January-February 2020
ISSN: 0193-4120 Page No. 16003 - 16009

Systematic Literature Review -Software Effort Estimation Using Machine Learning Techniques

Mandeep Kaur Sandhu, Research Scholar, Computer Science and Engineering, Chandigarh University, Gharuan

Article Info
Volume 82
Page Number: 16003 - 16009
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 28 February 2020*

Abstract:
A measure of accurate effort estimation plays a vital role in the software development process. It directly affects the cost, time andhuman power required for software developments. This paper introduces a systematic literature review of machine learning techniques used for measuring software effort estimation. This paper offering different machine learning techniques that are used single or with combination for measure software effort estimation. The outcomes of the this review has concluded conspicuous fashions of machine learning techniques, performance metrics, datasets, year of publication , etc. used for software effort estimation

Keywords: *Effort Estimation, Software, Machine Learning*

INTRODUCTION

Software effort estimation plays a vital role in software development from the late 1970's. A lot of work was carried out in software estimation. Estimation that is more accurate is still a huge concern in the industries. Overestimation and underestimation affect the industries in a very dangerous manner. The overestimation leads to more men's power, cost and time it can affect the budget of the industries whereas underestimate may lead less number of workers, quality and missing deadlines. A key to the success of a software development project is an accurate estimation. Estimation in the form of effort, time and cost. Various estimation methods are involved in software process development. Researchers have proposed various methods for predicting the estimation more accurately, for what is known as the software development effort estimate (SDEE). All the techniques aim to predict a more accurate estimation, which help in software development. Early for predicting estimation depends on standard formulas, historical data, expert advice, experience methods come under this approach is COCOMO model, function point analysis and software life cycle model or use case point. Regression analysis based on human experience, advice, judgments highlighted by Jorgensen, which covers almost eleven estimation methods.

Machine learning methods used for effort estimation by researchers since 1991. By using these methods human spent more time on other functions of software development rather than estimating the proposed software. This paper focus on a systematic literature review on identifying prevalent tends in machine learning methods for effort estimation. This includes studies from 2000 to 2019. The studies from major database like 1) Web of science 2) IEEE digital library 3) Willey online library 4) ACM based on inclusion/ exclusion and quality assessment criteria

Process used for Systematic literature review (SLR)

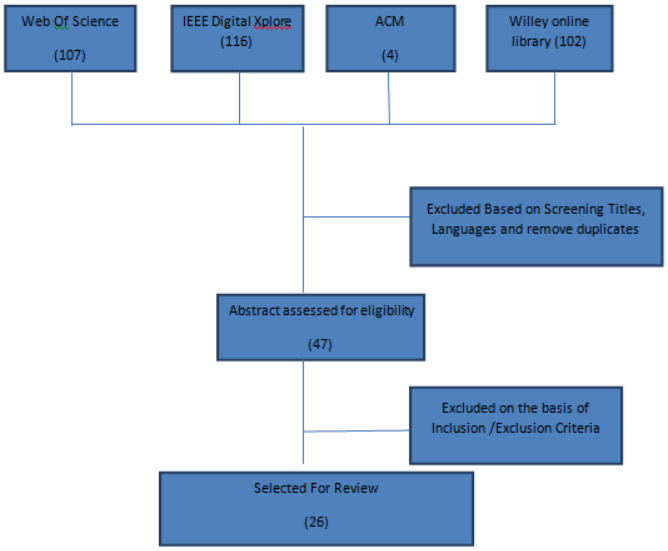
The process used to conduct the SLR study considers the recommendations and guidelines given by B. kitcheham [1]. The review method includes the following steps

- Research query identification
- Data source/ search procedure

January-February 2020
ISSN: 0193-4120 Page No. 16003 - 16009

- Process of selecting the study
- Quality valuation
- Data taking out
- Data synthesis

In this study, we independently performed database searches. We set up some research questions that meet the objectives of SLR after discussion with the team. A team consists of Authors, Experts in the field. After this, we designed a search procedure to select the studies that will help in answering our research questions. A detailed description of the search strategy is given in figure 1.



