Dear editors of Special Issue JIDM/KDMile-2014:

We are sending the revised manuscript entitled “Information Gain Feature Selection for Multi-Label Classification”, previously accepted at KDMile 2014.

Based on the advice received, we prepared a revised manuscript, considering the reviewers' comments, and submit a list of responses to these comments.

We would like to thank you for managing the process of the submission of this paper. We hope you and the reviewers now find the paper to be suitable for publication in Journal of Information and Data Management (JIDM) and look forward to your decision.

Best regards, the authors.

For all reviewers:

First of all, we would like to thank the reviewers for all comments. They were very useful and helped us to improve our manuscript. We have analyzed each of the comments, and the responses are included below.

For reviewer #1:

(1) There are two main issues that should be accounted for in the final version: Explain how the proposed method (MLInfoGain) works.

Answer: We have improved the explanation of the MLInfoGain technique in section 4. The relevant paragraph is the following: “The feature selection algorithm works as follows: it receives as input a multi-label data set. Then it computes the Entropy.ML measure defined before for each feature. Next, all the scores are sorted in a ranking. In order to have a list of selected features as an output, it is necessary to inform the number of selected features. This can be either a percentage of the total number of features or a score threshold to split the ranking. In this work we have opted for a percentage of features, in order to compare each technique with equal conditions.”

(2) Perform a statistical significance analysis of the results presented.

Answer: We have introduced a statistical analysis based on Friedman and Nemenyi post-hoc tests.

Also, all minor corrections were assessed.

For reviewer #2:

(1) There is much to learn in a thorough experimental analysis for this scenario. I just feel that the authors spent too much space detailing background information (4 pages plus a full page for references), whereas the experiments had to be summarized in a very short space. I think it would be important to find extra space for the analysis, and then show the results for all algorithms which were allegedly evaluated.

Answer: We have improved the experimental evaluation section with summarized results for all multi-label classifiers and new results for larger data sets.

(2) Furthermore, I think the paper should employ non-parametric statistical tests to verify the significance of the results (though one could argue that such a detailed analysis would only be needed for a journal paper). Significance tests add weight and credibility to empirical analyses.

Answer: We have introduced a statistical analysis based on Friedman and Nemenyi post-hoc tests.

(3) Finally, since the main contribution of the work is not necessarily its innovation value or originality (but the analysis and insight/lessons learned), I think the authors should make available online a .zip package with the whole source code of the experimental setup, plus detailed results' data, in order to allow proper reproducibility (insert the link for the online material within the text).

Answer: We pretend to include our Multi-label Information Gain adaptation for feature selection in the Mulan package (either in its core code or as a bundle package). This way all results can be reproduced exactly.

(4) Minor typos below (though I have to say the paper is a pleasure to read):

Answer: All typos were corrected. Thanks for pointing them out.

For reviewer #3:

(1) a discretização foi feita para cada rótulo, de maneira independente? No caso da relevância binária entendo que sim, mas e para os demais?

Answer: A discretização é feita no momento em que a métrica Information Gain é chamada para criar o ranking de atributos. Esta métrica considera valores de classe individualmente, então foi feita para cada rótulo. Para os algoritmos de transformação (LP, BR e Copy) foi feita a discretização supervisionada logo após a transformação (pois há a presença de uma classe única) e não-supervisionado para a adaptação (MLInfoGain), que não considera o valor da classe (pois nesse caso há múltiplas classes).

(2) o mesmo para a seleção de atributos. No caso da relevância binária, cada rótulo tem um subconjunto selecionado? Como funciona para os outros?

Answer: Na relevância binária os rankings são combinados (utilizando o método de agregação max), fazendo com que o resultado de seleção de atributos seja apenas um. Todos os demais métodos também têm como saída apenas um subconjunto de atributos.

(3) na tabela II, não ficou claro qual algoritmo multirrótulo foi usado como baseline na coluna "No Sel."

Answer: O algoritmo de classificação foi o BRKNN para todas as técnicas da tabela.

(4) a última linha da tabela II não deveria ser chamada de "Beats baseline", uma vez que podem também ocorrem empates.

Answer: Foi alterado para <= baseline.

For reviewer #4:

(1) I would give a better evaluation of the paper if the authors sold the idea of a faster multi-label feature selection.

Answer: We have included more results for the larger data sets, confirming that the proposed MLInfoGain performs indeed faster than the other assessed methods. We have added an emphasis on this benefit.

For reviewer #5:

(1) The work is very well organized and written. The problem is clearly defined and contextualized and the contribution of the work was properly highlighted. Very good contribution to KDMiLe.

Answer: We have included more results and a statistical analysis, based on Friedman and Nemenyi post-hoc tests. Thus, we expect to make the contribution even better.