Institute of Mathematics and Computational Sciences

ISSN - 0103-2569

A systematic review to identify feature selection publications in multi-labeled data

Newton Spolaôr
Maria Carolina Monard
Huei Diana Lee

N° 374

ICMC TECHNICAL REPORT

São Carlos, SP, Brazil
February/2012

A systematic review to identify feature selection publications in multi-labeled data[*] [†]

Newton Spolaôr[1,2]
Maria Carolina Monard[1]
Huei Diana Lee[2]

[1]Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
Laboratório de Inteligência Computacional
CEP 13560-970 - São Carlos, SP, Brasil

[2]Universidade Estadual do Oeste do Paraná
Centro de Engenharias e Ciências Exatas
Laboratório de Bioinformática
CEP 85867-900 - Foz do Iguaçu, PR, Brasil

February, 2012

[*]This work was supported by the São Paulo Research Council - FAPESP.
[†]Available on-line at: http://www.icmc.usp.br/~biblio/relatorios_tecnicos.php.

Abstract

Feature selection enables the identification of important features in data sets, contributing to an eventual increase in the quality of the knowledge extracted from them. A kind of data of growing interest is the multi-labeled one, which has more than one label for each data instance. However, there is a lack of reviews about publications of feature selection to support multi-label learning. To this end, the systematic review process can be useful to identify related publications in a wide, rigorous and replicable way. This work uses the systematic review process to answer the following research question: *what are the publications of feature selection in multi-labeled data?* The systematic review process carried out in this report enabled us to select 49 relevant publications and to find some gaps in the current literature, which can inspire future research in this subject.

Keywords: Systematic Review, Feature Selection, Feature Importance Measures, Multi-label Machine Learning

This document was written with the LaTeX text editor. The system of citation of bibliographical references uses the *Apalike* style of the bibTeX system.

Contents

1 Introduction

A research process can be specified as a sequence of activities that allows to obtain knowledge related to a subject. Bibliographical research, part of the research process, can be performed in a relatively more elaborated way through the Systematic Review (SR) process (Kitchenham, 2007).

The SR process enables to answer Research Questions (RQ) about a subject through previously specified activities to identify, select and evaluate publications (Castro et al., 2002). To this end, it explores the literature searching for relevant pieces of work in a fair, rigorous and replicable way. Specifically, in the evaluation of these pieces of work, the SR process can or cannot include meta-analysis, a form of study that synthesizes the results of the review through statistical techniques.

In Computer Science, there are several applications of the SR process in subjects related to the area of Software Engineering (Guessi et al., 2011; Kitchenham et al., 2010). Recently, there were some applications of this process related to the area of Artificial Intelligence, such as Feature Selection (FS) (Spolaôr et al., 2010).

The objective of FS can be defined as a search for important features in a given domain. As a consequence of the possible reduction of the number of features needed to represent the data, the "curse of dimensionality" effects can be reduced and the quality of knowledge obtained from learning can be improved (Liu and Motoda, 2008). Feature selection has traditionally been applied to the single-label problem, which has a unique label or class associated to each instance. Recently, FS has also been applied to the multi-label problem, where each instance is related to more than one label. Image and video annotation, bioinformatics and classifications of musics into emotions are tipical examples of multi-label problems (Tsoumakas et al., 2009). Furthermore, feature selection for multi-labeled data is the ongoing PhD research subject of the first author of this report.

However, there is a lack of reviews about publications of feature selection to support multi-label learning. Thus, this work contributes to reduce this gap through the use of the systematic review process, which does not include meta-analysis.

This work is organized as follows: Section 2 briefly describes the SR process. Section 3 describes the application of this method to identify publications of FS in multi-labeled data, and Section 4 presents the final conclusions.

2 Systematic Review Process

One of the first systematic review was published in 1955 (Pearson, 1904). It consists of a study about a clinical situation related to the placebo effect (Beecher, 1955). The popularity of this type of research in areas such as Medicine increased in the 80's and 90's (Castro et al., 2002). This growing interest is due to the elaboration of guidelines in Medicine (Higgins and Green, 2009), Social Science (Petticrew and Roberts, 2006) and Computer Science (Kitchenham, 2007; Biolchini et al., 2005) fields, among other reasons.

However, in Computer Science there are only recent applications of the systematic review process. For example, in the area of Software Engineering there are several systematic reviews, which were analyzed by Kitchenham et al. (2010). In the area of Artificial Intelligence, Spolaôr et al. (2010) performed one of the first SR to address publications of feature selection in single-labeled data. The results of this SR show a growing interest in the use of metaheuristics to perform single-label feature selection.

The systematic review process is composed of three steps (Kitchenham, 2007).

