Adoption of Machine Learning Techniques to Perform Secondary Studies: A Systematic Mapping Study for the Computer Science Field

Leonardo Sampaio Cairo[1], Glauco de Figueiredo Carneiro[1] and Bruno C. da Silva[2]
[1]*Universidade Salvador (UNIFACS), BA, Brazil*
[2]*California Polytechnic State University (Cal Poly), San Luis Obispo, CA, U.S.A.*

Keywords: Machine Learning, Text Mining, Systematic Mapping, Systematic Literature Review, Secondary Studies.

Abstract: **Context:** Secondary studies such as systematic literature reviews (SLR) have been used to collect and synthesize empirical evidence from relevant studies in several areas of knowledge, including Computer Science. However, secondary studies are time-consuming and require a significant effort from researchers. **Goal:** This paper aims to identify contributions derived from the adoption of machine learning (ML) techniques in Computer Science SLRs. **Method:** We performed a systematic mapping study querying well-known repositories and first found 399 studies as a result of applying the search string in each of the selected search engines. Following the research protocol, we analyzed titles and abstracts and applied inclusion, exclusion and quality criteria to finally obtain a set of 17 studies to be further analyzed. **Results:** The selected papers provided evidence of relevant contributions of the machine learning usage in performing secondary studies. We found that ML techniques have not been applied yet to all the stages of a SLR. Typically, the preferred stage to apply ML in an SLR is the study selection phase (typically the initial phase). For assessing the effectiveness of ML support while performing SLRs, researchers have provided a comparison either across different ML techniques tested or between manual and ML-supported SLRs. **Conclusion:** There is significant evidence that the use of machine learning applied to SLR activities (especially the study selection activity) in Computer Science is feasible and promising, and the findings can be potentially extended to other research fields. Also, there is a lack of studies exploring ML techniques for other stages than study selection.

1 INTRODUCTION

Secondary studies are typically performed as Systematic Literature Reviews (SLR). An SLR is a research method for identifying, evaluating and interpreting relevant research papers available focusing on a specific topic, thematic area, or phenomenon of interest. There are many reasons to perform an SLR, such as summarizing existing evidence for a treatment or technology and identifying gaps in current research to suggest areas for additional research. Furthermore, they can be a mean to examine to what extent the empirical evidence supports/contradicts theoretical hypotheses (Kitchenham and Charters, 2007).

Systematic Mapping Studies (SMS) are also classified as secondary studies. An SMS has the goal to review primary studies related to specific topic, thematic area, or phenomenon of interest represented as research questions (RQs) to integrate/synthesize evidence related to those RQs. The result of performing secondary studies is mainly the potential ability to combine data from several studies and provide both a panoramic and in-depth characterization from the perspective of the target research questions. These benefits can at least partly explain why secondary studies have been gaining popularity over the years.

The number of research studies published in Computer Science is continually expanding, and secondary studies have become essential tools for researchers to keep up to date in their particular fields. However, SLRs require considerable effort (Petersen et al., 2008), especially in the cases when the SLR activities are performed manually. For this reason, automating the activities of a SLR can reduce the required effort and also increase the coverage of evidence to support the answers for the stated research questions.

Therefore, we performed a systematic mapping study guided by the following research questions: **RQ1:** Which machine learning techniques have been applied by researchers and practitioners to support the execution of SLRs in Computer Science? **RQ2:** How

Cairo, L., Carneiro, G. and C. da Silva, B.
Adoption of Machine Learning Techniques to Perform Secondary Studies: A Systematic Mapping Study for the Computer Science Field.
DOI: 10.5220/0007780603510356
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 351-356
ISBN: 978-989-758-372-8

have researchers and practitioners evaluated the effectiveness of the machine learning techniques to support SLRs in Computer Science?

The goal of RQ1 is to provide an overview on how existing machine learning techniques cope with the automation of activities in secondary studies. It considers the adoption of both specialized supervised and unsupervised ML algorithms to target the aforementioned automation. In addition, we are interested in the evaluation of effectiveness of the machine learning support to this automation (RQ2).

The remainder of this paper is organized as follows. Section 2 describes the research method used in this systematic mapping. In Section 3, we discuss the results based on evidence obtained from the literature. Section 4 presents the conclusion, threats to validity, and scope for future work.

2 THE METHODOLOGY

This section describes the methodology applied in the planning, execution and documentation phases of our systematic mapping. Unlike an unstructured review, this mapping follows a precise and rigorous sequence of methodological steps to review the literature available in electronic databases.

