Review Form

Submission #14: IITP: Multiobjective Differential Evolution based Twitter Named Entity Recognition
paper types: System Description Paper for NER Shared Task

Reviewer: Zhiqiang Toh
Secondary Reviewer:

Summary Ranking

Please evaluate the submission according to the criteria below.

Evaluation Category	Enter Your Score
APPROPRIATENESS (1-5) Does the paper fit in WNUT 2015? Please answer this question in light of the desire to accept a diverse set of papers on noisy user-generated text. 5: Certainly. 4: Probably. 3: Unsure. 2: Probably not. 1: Certainly not.	5
CLARITY (1-5) For the reasonably well-prepared reader, is it clear what was done and why? Is the paper well-written and well-structured? 5 = Very clear. 4 = Understandable by most readers. 3 = Mostly understandable to me with some effort. 2 = Important questions were hard to resolve even with effort. 1 = Much of the paper is confusing.	4
ORIGINALITY (1-5) How original is the approach? Does this paper break new ground in topic, methodology, or content? How exciting and innovative is the research it describes? Note that a paper could score high for originality even if the results do not show a convincing benefit. 5 = Surprising: Significant new problem, technique, methodology, or insight -- no prior research has attempted something similar. 4 = Creative: An intriguing problem, technique, or approach that is substantially different from previous research. 3 = Respectable: A nice research contribution that represents a notable extension of prior approaches or methodologies. 2 = Pedestrian: Obvious, or a minor improvement on familiar techniques. 1 = Significant portions have actually been done before or done better.	4
SOUNDNESS / CORRECTNESS (1-5) First, is the technical approach sound and well-chosen? Second, can one trust the claims of the paper -- are they supported by proper experiments and are the results of the experiments correctly interpreted? 5 = The approach is very apt, and the claims are convincingly supported. 4 = Generally solid work, although there are some aspects of the approach or evaluation I am not sure about. 3 = Fairly reasonable work. The approach is not bad, and at least the main claims are probably correct, but I am not entirely ready to accept them (based on the material in the paper). 2 = Troublesome. There are some ideas worth salvaging here, but the work should really have been done or evaluated differently. 1 = Fatally flawed.	4
REPLICABILITY (1-5) Will members of the ACL community be able to reproduce or verify the results in this paper? Members of the ACL community: 5 = could easily reproduce the results and verify the correctness of the results. 4 = could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method. 3 = could reproduce the results with some difficulty. The settings of parameters are underspecified or subjectively determined; the training/evaluation data are not widely available. 2 = would be hard pressed to reproduce the results. The contribution depends on data that are simply not available outside the author's institution or consortium; not enough details are provided. 1 = could not reproduce the results here no matter how hard they tried.	3
MEANINGFUL COMPARISON (1-5) Do the authors make clear where the problems and methods sit with respect to existing literature? Are the references adequate? For empirical papers, are the experimental results meaningfully compared with the best prior approaches? 5 = Precise and complete comparison with related work. Good job given the space constraints. 4 = Mostly solid bibliography and comparison, but there are some references missing. 3 = Bibliography and comparison are somewhat helpful, but it could be hard for a reader to determine exactly how this work relates to previous work. 2 = Only partial awareness and understanding of related work, or a flawed empirical comparison. 1 = Little awareness of related work, or lacks necessary empirical comparison.	3
IMPACT OF RESOURCES (1-5) 5 = Enabling: The newly released resources should affect other people's choice of research or development projects to undertake. 4 = Useful: I would recommend the new resources to other researchers or developers for their ongoing work. 3 = Potentially useful: Someone might find the new resources useful for their work. 2 = Documentary: The new resources are useful to study or replicate the reported research, although for other	1

purposes they may have limited interest or limited usability. (this is a positive rating). 1 = No usable resources released. (most submissions).	
IMPACT OF IDEAS OR RESULTS (1-5) How significant is the work described? If the ideas are novel, will they also be useful or inspirational? Does the paper bring any new insights into the nature of the problem? 5 = Will affect the field by altering other people's choice of research topics or basic approach. 4 = Some of the ideas or results will substantially help other people's ongoing research. 3 = Interesting but not too influential. The work will be cited, but mainly for comparison or as a source of minor contributions. 2 = Marginally interesting. May or may not be cited. 1 = Will have no impact on the field.	4 ▼
RECOMMENDATION (1-5) In deciding on your ultimate recommendation, please think over all your scores above. But remember that no paper is perfect, and remember that this is a workshop, therefore works in progress or partially evaluated could be interesting for discussion and for fostering further advances in the field. Remember also that the author has a couple of weeks to address reviewer comments before the camera-ready deadline. Should the paper be accepted or rejected? 5 = This paper changed my thinking on this topic and I'd fight to get it accepted; 4 = I learned a lot from this paper and would like to see it accepted. 3 = Borderline: I'm ambivalent about this one. 2 = Leaning against: I'd rather not see it in the workshop. 1 = Poor: I'd fight to have it rejected.	4 ▼
REVIEWER CONFIDENCE (1-5) 5 = Positive that my evaluation is correct. I read the paper very carefully and am familiar with related work. 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings. 3 = Pretty sure, but there's a chance I missed something. Although I have a good feel for this area in general, I did not carefully check the paper's details, e.g., the math, experimental design, or novelty. 2 = Willing to defend my evaluation, but it is fairly likely that I missed some details, didn't understand some central points, or can't be sure about the novelty of the work. 1 = Not my area, or paper is very hard to understand. My evaluation is just an educated guess.	3 ▼

Detailed Comments

Please supply detailed comments to back up your rankings. These comments will be forwarded to the authors of the paper. The comments will help the committee decide the outcome of the paper, and will help justify this decision for the authors. Moreover, if the paper is accepted, the comments should guide the authors in making revisions for a final manuscript. Hence, the more detailed you make your comments, the more useful your review will be - both for the committee and for the authors.

```
This paper describes a Twitter NER system that uses a combination of features.
The optimal feature set is determined using multi-objective differential
evolution, using precision and recall as the two objective functions.

The paper is well-written and structured. It would be nice if there are more
details of the algorithm, e.g. high-level pseudocode.

Figure 1 illustrates the proposed methodology with two steps. It is unclear
what Step 1 is doing. It is stated Step 1 generates many models based
on the various feature combinations, and subsequently selects the best model
based on the development data. This seems like some form of feature selection,
which is also done in Step 2. Is Step 1 performing some preliminary (or simple)
feature selection, while Step 2 performs more fine-grained (or complex) feature
selection? Some additional description of Step 1 will benefit the reader.

Some minor issues:

- Provide additional description or reference to the crowding distance sorting
algorithm.
- The acronym "MOO" was used without referring to the original (long) form.
```

Confidential Comments for Committee

You may wish to withhold some comments from the authors, and include them solely for the committee's internal use. For example, you may want to express a very strong (negative) opinion on the paper, which might offend the authors in some way. Or, perhaps you wish to write something which would expose your identity to the authors. If you wish to share comments of this nature with the committee, this is the place to put them.

Submit

START Conference Manager (V2.61.0 - Rev. 5183)