Step 1. Planning;

Step 2. Conducting;

Step 3. Reporting.

Step 1 involves specifying the research questions that must be answered and the creation of a protocol. The activities that integrate this protocol are carried out in Step 2 to identify a set of publications related to the researched subject. The last step is responsible to report the results obtained. These results are usually reported in PhD thesis, technical reports, articles or other formats.

Each step is composed of several activities described next, which can be executed concomitantly, allowing to improve themselves.

Step 1. Planning:

- Identification of the need for a review;
- Commissioning a review (optional);
- Specifying the research questions;
- Developing a review protocol;
- Evaluating the review protocol (optional).

Step 2. Conducting:

- Identification of research;
- Selection of publications;
- Study of quality assessment;
- Information extraction;
- Information synthesis.

Step 3. Reporting:

- Specifying the dissemination mechanisms;
- Formatting the main report;
- Evaluating the report (optional).

Specifying the research questions in Step 1 is one of the main activities carried out in the systematic review process, since it guides the development of the criteria contained in the protocol, the scope of the bibliographical review and the activities to be carried out in the other steps. At the end of the SR process, these research questions should be answered, highlighting its importance.

In Medicine, a practical way to perform this activity is by organizing the questions according to the medical concepts of Population (PO), Intervention (IE) and Outcome or evaluation metric (OU) (Petticrew and Roberts, 2006). These concepts are explained next, as well as their possible analogy to Computer Science:

PO. A patient group affected by the treatments under study.

An analogy for Computer Science could specify this concept as an application domain.

IE. Two or more treatments that are used in the population.

In Computer Science would be, for example, the mechanisms or methods which address a problem under study.

OU. Clinical and economic factors which enable the comparison between treatments.

In Computer Science could be a set of metrics that are used to compare the result obtained by algorithms which solve a given problem.

After formulating the research questions, it is possible to develop a review protocol to minimize potential bias during the application of the systematic review process (Kitchenham, 2007). A protocol is basically composed of the background on the subject studied and the description of the strategies which are used in the Step 2 of the SR process. An advantage of having a protocol is to support the replication of the SR process.

Identifying the publications which compose the research in a subject is an important activity carried out in Step 2. It is performed through the use of search strategies, which can be developed based on earlier systematic review and previous tests. An approach to specify a search strategy is based on the structure of the research questions (RQ). For each concept, topics should be established, such as words, lists of expressions[1] and its synonyms. The topics of each concept can then be grouped through the boolean operator OR and joined through the boolean operator AND.

The research question described in the following exemplifies this approach, which provides support to the generation of the three lists described in Table 1 (Spolaôr et al., 2010).

RQ What medical procedures and techniques can be used to identify evidences of breast cancer?

Population	Intervention	Evaluation Metrics
Breast cancer Carcinoma Malign tumor	Breast self-examination Mammogram Mammography	Sensitivity Specificity

Table 1: Lists of topics related to a research question.

The use of this approach allows to create the Search String (SS) described next.

SS (("breast cancer" OR carcinoma OR "malign tumor") AND ("breast self-examination" OR mammogram OR mammography) AND (sensitivity OR specificity))

The selection of publications is another important activity of the systematic review process. It can be performed with the support of inclusion/exclusion criteria and only publications that can answer the research questions should be kept.

An example of the exclusion criterion is the deletion of publications that were published before a specific year. Nevertheless, it is interesting to adopt a conservative posture during this activity, as the careless exclusion of a relevant publication implies in loss of information, which may affect the quality of the SR process. To check if a publication suits the selection criteria, it would be necessary to read the title, the abstract and the other parts of the publication, eventually including the whole publication. Although a publication with

[1] Sequence of words delimited by quotation marks.

a well structured abstract could support the selection activity, some abstracts are not well structured and the whole publication should be read.

The quality assessment of publications is performed with the use of the quality criteria, usually through checklists, which can be based in some models found in the literature (Fink, 2004). In this context, "quality" means the methodological merit of a study. These criteria contribute, for example, to correlate differences among the results in different publications and among the quality of these results. They also might suggest future trends on the subject of the systematic review process. The use of statistical tests in publications under analysis is an example of these criteria.

Checklists provide support to the use of two approaches (Kitchenham, 2007):

1. Specifying more detailed selection criteria;
2. Support to the analysis and synthesis of the information obtained from pieces of work that can answer the research questions.

The first approach demands a separated form to extract information from the new selected publications, while the second one allows the use of a unique form.

The synthesis activity in Step 2, which can be quantitative or qualitative, supports the summarization and organization of the information extracted from the publications. The first one enables the use of meta-analysis and considers numerical information, such as size of samples, accuracy and standard deviation. Thus, differences among publications can be highlighted. On the other hand, qualitative synthesis allows to identify similarities among publications and can use approaches as the one suggested by Noblit and Hare (1988).