Our goal is to analyze the current state-of-the-art on machine learning techniques applied to secondary studies in the area of computer science. Thefore, this mapping study intended to answer the research questions we introduced in the previous section.

2.1 Search for Primary Studies

We target the search on the following digital databases: ACM Digital Library, IEEE Xplore, and Scopus. ACM and IEEE digital libraries are the most relevant ones in Computer Science (Zhang et al., 2011) whereas Scopus is the world largest database for peer-reviewed research literature. After the fine tuning and preliminary analysis of retrieved results, we ended up with the following search string:

(“systematic literature review” OR “slr” OR “systematic review” OR “systematic mapping” OR “mapping study” OR ”secondary study”) AND (“machine learning” OR “text mining” OR “nlp” OR “natural language processing” OR “text analytics” OR “information retrieval”)

Regarding the period covered in our search, we covered all papers published in peer-reviewed magazines, journals, and conferences until March 2018, when we last applied the Search String.

2.2 Selection of Primary Studies

The following steps guided the selection of primary studies.

Stage 1 - Results Obtained from Automatic Applying Our Search String on the Digital Libraries. We converted the search string to each specific syntax of the repositories, and we always applyed the search string to the title, keywords and abstract.

Stage 2 - Reading Titles and Abstracts to Identify Qualifying Studies. Identification of eligible studies, based on the title, abstract and content analysis in some cases, ruling out studies that were clearly irrelevant to the review. This activity was performed by the three researchers co-authors of this paper. When we raised questions about the eligibility of a study, we marked the paper for further discussion, and then we went over all the marked papers to a debate over raised questions. In the end, we debated over the papers that had a different classification in this stage in order to come to a consensus.

Stage 3 - Applying Inclusion and Exclusion Criteria When Reading the Full Text. We defined that entries must meet all of the following *Inclusion Criteria* listed bellow: *IC1:* Published papers describing the use of ML techniques in the execution of Computer Science secondary studies. *IC2:* When several papers report the same study, only the most recent one should be included. *IC3:* Papers published in peer-reviewed computer science conferences, magazines, and journals. *IC4.* Works written in English. Regarding the **Exclusion Criteria**, this Systematic Mapping discarded papers that met at least one of the following: *EC1:* Studies that do not describe the Machine Learning technique used. *EC2:* Studies that are only available in a summary form or presentation notes (slides). *EC3.* Book chapters and other materials that have not undergone peer-review.

Stage 4 - Obtaining Primary Studies and Performing a Critical Evaluation. We obtained a list of primary studies which was subsequently subject to critical examination using the following **Quality Criteria** (QC): *QC1:* Is the document based on research or is it just a “lessons learned” report based on expert opinion? *QC2:* Is there an adequate description of the context in which the search was performed? *QC3:* Is there a control group with which to compare treatments? *QC4:* Is there a clear statement of results?

2.3 Conducting the Review

We initiated the review with an automatic search in the repositories, followed by a manual search on references of backward snowballing (Jalali and Wohlin, 2012), to identify potentially relevant studies. Next, the inclusion and exclusion criteria were applied. It was necessary to adapt the search string according to the syntax of each repository. The manual search consisted of studies previously known and published in conference proceedings and computer science journals and/or secondary studies that were included by the authors while researching the theme in different repositories.

During the course of this review, we used the tool Mendeley [1] to manage references collaboratively. All annotations and classifications were performed using Mendeley by using tags, folders and annotations made directly in the PDF files of each study.

2.4 Data Extraction

During this phase, we extracted data from each of the 17 primary studies to answer the research questions of this systematic mapping. We registered the obtained data as Mendeley notes e exported them using the Bibtex format supported by JabRef [2]. We organized the extracted data in HTML format with the following fields: Study Identification (S1, ..., S17), Authors, Study Title, Abstract, Research Questions, Year, Journal/Conference/Periodical and Source Repository.

2.5 Potentially Relevant Studies

We included in our Mendeley library the results obtained from the automatic and manual search. At this phase, a total of 399 studies were recorded, 395 of the automated search plus four manual search (Phase 1). Then we read titles and abstracts to identify relevant studies, resulting in 48 studies (Phase 2). In Step 3, we read the introductions and methodology and conclusions sections. Then we applied the quality criteria carrying 17 studies forward to the subsequent stage. In Step 4 the answers for the the research questions were obtained. Table 1 summarizes the paper selection and review process in numbers.

