From: icdm-inform@wi-lab.com
Subject: ICDM 2016 acceptance decision: regular paper
Date: September 9, 2016 at 1:25 PM
To: sscott@cse.unl.edu, ymo@cse.unl.edu, ddowney@eecs.northwestern.edu

I

Dear authors,

Your submission

DM1030, titled Learning Hierarchically Decomposable Concepts with Active Over-Labeling

has been accepted for presentation at the IEEE ICDM 2016 conference as a *regular* paper. Congratulations!

This year, only 19.6% of submissions were accepted to the conference (8.5% regular papers and 11.1% short papers). Regular papers will have up to 10 pages and short papers up to 6 pages in the proceedings based on the IEEE double column format.

This email contains a number of important instructions. Please read it carefully since it covers many details. The reviews are attached at the end of this message.

- *Camera-ready papers are due by Sep. 30, 2016.*

Please take the review reports into account when preparing your final version. Remember that reproducibility is a central scientific principle. Therefore, please make sure that the code and the data are available (where needed, add them to a suitable repository, such as GitHub, Google Sites, or your homepage) and provide pointers to the appropriate URLs as a footnote in the final version of your paper. This will ensure that the community can benefit from your work, and may also attract additional citations to it.
A follow-up email containing detailed instructions for camera ready preparation will be sent next week.

- *The author registration deadline is Oct. 3, 2016.*

ICDM is a forum for presenting and discussing current research in data mining. Therefore, every accepted paper must have at least one registered author by Oct. 3, in order for the paper to be included in the conference proceedings and the program.

Information on registration is available at:
http://icdm2016.eurecat.org/registration/

See the above URL on the policy for authors of multiple accepted papers.

The conference dates are Dec. 13-15, 2016, and the venue is Barcelona, Catalonia. Please follow the conference website at http://icdm2016.eurecat.org/ for more information on the program, including workshops and participation. Visa letter information will be announced on the conference website soon.

Congratulations again! We look forward to seeing you at the IEEE ICDM 2016 in December.

Francesco Bonchi and Josep Domingo-Ferrer
IEEE ICDM 2016 PC Co-chairs

======

--======== Review Reports ========--

The review report from reviewer #1:

*1: Is the paper relevant to ICDM?
 [_] No
 [X] Yes

*2: How innovative is the paper?
 [_] 6 (Very innovative)
 [X] 3 (Innovative)
 [_] -2 (Marginally)
 [_] -4 (Not very much)
 [_] -6 (Not at all)

*3: How would you rate the technical quality of the paper?
 [_] 6 (Very high)

[_] 6 (Very high)
[X] 3 (High)
[_] -2 (Marginal)
[_] -4 (Low)
[_] -6 (Very low)

*4: How is the presentation?
[_] 6 (Excellent)
[X] 3 (Good)
[_] -2 (Marginal)
[_] -4 (Below average)
[_] -6 (Poor)

*5: Is the paper of interest to ICDM users and practitioners?
[X] 3 (Yes)
[_] 2 (May be)
[_] 1 (No)
[_] 0 (Not applicable)

*6: What is your confidence in your review of this paper?
[X] 2 (High)
[_] 1 (Medium)
[_] 0 (Low)

*7: Overall recommendation
[_] 6: must accept (in top 25% of ICDM accepted papers)
[X] 3: should accept (in top 80% of ICDM accepted papers)
[_] -2: marginal (in bottom 20% of ICDM accepted papers)
[_] -4: should reject (below acceptance bar)
[_] -6: must reject (unacceptable: too weak, incomplete, or wrong)

*8: Summary of the paper's main contribution and impact
The authors present a new approach for learning hierarchically decomposable concepts, which learns a high-level classifier by separately learning multiple finer-grained classifiers, and then combining the results. They also address the additional cost issue by suggesting a "cost-sensitive active learner". Early experimental results are promising.