Other examples of research questions and criteria, as well as a wider introduction to the systematic review process, are described in (Spolaôr et al., 2010).

In the following, the use of the systematic review process to search for publications about feature selection to support multi-label learning is described.

3 Systematic Review Application

The SR process related to this report focuses on the identification of publications that involve the use of feature selection in multi-labeled data. This process was carried out during 10 days (July 26 - August 05 of 2011) at the Institute of Mathematics and Computer Science, University of São Paulo. In what follows, the three steps of the systematic review process are described.

3.1 Planning

As already mentioned, the motivation to perform a systematic review on feature selection in multi-labeled data includes the need to increase the background on this subject, as well as to identify related publications. It should be observed that there were not found previous SR related to publications about feature selection to support multi-label learning. The protocol developed in this report is based on a previous protocol which was used to investigate the use of metaheuristics to perform feature selection in single-labeled data (Spolaôr et al., 2010).

Initially, as this is the first SR on feature selection for multi-labeled data, a unique research question was specified.

RQ What are the publications of feature selection in multi-labeled data?

In the future, in the case of identifying a great number of publications using this unique research question, it would be possible to refine it, as well as formulating more specific research questions.

The protocol of the SR was developed using the initial background described by Spolaôr et al. (2011). The search string was obtained according to the concepts of population (41 topics) and intervention (70 topics). The population was related to multi-label learning because it is a problem that can be studied with the support of feature selection. In other words, feature selection was considered as the intervention mechanism which contributes to the study of multi-labeled data. The topics and search string used are described in the Appendix.

The sources selected to find the publications were: ACM Portal[2], CiteSeer^{X}[3], IEEE Xplore[4], ScienceDirect[5], Scopus[6], Wiley Interscience[7] and Web of Science[8]. In some cases, the search string was adapted to suit site limitations, such as the maximum number of topics.

The adaptations include decomposition of the search string into smaller ones and the posterior union of the results. Furthermore, the scope of the search string was limited to the title, abstract and keywords, whenever the source supported this requirement.

The activity to select the relevant publications consists of two main procedures. First of all, the title and abstract of the retrieved publication is read.

[2] http://portal.acm.org
[3] http://citeseerx.ist.psu.edu
[4] http://ieeexplore.ieee.org
[5] http://www.sciencedirect.com
[6] http://www.scopus.com
[7] http://onlinelibrary.wiley.com
[8] http://isiknowledge.com

If the information in the title and abstract is not enough to select the publication, other sections of the publication are read to make a decision. In this work we specified the 14 selection criteria described next, all of them being exclusion criteria.

- Publications that do not suit the RQ;
- Duplicated publications by the same authors[9]. In this case, only one is kept;
- Publications which focus on selection of rules, prototypes, classifiers, parameters, architectures and reducts, as well as selections do not related to feature selection;
- Publications that focus on feature extraction[10];
- Publications that also perform label selection;
- Publications that also perform feature clustering;
- Publications that only perform simplistic feature selection based on frequency[11];
- Publications that do not address explicitly multi-labeled data;
- Tutorial slides;
- Publications composed of only one page (abstract papers), posters, presentations, proceedings and program of scientific events;
- Publications hosted in web pages which are not accessed through the account of the University of São Paulo;
- Publications written in a language different than English.

After the selection of the publications, nine quality criteria were applied according to the second approach described in Section 2. These quality criteria are described next.

1. Is the feature selection method related to multi-label learning the main goal of the publication?
2. Is the main feature selection method proposed in the publication compared to other feature selection methods?

[9] Similar title, abstract, results or text.
[10] Feature construction, which increases the data dimensionality description.
[11] Measures as feature frequency in the set of documents exemplify this scenario.

3. Does the evaluation strategy of the method described in the publication involve a statistical analysis of significance? (Crombie, 1996)

4. Is the filter approach the one used for feature selection? (Spolaôr et al., 2010)

5. Are the experimental results compared to previous results? (Crombie, 1996)

6. Is the parameter setting of the feature selection method described in the publication justified? (Ali et al., 2009)

7. How many data sets are used in the publication?

8. Are there synthetic data sets used in the publication?

9. Is label dependence considered by the feature selection method used in the publication?

The form constructed after this activity consists of a LibreOffice[12] electronic sheet with 23 columns. The use of an electronic sheet enables a simple analysis of the quality criteria. For example, counting the data sets used in each publication, which satisfy a quality criteria, can be easily performed in this environment. The 23 columns of the LibreOffice's form are described next.

