Eduardo Gade Gusmão <eggduzao@gmail.com>

IJCNN 2014 Paper #14305 Decision Notification

4 mensagens

Cesare Alippi <cesare.alippi@polimi.it> 4 de março de 2014 12:42
Responder a: Cesare Alippi <cesare.alippi@polimi.it>
Para: eduardo.gusmao@rwth-aachen.de, marcilio.desouto@univ-orleans.fr

Dear Author(s),

Congratulations! On behalf of the IJCNN 2014 Program Chairs, we are pleased to inform you that your paper:

Paper ID: 14305
Author(s): Eduardo Gusmao and Marcilio de Souto
Title: Issues on Sampling Negative Examples for Predicting Prokaryotic Promoters

has been accepted for presentation at the 2014 International Joint Conference on Neural Networks and for publication in the conference proceedings published by IEEE. This email provides you with all the information required to complete your paper and submit it for inclusion in the proceedings. A notification of the presentation format (oral or poster) and timing of that presentation will be sent in a subsequent email.

Here are the steps:

1. Please address the attached REVIEWERS' COMMENTS which are intended to improve the final manuscript. Final acceptance is conditional on appropriate response to the requirements and comments.

2. Please prepare your manuscript for final camera ready submission by following the formats described on the conference web site and using the IEEE templates:

http://www.ieee-wcci2014.org/Paper-Submission.htm

The use of the template is compulsory: any discrepancy from it will be automatically detected by the system and you will need to provide a compliant submission. Once you are ready to submit it, please go to:

http://ieee-cis.org/conferences/ijcnn2014/upload.php?PaperID=14305

to submit your final camera-ready paper. On this page you will need to use the following password:

t7559my6

Please do adhere the strict deadline for final manuscript submission April 15, 2014. Any papers submitted after this date will not be included in the proceedings. The paper must be re-submitted even if the reviewers indicated that no changes are required.

3. In order for your paper to be published in the conference proceedings, a

signed IEEE Copyright Form must be submitted for each paper. IJCNN 2014 has registered to use the IEEE Electronic Copyright (eCF) service. The confirmation page shown after submitting your final paper contains a button linking directly to a secure IEEE eCF site which allows electronic completion of the copyright assignment process. In case it fails, please have the completed IEEE Copyright Form, found at http://www.ieee.org/web/publications/rights/copyrightmain.html, emailed it to Publication Co-Chair, Zhigang Zeng <zgzeng@hust.edu.cn>.

IMPORTANT: No paper can be published in the proceedings without being accompanied by a Completed IEEE Copyright Transfer Form. You must complete and submit this form to have your paper included in the conference proceedings.

4. Register for the conference at http://www.ieee-wcci2014.org by clicking on the conference registration link on the main page. Please read carefully all details in there.

IMPORTANT: Each paper MUST have a corresponding registered author to be included in the proceedings. Papers that do not have an associated registered author will NOT be included in the proceedings. The strict deadline for author registration is April 15, 2014 so be sure to register by that time otherwise your paper will not be included in the proceedings. No extensions will be given; please ensure that you complete your registration on time.

For questions regarding the conference program please contact Cesare Alippi <cesare.alippi@polimi.it>. For questions regarding Special Sessions contact Stefano Squartini <s.squartini@univpm.it>. For any other issue contact the conference secretariat at wcci2014@gmail.com.

Sincerely,
Cesare Alippi <cesare.alippi@polimi.it>
IJCNN 2014 Program Chair

REVIEWERS' COMMENTS

REVIEW NO. 1

Originality: Reject
Significance of topic: Weak Reject
Technical quality: Strong Reject
Relevance to IJCNN 2014: Reject
Presentation: Reject
Overall rating: Reject

Reviewer's expertise on the topic: High
Suggested form of presentation: Poster
Best Paper Award nomination: No

Comments to the authors:
This work proposes the use of supervised learning to identify the promoter

regions of Prokaryotes. The "novel" contribution apparently is from the use of

different "negative" datasets. As such, there is little relevance to IJCNN other

than being a simple application paper that uses machine learning techniques.

Some technical comments:

- Why are the authors focusing on "promoter region" rather than the "regulatory region" in general?
- The rationale for this work is extremely weak.
- It is unclear to this reviewer why this is treated as a binary problem as promoter regions are usually not treated this way. (See below.)
- E.Coli has been well studied and the regulatory regions are well known. As such, it is not a good choice for the proposed methodology. That is, it is clear what is the promoter region and the rest would then simply be non-promoter region. There is a huge imbalance between promoter and non-promoter regions with the former being a very small percentage of the genome. It is obvious that only COD1 would have some relevance since it is in the region of a promoter.
- The authors need to at least present a second case study to illustrate their point.
- The proposed methodology seems very ad-hoc from combination of previous methods. In particular, this work mostly follows that of Ref 18.
- Use of the term "accuracy metrics" is misleading and should be changed to "performance metrics"; the "Correct Rate" is actually the "Accuracy"; also, Area Under the Curve of the ROC curve should be used as well.
- The authors should NOT use the default parameters only and should have done a sensitivity analysis instead.