2.6 Summary of Results

The aim of the synthesis was to group the information extracted from the studies in order to: identify the main techniques, algorithms, validation strategies and metrics related to the research questions. Meta-ethnographic methods were used to synthesize the data extracted from the primary studies (Noblit and Hare, 1988). In the first phase of the synthesis the main concepts of each study were identified using the author's original terms. The key concepts were then tabulated to allow comparison between studies. In the next section we present the results and the respective discussion of the findings obtained from the selected primary studies.

[1]https://www.mendeley.com, Reference Management Software and Researchers Network

[2]http://www.jabref.org, Graphic Application in Java to manage bibtex (.bib) databases.

Table 1: Paper selection and review process in numbers.

Repository	Result (1)	Incl.(2)	(2)/(1)%
ACM Digital Library	183	9	4,92%
IEEE Xplorer	83	4	4,82%
SCOPUS	268	13	4,85%
Manual Inclusion	4	2	50,00%
Total	**399**	**17**	**4,26%**

3 RESULTS AND DISCUSSION

In this section, we present the findings related to research questions 1 and 2. These findings provide evidence that machine learning techniques have been applied in computer science secondary studies in a somewhat successful way. Each of the two nodes of the mental map presented in Figure 1 summarizes findings to answer both of our research questions.

3.1 ML Techniques to Support Secondary Studies in Computer Science (RQ1)

According to Figure 1 illustrates, the literature has registered a number of machine learning techniques to support secondary studies in Computer Science: Bag Of Words (BoW), Decision Tree (DT), Hybrid Feature Selection Method (HFSM), Support Vector Machine (SVM), Term Frequency (Inverse Document Frequency (TF- IDF), Visual Text Mining (VTM) and its variations.

The Bag Of Words (BoW) adopted by studies S1, S3, S4, S9, S11, S12 and S15, technique focuses on the preprocessing of input data to convert them into a vectorial representation based on counting the number of appearance of a set of selected words extracted from the primary studies, as proposed by Salton and colleagues (Salton et al., 1975).

According to S12, Suitable algorithms for study selection are then, e.g. decision trees, logistic regression, and Naive Bayes. Overall, the most popular ones are VTM, TF-IDF, and BoW.

These results are aligned with the results already

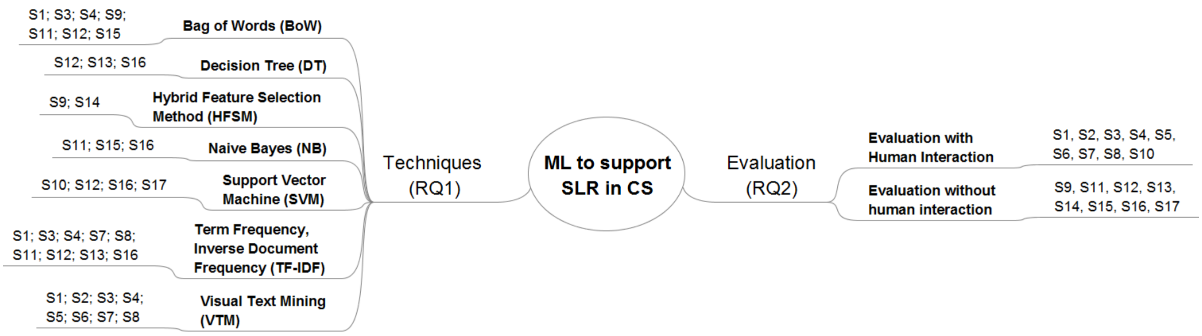

Figure 1: Results of selected studies.

presented in (Hamad and Salim, 2014) regarding ML algorithms (NB, DT, and SVM) and the use of VTM. Even though the past studies covered on Hamad and Salim (2014) have not specifically focused on computer science studies, our review indicates how useful these algorithms are to the research community, whether they are computer science or not.

In addition, a large number of studies using VTM for SLR in computer science (47% of selected primary studies) matches with the findings of (O'Mara-Eves et al., 2015) and (Olorisade et al., 2016).

However, in spite of the similarities previously reported, for automation of SLR steps in computer science, the most used algorithm in the primary studies was TF-IDF, which that was present in 53% of the selected primary studies.