1. Publication's ID;

2. Feature selection approach used;

3. Year of publication;

4. Authors' name;

5. Name of other feature selection methods which are used to compare with the main feature selection method proposed in the publication;

6. Comparison with previous published results;

7. Parameter setting of the feature selection method;

8. Description of the data sets used: name, number of examples, number of features, number of classes, type[13], data pre-processing employed and domain;

9. Scope of the feature selection method[14];

[12] http://www.libreoffice.org
[13] Benchmark, real or synthetic.
[14] Global or local (Esuli et al., 2006).

10. Multi-label learning approach used by the feature selection method[15];
11. Strategy, learning algorithm and measures used to evaluate the feature selection method;
12. Publication's source;
13. Investigation of label dependence in multi-label learning by the feature selection method;
14. Justification of the parameter setting used;
15. Feature importance measures used;
16. Best result obtained for each evaluation measure;
17. Search method used to identify subsets of features;
18. Motivation for using feature selection methods;
19. Main objective of the publication regarding the feature selection method;
20. Main objective of the publication;
21. Restrictions related to the application of the proposed feature selection method;
22. Name and confidence interval of the statistical test applied;
23. Observations.

We performed a qualitative synthesis of the information described in the form, according to the majority of systematic reviews carried out in the area of Software Engineering (Brereton et al., 2007). Some quality criteria of the qualitative synthesis, which we consider important for the development of our work in feature selection to support multi-label learning, are highlighted in Section 3.3.

3.2 Conducting

The use of the search string in the sources selected allowed us to identify more than a thousand publications. However, in many cases, the same publication was found by two or more sources. Therefore, semi-automatic removal of publications with the same title was carried out. To this end, a simple computational framework was implemented in order to remove publications with

[15]Problem transformation or algorithm adaptation (Tsoumakas et al., 2009).

identical titles. Afterwards, the exclusion criteria described in Section 3.1 were applied, resulting in an aggressive reduction of the number of publications.

Table 2 summarizes these results. It shows the number of publications (♯publications) found after the first activity, as well as the number of publications and the percentage of reduction (%reduction) after applying each one of the exclusion criteria procedures. It can be observed that after applying our computational framework, the original number of publications found, 1630, was reduced to 885 (45.7% reduction). Two more publications were removed manually (883) and, after analysing the publication title, this number was reduced to 645. Next, the abstracts were analyzed and 214 publications were left to be considered. Finally, after reading other sections of these 214 publications, only 47 of them were selected, which represent 2.9% of the original publication. In other words, the application of the exclusion criteria procedure yielded a 97.1% reduction.

	♯publications	%reduction
Publications identified on the first activity	1630	0
Unique publications (automatic)	885	45.7
Unique publications (manual)	883	45.8
Selection (after reading the title)	645	60.4
Selection (after reading the abstract)	214	86.9
Selection (after reading the other sections)	47	97.1

Table 2: Summary of the application of the exclusion criteria procedures.

Despite the initial identification of a high number of publications, there were two publications addressing the research question, previously known by us, which were not found. More specifically, the publications (Trohidis et al., 2008), which was not indexed in the searched sources, and (Zheng et al., 2004), which was not addressed by the search string. These two publications were included manually in the 47 publications selected, totalizing 49 publications.

The next activity performed was the quality assessment of the 49 final publications. This activity was carried out using the information extracted from the 49 publications, already structured in the electronic sheet. Next section describes the synthesis activity performed on the information extracted from all the 49 publications, as well as the manner we are using to reporting to the community our systematic review results.

3.3 Reporting

The following dissemination mechanisms were selected to report the results:

1. Hosting the results in a web site at University of São Paulo[16];

2. Disseminating the site to the community, with special attention to our research collaborators from UFABC[17] and UNIOESTE[18]. The eletronic sheet form holding the information extracted from the selected publications can also be obtained from the authors of this systematic review.

The synthesis of the 49 selected publication is qualitative as well as quantitative. Figure 1 presents the quantitative summary of the number of publications during the last years, where it is possible to observe a general trend of growing interest in the publications of feature selection methods for multi-labeled data. The 38 papers published since 2007 represent nearly 77.6% of the total ones selected.