*9: Justification of your recommendation
This is a solid paper, well motivated and presented. It contains both

This is a solid paper, well-motivated and presented. It contains both technical contributions and evidence of practical significance.

*10: Three strong points of this paper (please number each point)
 - Well motivated and presented
- Strong technical contributions
- Promising early experimental results

*11: Three weak points of this paper (please number each point)
 - The evaluation part could be more comprehensive

*12: Is this submission among the best 10% of submissions that you reviewed for ICDM'16?
 [X] No
 [_] Yes

*13: Would you be able to replicate the results based on the information given in the paper?
 [X] No
 [_] Yes

*14: Are the data and implementations publicly available for possible replication?
 [X] No
 [_] Yes

*15: If the paper is accepted, which format would you suggest?
 [X] Regular Paper
 [_] Short Paper

*16: Detailed comments for the authors
 The authors present a new approach for learning hierarchically decomposable concepts, which learns a high-level classifier by separately learning multiple finer-grained classifiers, and then combining the results. They also address the additional cost issue by suggesting a "cost-sensitive active learner". Early experimental results are promising.

This is a solid paper, well-motivated and presented. It contains both technical contributions and evidence of practical significance. The evaluation part, however, could be more comprehensive.

===
The review report from reviewer #2:

*1: Is the paper relevant to ICDM?
 [_] No
 [X] Yes

*2: How innovative is the paper?
 [X] 6 (Very innovative)
 [_] 3 (Innovative)
 [_] -2 (Marginally)
 [_] -4 (Not very much)
 [_] -6 (Not at all)

*3: How would you rate the technical quality of the paper?
 [_] 6 (Very high)
 [X] 3 (High)
 [_] -2 (Marginal)
 [_] -4 (Low)
 [_] -6 (Very low)

*4: How is the presentation?
 [_] 6 (Excellent)
 [X] 3 (Good)
 [_] -2 (Marginal)
 [_] -4 (Below average)
 [_] -6 (Poor)

*5: Is the paper of interest to ICDM users and practitioners?
 [X] 3 (Yes)
 [_] 2 (May be)
 [_] 1 (No)
 [_] 0 (Not applicable)

*6: What is your confidence in your review of this paper?
 [X] 2 (High)
 [_] 1 (Medium)

[_] 0 (Low)

*7: Overall recommendation
[_] 6: must accept (in top 25% of ICDM accepted papers)
[X] 3: should accept (in top 80% of ICDM accepted papers)
[_] -2: marginal (in bottom 20% of ICDM accepted papers)
[_] -4: should reject (below acceptance bar)
[_] -6: must reject (unacceptable: too weak, incomplete, or wrong)

*8: Summary of the paper's main contribution and impact
The authors describe a new strategy, called "active over-labeling," for learning hierarchically decomposable concepts. Instead of learning broader or coarse-grained concepts directly, the authors propose actively soliciting labels at a subconcept or finer grained level. They also propose a variant of an existing multi-armed bandit method that can dynamically determine the proportions of labels that are sought at various levels of a hierarchy. They show that active over-labeling is more effective than state-of-the-art active and passive learning methods in some situations.

*9: Justification of your recommendation
The authors have developed a method that appears to be unique in its suggestion that labels be sought at a finer level of granularity. The methods are presented clearly and convincingly. They evaluate the effectiveness of their method with various fine label acquisition costs and show that their BANDIT method can effectively pick an effective mix of labels at various levels to solicit. The active over-labeling method appears to be a unique method that outperforms state-of-the-art active and passive learning methods for hierarchical concepts. The graphs could be improved, more future work could be discussed, and some descriptions of the results in the paper do not exactly correspond with the results shown in the graphs and tables, but these issues can be fixed in editing.

There are some minor weaknesses in the experimental evaluation.

But, this paper should be accepted.

*10: Three strong points of this paper (please number each point)
1. Appears to be a unique and effective new technique for learning hierarchically decomposable concepts.