1. Research query identification

For identify the research question we follow the research strategy. Data extraction and analysis all depend on the research question. The research question for our systematic literature review is as follows:

1. Were selection criteria clearly described
2. Type of MI technique used
3. Were there an additional technique used?
4. Were experiments applied in a sufficient database?
5. Were the performance metric used to measure the accuracy of the proposed method?
6. In which Year Research article is published?

2. Data source/ search procedure

A data source for search procedure includes prominent electronic libraries such as web of knowledge, willey online library, IEEE digital Xplore, ACM Digital Library. The search procedure includes searching for keywords related to our study. We follow the Prisma method for our research.

For each database, enter each key search term individually. Authors combine all terms in different combinations using boolean operators like AND or OR as appropriate. Apply our limits such as year of publication, Language, Techniques and so on.

Table 1. shows no. of articles for SLR.

NO. of articles			
Library	**Total Papers**	**Accepted**	**Rejected**
WOS	107	14	93
IEEE	116	9	107
ACM	4	3	1
Willey	102	1	101
Total	329	26	303

3. Process of selecting the study(inclusion/ exclusion criteria)

An inclusion and Exclusion criterion has been defined to precisely assess the quality of literature available. Author reviewed the papers that are available from the selected database and discussed for selection and rejection decisions.

Inclusion (Paper Included for our study)

- Papers published from the year 2005 to 2019 year with full text.
- Papers in which MI techniques used for computing software effort estimation
- Papers in which performance measuring metric used
- Study apply on valid database

Exclusion(Paper excluded from our SLR)

- Papers did not satisfy the above inclusion criteria

January-February 2020
ISSN: 0193-4120 Page No. 16003 - 16009

4. Quality assessment

No outer factor is incorporated for estimating the quality assessment. We select papers from a reputed database as expert researchers related to quality journals and conferences have audited them. The peer-reviewed papers were viewed as adequate

5. Data taking out

Data taking out depends on our research question that was created by the author by thinking about the accompanying significant properties. The properties consider in extraction are as per the following

- Publication Year.
- Machine Learning Techniques
- Datasets Used
- Estimation Performance

6. Data Synthesis

The data synthesis procedure incorporates gathering the information and finishing with appropriate responses according to research queries. Data Synthesis has been performed by dissecting the literary works through various statistics.

RESULTS OF THE SYSTEMATIC REVIEW

The outcome of the review is addressed in the form of answers to the research queries.

RQ1.Were selection criteria clearly described

RQ3.Were there an additional technique used?

RQ4.Were experiment applied on sufficient database?

RQ5.Were performance metric used to measure the accuracy of proposed method?

Study No.	RQ1	RQ3	RQ4	RQ5
S1	YES		YES	YES
S2	YES	YES	YES	YES
S3	YES	YES	YES	YES
S4	YES		YES	YES
S5	YES	YES	YES	YES
S6	YES		YES	YES
S7	YES	YES	YES	YES
S8	YES		YES	YES
S9	YES	YES	YES	YES
S10	YES	YES	YES	
S11	YES	YES		YES
S12	YES	YES	YES	
S13	YES			
S14	YES	YES		YES
S15	YES	YES	YES	
S16	YES	YES		
S17	YES	YES	YES	YES
S18	YES	YES	YES	YES
S19	YES	YES	YES	YES
S20	YES	YES	YES	YES
S21	YES			
S22	YES	YES	YES	YES
S23	YES	YES		YES
S24	YES	YES	YES	
S25	YES	YES		YES
S26	YES	YES	YES	YES

Table II listtheAnswer of RQ1,RQ3,RQ4 and RQ5

RQ2. Type of MI technique used?