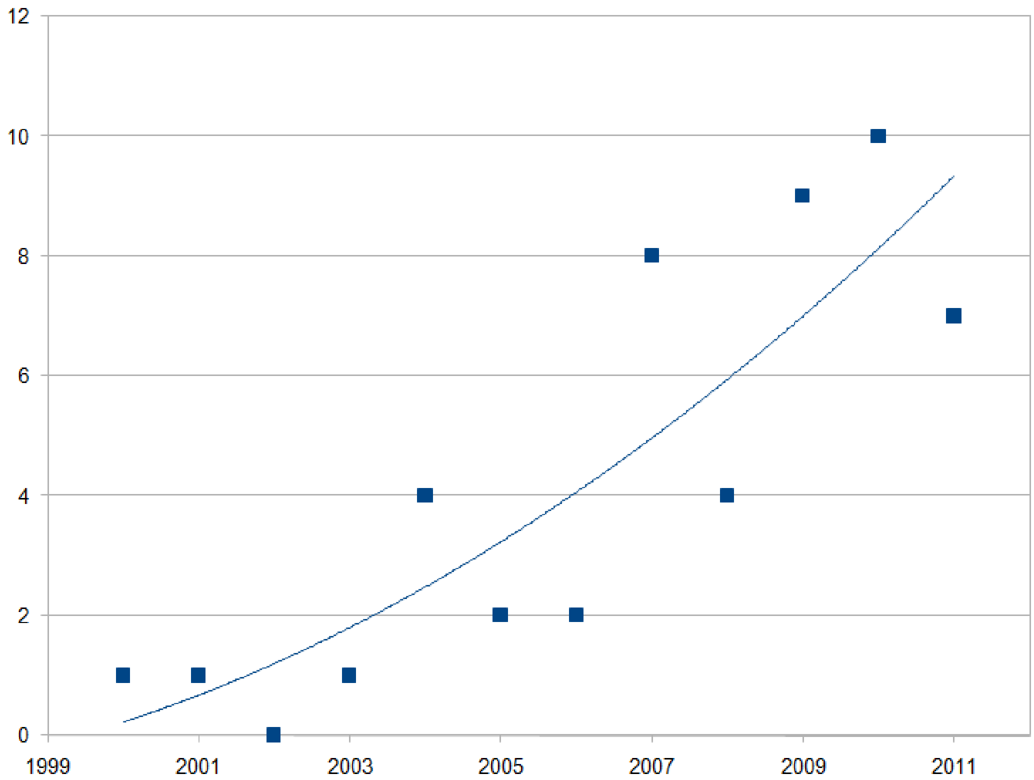

Figure 1: Trend in the number of publications $\times$ year of publication.

The quality criteria enable us to make some interesting observations related to the 49 publications[19] found by the SR process.

- In most publications (83.7%), the proposed method is compared to other feature selection methods. Few publications (20.4%) perform statistical

[16]http://www.icmc.usp.br/~biblio/relatorios_tecnicos.php
[17]http://dgp.cnpq.br/buscaoperacional/detalhegrupo.jsp?grupo=IWU4103700AHR2
[18]http://www.foz.unioeste.br/labi
[19]See (Tsoumakas et al., 2009; Liu and Motoda, 2008) for more detail about feature selection and multi-label learning.

analysis of significance. Only seven publications (14.3%) compare their results with experimental results already published;

- Only 23 publications (46.9%) consider feature selection for multi-labeled data as its main goal. Other publications, for example, focus in the development of methods for multi-label classification or the transformation of multi-labeled data to a single-labeled one. In many cases, an algorithm for feature selection is only applied in some of the data sets to address a specific aspect of the work;

- Most of the selected publications (61.2%) apply multi-label feature selection according to the filter approach. This result differs from the one observed in single-label feature selection (Spolaôr et al., 2010), possibly because many multi-labeled data sets for text categorization have hundreds or thousands of features. Therefore, the lower computational cost usually related to the filter approach can be more useful in these cases;

- The parameter setting of the feature selection method is justified in 17 publications (34.7%). Many of them are related with the definition of the number of features to be selected. Parameters setting justifications could be useful to guide future work;

- Nearly half of the publications (42.9%) investigate two or more data sets using feature selection methods. However, only in few cases data sets of different kind (synthetic, benchmark or real) are used. From the three publications (6.1%) that use synthetic data sets, only two of them (4.1%) use other kind of data set (benchmark or real). It is worth observing that there is a lack of publications that address synthetic multi-labeled data sets;

- There is also a lack of publications that address an important issue: label dependence together with feature selection methods. Only four publications (8.2%) address both aspects, although relevant references on multi-label learning sustain that the study of label dependence is an important issue, which should be taken into account to search for models with better performance (Tsoumakas et al., 2009).

Among all quality criteria, the one related to label dependence is one of the most prominent due to its importance for the multi-label learning community. Thus, the strategies adopted in the four publications concerned with label dependence are described next.