2. Very clear explanation of the topic and the technique.
3. Strong justification for over-labeling with description of learning-theoretic advantages.

*11: Three weak points of this paper (please number each point)
1. For performing experiments on document classification value of regularization parameter is fixed to $\lambda=0.1$. Is there any reason why this is done? Explanation should be provided.
2) Computational cost analysis is missing. It is one of the important concern especially when Time sheet is no longer avaliable
number of fine grained concepts become huge.
3. The graphs are hard to read. The graphs are too small, making the text hard to read. Additionally, the graphs with BANDIT vs. the FFR values have many lines with similar colors, making it difficult to read.

TODO
*12: Is this submission among the best 10% of submissions that you reviewed for ICDM'16?
[X] No
[_] Yes

*13: Would you be able to replicate the results based on the information given in the paper?
[X] No
[_] Yes

*14: Are the data and implementations publicly available for possible replication?
[X] No
[_] Yes

*15: If the paper is accepted, which format would you suggest?
[X] Regular Paper
[_] Short Paper

*16: Detailed comments for the authors
This paper introduces the “active over-labeling” method for learning hierarchically decomposable concepts, and also shows the use of a modified multi-armed bandit technique to automatically balance the proportion of labels sought at each level of the hierarchy

proportion of labels sought at each level of the hierarchy.

I enlarged the figures a little bit. 3 in a row is still too small, but position vertically taking too much space

Strong points:
The topics and methods in this paper are explained very clearly. Additionally, strong justification for over-labeling is provided in the section on learning-theoretic advantages. The proposed methods seem to be a unique and effective solution within the problem domain.
Weak points:
The graphs are a bit hard to read. Due to their small size, the text is very small. Also, in the BANDIT vs. FFR graphs, many of the lines have very similar colors and are thus hard to distinguish. BANDIT should at least have an easily distinguishable color; at the moment it looks similar to FFR[1.0]. The tables are helpful, but the paper reports that for Tables 2 and 3, BANDIT has the second lowest rank, when in fact it has the third lowest in each. Additionally, the paper states that FFR[0.0] is not affected by the fine cost, which makes sense. However, the graphs show that the FFR[0.0] line varies slightly with the fine cost, which doesn't seem to make sense and conflicts with the account given in the paper. This should be explained or rectified. Maybe a little more about future work could be written, especially since active over-labeling is a unique new method.

The fluctuation in FFR[0.0] will be smoother when averaged across more repeat

Author should consider adding following results to the experimental section:

Doug, when using logit for RCV1 with TF-IDF, how important is tuning lambda? Is there a paper that addresses this? If we can cite something that says lambda=0.1 is expected to do well, perhaps we don't need to tune it, since it's not a major part of the paper.

(i) Regularization parameter λ is fixed to 0.1 in the experiments which is not good in essence.
Why not tune the parameters? If data is scarce for tuning parameter than results should be

acknowledged that this parameter can be further tuned

demonstrated using different values of λ. May be the results might be completely different with
different lambda values.

(ii) Leaning fine grained concepts requires more computation time to learn parameters
corresponding to each of these concepts. Adding computational time of different approaches
might be helpful to understand the increase in runtime and also determine the feasibility of the
proposed approach for datasets with large number of fine grained classes.

Yuji can look up the run time per CV fold for coarse, for 0.1 fine, 0.2 fine, etc., and report them in a table. This should address this question on how much additional computational effort is required for fine. Also, report same info for BANDIT.

Some fine-grained label learners might overfit, but are we seeing any performance problems in the coarse-grained learning? We don't care if a fine class A instance is misclassified as B so long as it is one or the other, since both yield correct coarse classification. Do we have safety in numbers due to the large number of fine-grained classifiers?

(iii) Although adding performance has been shown to improve performance: there are few
concerns which requires to be explained in the discussion section: Possibility of over-fitting the
models with very few examples corresponding to fine grained classes, what if the fine grained
mentioned 10 examples minimum before participating votes
concepts can belong to multiple nodes like museum in the Figure 1. In such cases there is
Can't happen; see Footnote 1
possibility of learned models getting confused which may result in performance deterioration
instead of improvement.