3.2 How ML Techniques Have Been Assessed (RQ2)

According to the data extracted from each of the primary studies, researchers have assessed their techniques by performing exploratory studies carried out using one of the two models as follows:

Evaluation with Human Interaction: In S1, S2, S3, S4, S5, S6, S7, S8, and S10, the authors simulated an SLR with groups of researchers that were divided into researchers following a traditional SLR process (manual SLR) and researchers following an assisted SLR approach (using the technique in question). In the end, the results of these groups were compared. Following this approach, the results were susceptible to the degree of knowledge and experience of the involved researchers.

Evaluation without Human Interaction: In S9, S11, S12, S13, S14, S15, S16 and S17, the authors used SLRs already published to generate a corpus (set of texts extracted from each selected primary study) on which the author of the proposal, without the participation of third parties, applied her/his technique. In the end, the evaluation result was compared with the result of the original secondary study.

The two forms of evaluation proved to be very common in our study dataset. The evaluation model with human participants (the first mode described above) had one more paper compared to the other model. However, in the systematic review carried out by (Olorisade et al., 2016), the use of evaluation without human participation was the most commonly used.

Additionally, among the phases of a systematic review (Kitchenham and Charters, 2007) (SLR planning phase, search string construction, primary studies selection, and data extraction from primary studies), most of the studies, 14 out of 17, focused on support the automation of the primary studies selection phase, whereas the other 3 papers are secondary studies published a few years ago. This concentration of efforts in the selection phase of primary studies coincides with previous findings (Hamad and Salim, 2014) regarding the support of automation. This also reinforces the position of (Olorisade et al., 2016) that identified solid evidence on the effectiveness ML support at that SLR phase.

4 CONCLUSIONS

In this systematic mapping we found evidences that there is an increasing use of ML techniques to support the automation of some SLR activities. We highlight the following works to represent such evidences (Hamad and Salim, 2014), (O'Mara-Eves et al., 2015) and (Olorisade et al., 2016). Over the last four years, they have obtained more than 50 studies that have set out to automate activities of secondary studies. The use of ML algorithms has been shown to be promising based on the results reported, and as presented by (Hamad and Salim, 2014), (O'Mara-Eves et al., 2015) and (Olorisade et al., 2016), automation is increasing in several areas of knowledge other than computer science.

In addition, it becomes apparent the viability of automation techniques applied to the area of computer science. Although still in the experimental stage and with the need for some advancements, the application of ML to support SLRs has been effective enough to be further explored.

The studies we analyzed indicate consistency in their results and the keen interest in the paper selection phase demonstrates how important and costly this step is for researchers.

This systematic mapping supports the development of future work that can either propose new techniques for automating different phases of an SLR or improve the effectiveness of existing approaches.

In the following, we present the list of studies selected in this Systematic Mapping: **S1** (Felizardo et al., 2010), **S2** (Felizardo et al., 2011), **S3** (Felizardo et al., 2012), **S4** (Felizardo et al., 2014), **S5** (Feng et al., 2017), **S6** (Garcés et al., 2017), **S7** (Malheiros et al., 2007), **S8** (Mergel et al., 2015), **S9** (Ouhbi et al., 2016), **S10** (Piroi et al., 2015), **S11** (Rizzo et al., 2017), **S12** (Ros et al., 2017), **S13** (Rúbio et al., 2016), **S14** (Sellak et al., 2015), **S15** (Tomassetti et al., 2011), **S16** (Torres et al., 2013), **S17** (Yu et al., 2018).

REFERENCES

Felizardo, K. R., Andery, G. F., Paulovich, F. V., Minghim, R., and Maldonado, J. C. (2012). (S3) A visual analysis approach to validate the selection review of primary studies in systematic reviews. *Information and Software Technology*, 54(10):1079–1091.

Felizardo, K. R., Nakagawa, E. Y., Feitosa, D., Minghim, R., and Maldonado, J. C. (2010). (S1) An Approach Based on Visual Text Mining to Support Categorization and Classification in the Systematic Sapping. In *Proceedings of the 14th international conference on Evaluation and Assessment in Software Engineering, EASE 2010*, EASE'10, pages 1–10, Swinton, UK, UK. British Computer Society.

Felizardo, K. R., Nakagawa, E. Y., MacDonell, S. G., and Maldonado, J. C. (2014). (S4) A visual analysis approach to update systematic reviews. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*, EASE '14, pages 1–10. ACM.

Felizardo, K. R., Salleh, N., Martins, R. M., Mendes, E., MacDonell, S. G., and Maldonado, J. C. (2011). (S2) Using Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews. In *2011 International Symposium on Empirical Software Engineering and Measurement*, ESEM '11, pages 77–86. IEEE Computer Society.