The Table IIIlist the different machine learning approaches and strategies that have been received. The measurable after effects of methodologies demonstrate that ArtificalNeural Network, Fuzzy Logic, Support Vector Regression are most applied technique for software effort estimation. Each Approach has its own pros and cons as per which they are utilized.

Machine Learning Approaches	Statistics of Usage	Study Number
Artificial Neural	32.35 %	S2,S5,S6,S8,S13,S16,S17,S18,S20,S23,S24

January-February 2020
ISSN: 0193-4120 Page No. 16003 - 16009

Network		
Fuzzy Logic	11.76 %	S10,S14,S25,S26
Genetic Algorithm	5.88%	S3,S12
Analogy Based	2.94%	S9
Support Vector Regression	26.47 %	S2,S3,S4,S5,S8,S9,S11,S17,S19
Bayesian Network	8.82%	S15,S21,S23
Regression Tree	8.82%	S1,S5,S22
Case Based Reasoning	2.94%	S7

Table III.list the different machine learning techniques

RQ6.In which year research article published?

The Table IV below gives the quantity of research papers distributed on machine learning based effort estimation over the predetermined period. The appropriation of papers over years demonstrates that reliable efforts were made by researchers to improve the effort estimations utilizing new technologies and techniques. The significant distributing sources are Web of Science and IEEE advanced Xplore.

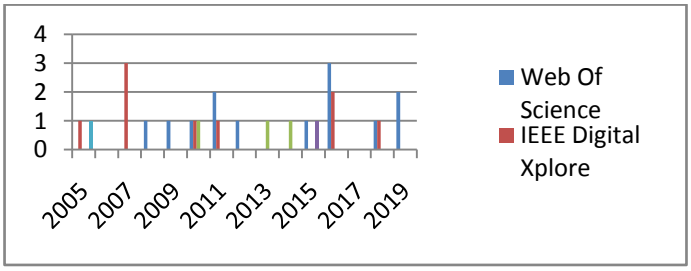


Table IV. Year of Publication

DatasetsUsed

The Table IV shows the list of datasets utilized in the research papers which are conspicuously NASA, COCOMO, ISBSG, TukutukuandDesharnais datasets. These are considered as benchmark and institutionalized datasets. Outcomes additionally uncovered that new datasets are by and large progressively utilized as to give better and generalized estimation.

Datasets	Statistics of Usage	Study No.
ISBSG	11.78%	S1,S8,S10,S24,S26
COCOMO	12.12%	S1,S2,S6,S7,S19,S26
TUKUTUKU	7.68%	S1,S4
DESHANAIS	9.60%	S1,S2,S3,S7,S17
ALBRECHT	4.80%	S1,S2,S7
FINNISH	1.92%	S9,S20
KERMERER	1.92%	S9,S20
MAXWELL	4.80%	S9,S20,S24
NASA	13.44	S2,S3,S7,S12,S15,S17,S18,S19,S22
OTHERS	31.68	S2,S5,S7,S11,S13,S14,S16,S18,S19,S21,S23,S24,S25,S26

Performance Measures

The aftereffects of the effort estimations have been assessed for exactness and blunder measures by contrasting the evaluated outcomes and the actual effort values. The generally used performance measures are Prediction (Pred), Magnitude Mean relative Error (MMRE) and Mean relative Error (MRE), [6]. In any case, numerous presentation measures can be utilized depending on the kind of issue and machine learning approach utilized. They are Magnitude of Error Relative to estimate (MER), Median of Magnitude of Relative Error (MdMRE), Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Magnitude of Error Relative to estimate (MMER), etc.

January-February 2020
ISSN: 0193-4120 Page No. 16003 - 16009

FINDINGS AND CONCLUSIONS

The review paper concludes that a large amount of research was carried out in software effort estimation using machine-learning techniques. From our SLR we conclude that the most commonly machine learning technique used are Artificial Neural Networksi.e 32.35%, Support Vector Regressioni.e 26.47% and Fuzzy Logicsi.e 11.76% software effort estimation and other Ml techniques like RT, CBR,GA and Analogy based are used less than 10%. The SLR further revealed the dataset mostly used are real life datasets and there is no specification way to validate the dataset. A very less work is done by using CBR techniques. In future work you can use a hybrid approach by using CBRwith any other ML or Non Ml technique for measuring the software effort estimate.

