presentation problems that Springer may raise.

Please, if you have any question, do not hesitate in contacting me.

Regards,
BSB2012 Conference Chairs

===== Review =====

*** Appropriateness for BSB (Is the paper relevant to the conference topics; meets length and language restrictions.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Originality (New results; novel approach or combination of approaches; has not been published before; brings innovations to research field(s); points out differences from related research.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Technical Quality (Soundness; depth; describes scope and limitations of the approach.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=average (3)

*** Readability (Clarity; connectivity between sections; appropriate details, title, abstract, figures, tables, grammar, structure, references.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Overall Rating (Overall rating for the paper.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Recommended Action (Accept or reject the paper.): 1: (strong reject) Serious problems. I'd argue to reject this paper. 2: (weak reject) Weak paper. But I wouldn't argue strongly *against* it. 3: (weak accept) OK paper. But I wouldn't argue strongly *for* it. 4: (strong accept) Good paper. I'd argue to accept it.

Evaluation=(strong accept) Good paper. I'd argue to accept it. (4)

*** Confidence in your judgment (confidence in your judgement; expertise in the field): 1: Low (I'm beginner in the filed) 2: Medium (I know the field) 3: High (I'm an expert in the field)

Evaluation=Medium (I know the field) (2)

*** Comments for the authors (Please add comments for the authors supporting your evaluation above.): In this paper authors compare distinct approaches to compose negative datasets for classification problems, which is a common issue when dealing with tasks of this

nature. This study concerns the specific application of identification of promoter regions.

In general, the paper is very well written and organized, and presents a good depth of discussion.

However, some points were left behind, at least for a non-expert reader. For instance, what are the categoric attributes when the primary sequence datasets are used for training and testing? This is not clear in the text.

Also, authors mention that a paired t-test was applied in results analysis, but in the discussion of both case studies is not clear whether this test was indeed performed. Variations are characterized as "large", "no great", "slightly lower", among others. But are the differences among computed metrics statistically significant?

At last, when comparing graphs on Figures 1 and 2, authors do not discuss the variation observed in the classifiers' performance regarding the CTRL dataset: a large increase is observed in the specificity (from ~70% to ~100%) when attributes are extracted with the vw Z-curve. It seems to me this is more than "no great variation", as described in the text. Conversely, the sensitivity decreases for this case (from ~90% to ~70%). How authors justify this behavior? This point should be adressed in paper discussion.

I truly support authors to review the above points in order to make them clearer for readers as well as to enrich the paper.

===== Review =====

*** Appropriateness for BSB (Is the paper relevant to the conference topics; meets length and language restrictions.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=average (3)

*** Originality (New results; novel approach or combination of approaches; has not been published before; brings innovations to research field(s); points out differences from related research.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=weak (2)

*** Technical Quality (Soundness; depth; describes scope and limitations of the approach.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=weak (2)

*** Readability (Clarity; connectivity between sections; appropriate details, title, abstract, figures, tables, grammar, structure, references.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=weak (2)

*** Overall Rating (Overall rating for the paper.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=weak (2)

*** Recommended Action (Accept or reject the paper.): 1: (strong reject) Serious problems. I'd argue to reject this paper. 2: (weak reject) Weak paper. But I wouldn't argue strongly *against* it. 3: (weak accept) OK paper. But I wouldn't argue strongly *for* it. 4: (strong accept) Good paper. I'd argue to accept it.

Evaluation=(weak reject) Weak paper. But I wouldn't argue strongly *against* it. (2)

*** Confidence in your judgment (confidence in your judgement; expertise in the field): 1: Low (I'm beginner in the filed) 2: Medium (I know the field) 3: High (I'm an expert in the field)

Evaluation=High (I'm an expert in the field) (3)

*** Comments for the authors (Please add comments for the authors supporting your evaluation above.): It is not clear what is the originality and how valid are the conclusions from this work. The two main methods used to generate the negative examples are from [8]. The other are either random or small variations. What are the novelties regarding the work from 8? Where these results biologically validated? The writing needs to be revised, since there are several errors. More details regarding the experiments are needed. Which decision tre algorithm was used? What were the parameter values used for the supervised learning techniques? Why these particular four techniques? How statistically significant are the results?

===== Review =====

*** Appropriateness for BSB (Is the paper relevant to the conference topics; meets length and language restrictions.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Originality (New results; novel approach or combination of approaches; has not been published before; brings innovations to research field(s); points out differences from related research.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Technical Quality (Soundness; depth; describes scope and limitations of the approach.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Readability (Clarity; connectivity between sections; appropriate details, title, abstract, figures, tables, grammar, structure, references.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=average (3)

*** Overall Rating (Overall rating for the paper.): 1:insufficient 2:weak 3:average 4:good 5:excellent

Evaluation=good (4)

*** Recommended Action (Accept or reject the paper.): 1: (strong reject)

Serious problems. I'd argue to reject this paper. 2: (weak reject) Weak paper. But I wouldn't argue strongly *against* it. 3: (weak accept) OK paper. But I wouldn't argue strongly *for* it. 4: (strong accept) Good paper. I'd argue to accept it.

Evaluation=(strong accept) Good paper. I'd argue to accept it. (4)

*** Confidence in your judgment (confidence in your judgement; expertise in the field): 1: Low (I'm beginner in the filed) 2: Medium (I know the field) 3: High (I'm an expert in the field)

Evaluation=Medium (I know the field) (2)

*** Comments for the authors (Please add comments for the authors supporting your evaluation above.): The paper performs a study on the effect of choosing different sets of negative data in promoter prediction datasets. The topic is interesting and the materials and methods used are sound and adequate.

Nonetheless, I expected more discussions from the results achieved. For instance, were the results for the CTRL negative data as expected? I would expect worse results by using such negative data, which would evidence better the problem of choosing proper negative data.

Some remarks:

- Shouldn't dataset CTRL have low similarity to POS data? - Please replace "two cases case" by "two cases" only in page 4. - Figures 1 and 2 are difficult to follow. They have too much information. Maybe the complement of the performance rates shown could be represented more clearly. - Please complete reference 4.