Doquire and Verleysen (2011). Study of feature dependence and label dependence. The first dependence is related to the selection of features subsets. The second one is related to the application of another problem transformation approach that has different characteristics, such as pruning, to simplify the data and to ensure that all classes are represented by a specific number of instances (Read, 2008);

Wei et al. (2009). Description of a feature selection method used in single-label data sets obtained from a multi-label data set. Each single-label data set includes some labels as features. This approach has similarities with another problem transformation method proposed recently (Cherman et al., 2010). However, Wei et al. (2009) did not describe in details how FS is performed in the data;

Zhang et al. (2009). Use of Ranking Loss (Tsoumakas et al., 2009) as a feature importance measure, enabling the optimization of quality between output labels and, consequently, the analysis of correlations between them. This is the unique publication that studies label dependence through the FS wrapper approach;

Trohidis et al. (2008). Analysis of label dependence through the use of the Label Powerset (Tsoumakas et al., 2009), which is an example of the problem transformation approach. Basically, this approach converts the multi-labeled data into a single-labeled multi-class data. Each unique set of labels in the training data is converted to a new single-label. A feature selection method is applied in the transformed data and contributes to obtain better results related to other approaches.

It is also important to comment about other aspects of the publications related to feature selection methods. The scope of the feature selection methods can be organized as global or local (Esuli et al., 2006). The global scope is related to the selection of the same subset of features for all categories, such as exemplified in our previous study (Spolaôr et al., 2011). The local scope can identify a unique subset in each category. It was identified with the systematic review carried out that there is little difference between the number of publications in each scope. In any case, it should be observed that the local scope has one disadvantage: no feature will be removed whenever the union of the identified subsets of features from all categories is equal to the full feature subset.

There are also differences related to the frequency of use of each multi-label approach together with the FS method. Two important approaches are the problem transformation and the algorithm adaptation. The second one consists in the use of a method for multi-label learning that addresses multi-labeled data directly. Most of the publications selected are classified according to the problem transformation approach, possibly because it enables the use of traditional single-label feature selection methods.

In the selected publications, text categorization is the most frequent data set domain/problem studied by FS methods for multi-label learning. As stated before, one reason could be that this domain has usually hundreds or thousands of features, for which FS could have relevant contributions. Other frequent domains are image annotation/classification, gene annotation and emotion analysis.

4 Final Highlights

In this report we described the use of the systematic review process to find publications related to feature selection to support multi-label learning. A brief introduction about the SR process, including examples for some of its activities, as well as relevant references for feature selection and multi-label learning, are presented.

The systematic review process is an interesting method for bibliographical research that allows a wide, rigorous and reproducible literature exploration. The results and highlights presented in this report enabled us to carry out a relevant investigation about the research subject treated. These advantages compensate the additional effort needed to carry out a systematic review process.

The systematic review process allowed us to select 49 publications that answer the research question. We showed a publication trend graph that suggests a growing interest in the subject of feature selection methods to support multi-label learning. However, only few of these publications satisfy several quality criteria, such as using synthetic data sets, comparison with previous published results and analysis of label dependence. These aspects represent gaps that can be addressed in the future by the community.

One limitation of this work is that sources as SpringerLink, Scirus and Google Scholar were not selected to identify publications, due to the fact that restrictions to the insertion of long search strings make the use of these sources complex. Another limitation is that two important publications were not identified by the SR process, motivating the improvement of the keywords for the search string.

The protocol proposed in this report, including the search string, the selection criteria and the quality criteria, could be used in forthcoming surveys and related research related to publications on feature selection in multi-labeled data. Furthermore, the application of the SR process in Artificial Intelligence related areas could also use portions of this report as an initial support.

Acknowledgements

This research was supported by the Brazilian Research Council FAPESP. We would like to thank Ana Carolina Lorena, Bianca Espindola and Everton Alvares Cherman for their collaboration.

A Appendix

Topics of population and intervention, as well as search string used into the SR process.

PO "multi-label", "multilabel", "multi label", "multiple label", "multiple labels", "label correlation", "label correlations", "correlation of label", "correlations of label", "correlation of labels", "correlations of labels", "label set", "label sets", "set of label", "sets of label", "set of labels", "sets of labels", "label relationship", "label relationships", "relationship of label", "relationships of label", "relationship of labels", "relationships of labels", "label dependence", "label dependencies", "dependence of label", "dependencies of label", "dependence of labels", "dependencies of labels", "label co-occurrence", "label co-occurrences", "co-occurrence of label", "co-occurrences of label", "co-occurrence of labels", "co-occurrences of labels", "label combination", "label combinations", "combination of label", "combinations of label", "combination of labels", "combinations of labels".