Other comments:

- This paper is somewhat difficult to understandable as there are numerous typos and/or grammatical inconsistencies. More careful proofreading, preferably by an

English specialist, is necessary should this work be accepted.

e.g. in the abstract, "univocal" should be "unequivocal"; on p.1 Introduction, "each of which has their ..." should be "each of which has its ..."; and many, many more.

- consecutive citations should simply be shown as a range; e.g. on p.1, [12], [13], [14], [15] should be shown as [12-15];

- The acrynom TSS (transcription start site) should be clearly defined before use.

REVIEW NO. 2

Originality: Neutral
Significance of topic: Weak Accept
Technical quality: Accept
Relevance to IJCNN 2014: Weak Accept
Presentation: Weak Accept
Overall rating: Accept

Reviewer's expertise on the topic: High
Suggested form of presentation: Poster
Best Paper Award nomination: No

Comments to the authors:
This paper compares various methods for constructing negative examples to predict prokaryotic promoters, and provides a well-documented guideline regarding this issue.

Comments:

1 Introduction: In the first paragraph, the authors should mention the recent advances of high-throughput experimental methods (e.g. RNA-seq) for annotating transcription start sites.

2. Materials and Methods: In the third paragraph at page 2, "all the learning methods used in our study were obtained from the Matlab", if the authors use the MATLAB statistics toolbox, please specify it and provide the version used.

3. Control negative examples: Since there are a correlation among nucleotides within the genome, the most widely used approach to make a background sequence is
to use a higher-order markov chain. I recommend to use the higher-order Markov chain trained from the whole genome sequence to make control negative examples rather than the 0-order Markov chain used in this study.

4. B. Variable-window Z-Curve feature extraction: The second paragraph of this section, which is excerpted from [18] with a slight modification, is not well written.
The notations and equation (1) (why is the uniform definition of the vw Z-curve variables is important?) are difficult to follow.
I recommend to remove this second paragraph or revise it extensively.

5. Experiments and results: If MEME E-value is significant, it can not be larger than 1.
At page 5, (MEME E-value=1.3X10^19) --> 1.3X10^-19, and in Fig. 2. 1.3X10^19 --> 1.3X10^-19, 9.6X10^51 --> 9.6X10^-51

6. There are many typos and grammatical errors in this paper. It should be carefully proof-read.

REVIEW NO. 3

Originality: Strong Accept
Significance of topic: Strong Accept
Technical quality: Accept
Relevance to IJCNN 2014: Accept
Presentation: Strong Accept
Overall rating: Strong Accept

Reviewer's expertise on the topic: Medium
Suggested form of presentation: Oral
Best Paper Award nomination: No

Comments to the authors:

The paper discussed the important issue of negative sampling. The topic is very important and the paper is well written so I recommend that it get accepted.

However there are some comments that need to be addressed.

1- There is no justification for the classifiers choice. Why did not you use MLP with binary representation of the nucleotide as well.

2- Using NB is always recommended as a statistical classifier however, I believe it should always be combined with a full Bayesian Classifier. If NB outperformed the full BC, that is an indication of either overfitting, or the prior destitution assumption is wrong.

3- The representation in equating 1 looks need, but there are simpler representations in the literature to use.

4- I do not see the real benefit of scenario 1 which has cross validation without a test set. That is strange, why did not have a test set with that as well?

5- A very useful addition to your paper would be why each algorithm performed the way it did.

6- The study looks so important, so it would be great if you give more space to the discussion and recommendation what a negative sample should be defined as when it comes to promoter regions.

7- Finally what is your plan moving to Eukaryotes? Because it is fairly easy when it comes to promoters in prokaryotes as most of the DNA is covered with coding areas.

Eduardo Gade Gusmão <eduardo.gusmao@rwth-aachen.de> 4 de março de 2014 14:10
Para: Marcilio Pereira De Souto <marcilio.desouto@univ-orleans.fr>

Marcilio,

tudo bem por ai?
marca uma data pra gente conversar. Pra mim pode ser qualquer dia / qualquer hora, porem de preferencia nao hoje (04.03.2013).

abs!

2014-03-04 12:42 GMT+01:00 Cesare Alippi <cesare.alippi@polimi.it>:
[Texto das mensagens anteriores oculto]

Marcilio de Souto <marcilio.desouto@univ-orleans.fr> 4 de março de 2014 15:43
Para: Eduardo Gade Gusmão <eduardo.gusmao@rwth-aachen.de>

oi eduardo,

que tal amanha as 16h?

abcs
--
Marcilio DE SOUTO
Professeur - Université d'Orléans
LIFO Bat. 3IA
Rue Léonard de Vinci, B.P. 6759
F-45067 ORLEANS Cedex 2
Tel :+33 (0)2 38 41 72 92
http://www.univ-orleans.fr/lifo/Members/Marcilio.DeSouto/
[Texto das mensagens anteriores oculto]

Eduardo Gade Gusmão <eduardo.gusmao@rwth-aachen.de> 4 de março de 2014 15:55
Para: Marcilio de Souto <marcilio.desouto@univ-orleans.fr>

ok Marcilio. Ate amanha. abs.

2014-03-04 15:43 GMT+01:00 Marcilio de Souto <marcilio.desouto@univ-orleans.fr>:
[Texto das mensagens anteriores oculto]