3) Minor blemishes in the paper. For example: In section 4.4 - It should be Figure 3 instead of
Figure 4.1; In experiments - performance repeated twice and so on.

==
The review report from reviewer #3:

*1: Is the paper relevant to ICDM?
[_] No
[X] Yes

*2: How innovative is the paper?
[_] 6 (Very innovative)
[X] 3 (Innovative)
[_] -2 (Marginally)
[_] -4 (Not very much)
[_] -6 (Not at all)

*3: How would you rate the technical quality of the paper?
[_] 6 (Very high)
[X] 3 (High)
[_] -2 (Marginal)
[_] -4 (Low)
[_] -6 (Very low)

*4: How is the presentation?
[_] 6 (Excellent)
[_] 3 (Good)
[X] -2 (Marginal)
[_] -4 (Below average)
[_] -6 (Poor)

*5: Is the paper of interest to ICDM users and practitioners?
[_] 3 (Yes)
[X] 2 (May be)
[_] 1 (No)
[_] 0 (Not applicable)

*6: What is your confidence in your review of this paper?
[_] 2 (High)
[X] 1 (Medium)
[_] 0 (Low)

*7: Overall recommendation
[_] 6: must accept (in top 25% of ICDM accepted papers)
[X] 3: should accept (in top 80% of ICDM accepted papers)
[_] -2: marginal (in bottom 20% of ICDM accepted papers)
[_] -4: should reject (below acceptance bar)
[_] -6: must reject (unacceptable: too weak, incomplete, or wrong)

*8: Summary of the paper's main contribution and impact
Authors focus on classification task of high-level concepts that can be decomposed into a hierarchy of finer-grained subconcepts. As an example of application domain satisfying such assumption authors give document classification problem or biological sequence analysis. They show how classifiers aimed at course-grained tasks in such domains can be improved by training on fine-grained labels.

*9: Justification of your recommendation
Well written paper but its application limited to small number of application domains

*10: Three strong points of this paper (please number each point)

1) Interesting problem
2) Paper rather easy to follow

*11: Three weak points of this paper (please number each point)
1) Limited applications
2) Application domains may not be easy to identify.

*12: Is this submission among the best 10% of submissions that you reviewed for ICDM'16?
[X] No
[_] Yes

*13: Would you be able to replicate the results based on the information given in the paper?
[X] No
[_] Yes

*14: Are the data and implementations publicly available for possible replication?
[X] No
[_] Yes

*15: If the paper is accepted, which format would you suggest?
[_] Regular Paper
[X] Short Paper

*16: Detailed comments for the authors
In multimedia domains we have quite opposite situation. It is much easier to learn high-level concepts (classifiers have higher accuracy) than lower-level concepts. For instance, in music domain, if we take Hornbostel-Sachs classification, it is much easier to learn concept "aerophone" than lower-level concept "aero-double-reed".
The same situation is in automatic assignment of high-level tags vs lower-level tags to a video or image.
What about hierarchy-based automatic indexing of documents represented as text?
I would like see some discussion added on that subject.
There is a lot of quite powerful algorithms developed for automatic indexing of multimedia.

Sounds like R3 is talking about predicting labels at all levels, not just coarse

Can we adopt some of these strategies in a hierarchy scenario presented in this paper?
As a good example of automatic indexing in multimedia domain, which is based on hierarchical attributes, we can take paper
"Multi-Label Automatic Indexing of Music by Cascade Classifiers", W. Jiang, Z.W. Ras, in Web Intelligence
and Agent Systems, International Journal, IOS Press, Vol. 11, No. 2, 2013, 149-170

==

Maybe cite paper above and contrast its multi-level approach to our coarse-only one (emaphasize difference). Doing multi-level could be future work.

cited along with a fine grained text classification paper