IE "feature selection", "feature reduction", "feature ranking", "attribute selection", "attribute reduction", "attribute ranking", "variable selection", "variable reduction", "variable ranking", "gene selection", "gene reduction", "gene ranking", "feature subset selection", "feature subset reduction", "attribute subset selection", "attribute subset reduction", "variable subset selection", "variable subset reduction", "gene subset selection", "gene subset reduction", "selection of feature", "selection of features", "reduction of feature", "reduction of features", "ranking of feature", "ranking of features", "selection of attribute", "selection of attributes", "reduction of attribute", "reduction of attributes", "ranking of attribute", "ranking of attributes", "selection of variable", "selection of variables", "reduction of variable", "reduction of variables", "ranking of variable", "ranking of variables", "selection of gene", "selection of genes", "reduction of gene", "reduction of genes", "ranking of gene", "ranking of genes", "selection of feature subset", "selection of feature subsets", "selection of attribute subset", "selection of attribute subsets", "selection of variable subset", "selection of variable subsets", "selection of gene subset", "selection of gene subsets", "reduction of feature subset", "reduction of feature subsets", "reduction of attribute subset", "reduction of attribute subsets", "reduction of variable subset", "reduction of variable subsets", "reduction of gene subset", "reduction of gene subsets", "ranking of feature subset", "ranking of feature subsets", "ranking of attribute subset", "ranking of attribute subsets", "ranking of variable subset", "ranking of variable subsets", "ranking of gene subset", "ranking of gene subsets", "dimensionality reduction", "reduction of dimensionality".

SS ("feature selection" OR "feature reduction" OR "feature ranking" OR "attribute selection" OR "attribute reduction" OR "attribute ranking" OR "variable selection" OR "variable reduction" OR "variable ranking" OR "gene selection" OR "gene reduction" OR "gene ranking" OR "feature subset selection" OR "feature subset reduction" OR "attribute subset selection" OR "attribute subset reduction" OR "variable subset selection" OR "variable subset reduction" OR "gene subset selection" OR "gene subset reduction" OR "selection of feature" OR "selection of features" OR "reduction of feature" OR "reduction of features" OR "ranking of feature" OR "ranking of features" OR "selection of attribute" OR "selection of attributes" OR "reduction of attribute" OR "reduction of attributes" OR "ranking of attribute" OR "ranking of attributes" OR "selection of variable" OR "selection of variables" OR "reduction of variable" OR "reduction of variables" OR "ranking of variable" OR "ranking of variables" OR "selection of gene" OR "selection of genes" OR "reduction of gene" OR "reduction of genes" OR "ranking of gene" OR "ranking of genes" OR "selection of feature subset" OR "selection of feature subsets" OR "selection of attribute subset" OR "selection of attribute subsets" OR "selection of variable subset" OR "selection of variable subsets" OR "selection of gene subset" OR "selection of gene subsets" OR "reduction of feature subset" OR "reduction of feature subsets" OR "reduction of attribute subset" OR "reduction of attribute subsets" OR "reduction of variable subset" OR "reduction of variable subsets" OR "reduction of gene subset" OR "reduction of gene subsets" OR "ranking of feature subset" OR "ranking of feature subsets" OR "ranking of attribute subset" OR "ranking of attribute subsets" OR "ranking of variable subset" OR "ranking of variable subsets" OR "ranking of gene subset" OR "ranking of gene subsets" OR "dimensionality reduction" OR "reduction of dimensionality"**) AND (**"multi-label" OR "multilabel" OR "multi label" OR "multiple label" OR "multiple labels" OR "label correlation" OR "label correlations" OR "correlation of label" OR "correlations of label" OR "correlation of labels" OR "correlations of labels" OR "label set" OR "label sets" OR "set of label" OR "sets of label" OR "set of labels" OR "sets of labels" OR "label relationship" OR "label relationships" OR "relationship of label" OR "relationships of label" OR "relationship of labels" OR "relationships of labels" OR "label dependence" OR "label dependencies" OR "dependence of label" OR "dependencies of label" OR "dependence of labels" OR "dependencies of labels" OR "label co-occurrence" OR "label co-occurrences" OR "co-occurrence of label" OR "co-occurrences of label" OR "co-occurrence of labels" OR "co-occurrences of labels" OR "label combination" OR "label combinations" OR "combination of label" OR "combinations of label" OR "combination of labels" OR "combinations of labels"**)**

References

Ali, S., Briand, L., Hemmati, H., and Walawege, R. K. P. (2009). A systematic review of the application and empirical investigation of search-based test-case generation. Technical report, Simula Research Laboratory. Available at: http://simula.no/research/se/publications/Simula.SE.293/simula_pdf_file. Cited in page 8.

Beecher, H. K. (1955). The powerful placebo. *Journal of American Medical Association*, 159:1602–1606. Cited in page 2.

Biolchini, J., Mian, P. G., Natali, A. C. C., and Travassos, G. H. (2005). Systematic review in software engineering. Technical report, Alberto Luiz Coimbra Institute - Graduate School and Research in Engineering/Federal University of Rio de Janeiro - Brazil. Available at: http://www.cin.ufpe.br/~in1037/leitura/systematicReviewSE-COPPE.pdf. Cited in page 2.