Feng, L., Chiam, Y. K., Abdullah, E., and Obaidellah, U. H. (2017). (S5) Using Suffix Tree Clustering Method To Support The Planning Phase Of Systematic Literature Review . pp 311 - 332. *Malaysian Journal of Computer Science*, 30(4):311–332.

Garcés, L., Felizardo, K., Oliveira, L., and Nakagawa, E. (2017). (S6) An Experience Report on Update of Systematic Literature Reviews. In *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE*, pages 91–96.

Hamad, Z. and Salim, N. (2014). Systematic literature review (SLR) automation: A systematic literature review. *Journal of Theoretical and Applied Information Technology*, 59(3):661–672.

Jalali, S. and Wohlin, C. (2012). Systematic literature studies: Database searches vs. backward snowballing. In *Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 29–38.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*.

Malheiros, V., Hohn, E., Pinho, R., Mendonca, M., and Maldonado, J. C. (2007). (S7) A Visual Text Mining approach for Systematic Reviews. In *Proceedings of the First International Symposium on Empirical Software Engineering and Measurement*, pages 245–254. IEEE.

Mergel, G. D., Silveira, M. S., and da Silva, T. S. (2015). (S8) A method to support search string building in systematic literature reviews through visual text mining. *Proceedings of the 30th Annual ACM Symposium on Applied Computing - SAC '15*, 13-17-Apri:1594–1601.

Noblit, G. and Hare, R. (1988). *Meta-Ethnography*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America.

Olorisade, B. K., de Quincey, E., Brereton, P., and Andras, P. (2016). A critical analysis of studies that address the use of text mining for citation screening in systematic reviews. In *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering - EASE '16*, volume 01-03-June of *EASE '16*, pages 1–11. ACM.

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., and Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews*, 4(1):5.

Ouhbi, B., Kamoune, M., Frikh, B., Zemmouri, E. M., and Behja, H. (2016). (S9) A hybrid feature selection rule measure and its application to systematic review. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services - iiWAS '16*, pages 106–114. ACM Press.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *Ease*, volume 8, pages 68–77.

Piroi, F., Lipani, A., Lupu, M., and Hanbury, A. (2015). (S10) DASyR(IR) - document analysis system for systematic reviews (in Information Retrieval). In *2015*

13th International Conference on Document Analysis and Recognition (ICDAR), ICDAR '15, pages 591–595. IEEE Computer Society.

Rizzo, G., Tomassetti, F., Vetrò, A., Ardito, L., Torchiano, M., Morisio, M., and Troncy, R. (2017). (S11) Semantic enrichment for recommendation of primary studies in a systematic literature review. *Digital Scholarship in the Humanities*, 32(1):195–208.

Ros, R., Bjarnason, E., and Runeson, P. (2017). (S12) A Machine Learning Approach for Semi-Automated Search and Selection in Literature Studies. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering - EASE'17*, pages 118–127. ACM.

Rúbio, T. R. P. M., Gulo, C. A. S. J. C., Rubio, T. T. R., and Gulo, C. A. S. J. C. (2016). (S13) Enhancing academic literature review through relevance recommendation: Using bibliometric and text-based features for classification. In *Iberian Conference on Information Systems and Technologies, CISTI*, volume 2016-July, pages 1–6.

Salton, G., Wong, A., and Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620.

Sellak, H., Ouhbi, B., and Frikh, B. (2015). (S14) Using rule-based classifiers in systematic reviews. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications &Services - iiWAS '15*, iiWAS '15, pages 1–5. ACM.

Tomassetti, F., Rizzo, G., Vetro, A., Ardito, L., Torchiano, M., and Morisio, M. (2011). (S15) Linked data approach for selection process automation in systematic reviews. *15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011)*, 2011(1):31–35.

Torres, J. A. S., Cruzes, D. S., and Salvador, L. N. (2013). (S16) Automatically locating results to support systematic reviews in software engineering. *CIbSE 2013: 16th Ibero-American Conference on Software Engineering - Memorias del 10th Workshop Latinoamericano Ingenieria de Software Experimental, ESELAW 2013*, (January):6–19.

Yu, Z., Kraft, N. A., and Menzies, T. (2018). (S17) Finding Better Active Learners for Faster Literature Reviews. *Empir Software Eng*.

Zhang, H., Babar, M. A., and Tell, P. (2011). Identifying relevant studies in software engineering. *Information and Software Technology*, 53(6):625–637.

..