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., and Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4):571–583. Cited in page 9.

Castro, A. A., Saconato, H., Guidugli, F., and Clark, O. A. C. (2002). A systematic review and meta-analysis course (in portuguese). Available at: http://www.virtual.epm.br/cursos/metanalise. Cited in page 1 and 2.

Cherman, E., Metz, J., and Monard, M. C. (2010). A simple approach to incorporate label dependency in multi-label classification. In *Mexican International Conference on Artificial Intelligence - Advances in Soft Computing*, pages 33–43. Cited in page 13.

Crombie, I. K. (1996). *The Pocket Guide to Appraisal.* British Medical Journal Books. Cited in page 8.

Doquire, G. and Verleysen, M. (2011). Feature selection for multi-label classification problems. *Lecture Notes in Computer Science*, 6691:9–16. Cited in page 12.

Esuli, A., Fagni, T., and Sebastiani, F. (2006). TreeBoost.MH: A boosting algorithm for multi-label hierarchical text categorization. *Lecture Notes in Computer Science*, 4209:13–24. Cited in page 8 and 13.

Fink, A. G. (2004). *Conducting Research Literature Reviews - From the Internet to Paper.* Sage Publication. Cited in page 5.

Guessi, M., de Oliveira, L. B. R., and Nakagawa, E. Y. (2011). Current state of representation of reference architectures. Technical report, University of São Paulo - Brazil. Available at: http://www.icmc.usp.br/~biblio/relatorios_tecnicos.php. Cited in page 1.

Higgins, J. P. T. and Green, S. (2009). Cochrane handbook for systematic reviews of interventions. The Cochrane Collaboration. Available at: http://www.cochrane-handbook.org/. Cited in page 2.

Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O. P., Turner, M., Niazi, M., and Linkman, S. (2010). Systematic literature reviews in software engineering - a tertiary study. *Information and Software Technology*, 52(8):792–805. Cited in page 1 and 2.

Kitchenham, B. A. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report, Evidence-based Software Engineering - United Kingdom. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.154.1446&rep=rep1&type=pdf. Cited in page 1, 2, 3, and 5.

Liu, H. and Motoda, H. (2008). *Computational Methods of Feature Selection*. Chapman & Hall/CRC. Cited in page 1 and 11.

Noblit, G. W. and Hare, R. D. (1988). *Meta-Ethnography: Synthesizing Qualitative Studies*. Sage Publications. Cited in page 5.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, pages 1243–1246. Cited in page 2.

Petticrew, M. and Roberts, H. (2006). *Systematic Review in the Social Sciences: A Practical Guide*. Wiley-Blackwell. Cited in page 2 and 3.

Read, J. (2008). A pruned problem transformation method for multi-label classification. In *New Zealand Computer Science Research Student Conference*, pages 143–150. Cited in page 13.

Spolaôr, N., Cherman, E. A., and Monard, M. C. (2011). Using ReliefF to select features in multi-labeled data (in portuguese). In *Latin-american Informatics Conference*, pages 960–975. Cited in page 6 and 13.

Spolaôr, N., Lorena, A. C., and Lee, H. D. (2010). A systematic review on multiobjective metaheuristics for feature selection (in portuguese). Technical report, Federal University of ABC - Brazil. Available at: cmcc.ufabc.edu.br/index.php/relatorios-de-pesquisa/14-2010/117-technical-report-032010.html. Cited in page 1, 2, 4, 5, 6, 8, and 12.

Trohidis, K., Tsoumakas, G., Kalliris, G., and Vlahavas, I. (2008). Multi-label classification of music into emotions. In *International Conference on Music Information Retrieval*, pages 1–6. Cited in page 10 and 13.

Tsoumakas, G., Katakis, I., and Vlahavas, I. (2009). Mining multi-label data. *Data Mining and Knowledge Discovery Handbook*, pages 1–19. Cited in page 1, 9, 11, 12, and 13.

Wei, Q., Yang, Z., Junping, Z., and Wang, Y. (2009). Semi-supervised multi-label learning algorithm using dependency among labels. In *International Conference on Machine Learning and Computing*, pages 112–116. Cited in page 13.

Zhang, M.-L., Peña, J. M., and Robles, V. (2009). Feature selection for multi-label Naive Bayes classification. *Information Sciences*, 179:3218–3229. Cited in page 13.

Zheng, Z., Wu, X., and Srihari, R. (2004). Feature selection for text categorization on imbalanced data. *Special Interest Group on Knowledge Discovery and Data Mining Explorations Newsletter*, 6:80–89. Cited in page 10.