References

[1] Braga PL, Oliveira AL, Meira SR. Software effort estimation using machine learning techniques with robust confidence intervals. In Hybrid Intelligent Systems, 2007; 352- 357.

[2] Fenton N, Bieman J. Software metrics: a rigorous and practical approach. CRC press; 2014.

[3] J. Wen, S. Li, and L. Tang, "Improve analogy-based software effort estimation using principal components analysis and correlation weighting," Proc. - Asia-Pacific Softw. Eng. Conf. APSEC, no. 2, pp. 179–186, 2009.

[4] K. Moløkken-Østvold, M. Jørgensen, S. S. Tanilkan, H. Gallis,A.C. Lien, and S. E. Hove, "A survey on software estimation in the norwegian industry," Proc. - Int. Softw. Metrics Symp., pp. 208–219, 2004.

[5] Kitchenham B. Charters. Guidelines for performing systematic literature review in software engineering. Keele, UK, Keele University Version 2.3; 2007.

[6] Mendes E. Improving software effort estimation using an expert- centered approach. In International Conference on Human- Centred Software Engineering Springer, Berlin, 2012; 18- 33.

[7] P. Sharma and J. Singh, "Systematic Literature Review on Software Effort Estimation Using Machine Learning Approaches," *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*, Jammu, 2017, pp. 43-47..

[8] S.D. Conte, H.E. Dunsmore, and V.Y. Shen, "Software Engineering Metrics and Models", Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, 1986

[9] Z. Abdelali, H. Mustapha, and N. Abdelwahed, "Investigating the use of random forest in software effort estimation," in *SECOND INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING IN DATA SCIENCES (ICDS2018)*, 2019, vol. 148, pp. 343–352.

[10] A. L. I. Oliveira, P. L. Braga, R. M. F. Lima, and M. L. Cornelio, "GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation," *Inf. Softw. Technol.*, vol. 52, no. 11, SI, pp. 1155–1166, Nov. 2010.

[11] P. L. Braga, A. L. I. Oliveira, and S. R. L. Meira, "A GA-based Feature Selection and Parameters Optimization for Support Vector Regression Applied to Software Effort Estimation," in *APPLIED COMPUTING 2008, VOLS 1-3*, 2008, p. 1788+.

[12] A. Corazza, S. Di Martino, F. Ferrucci, C. Gravino, and E. Mendes, "Applying Support Vector Regression for Web Effort Estimation using a Cross-Company Dataset," in *ESEM: 2009 3RD INTERNATIONAL SYMPOSIUM ON EMPIRICAL SOFTWARE ENGINEERING AND MEASUREMENT*, 2009, p. 191+.

[13] R. Malhotra and A. Jain, "Software Effort Prediction using Statistical and Machine Learning Methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 1, pp. 145–152, Jan. 2011.

[14] P. Rijwani and S. Jain, "Enhanced Software Effort Estimation using Multi Layered Feed Forward Artificial Neural Network Technique," in *TWELFTH INTERNATIONAL CONFERENCE ON COMMUNICATION NETWORKS, ICCN 2016 / TWELFTH INTERNATIONAL CONFERENCE ON DATA MINING AND WAREHOUSING, ICDMW 2016 / TWELFTH INTERNATIONAL CONFERENCE ON IMAGE*

January-February 2020
ISSN: 0193-4120 Page No. 16003 - 16009

AND SIGNAL PROCESSING, ICISP 2016, 2016, vol. 89, pp. 307–312.

[15] M. Azzeh, “Adjusted Case-Based Software Effort Estimation Using Bees Optimization Algorithm,” in *KNOWLEDGE-BASED AND INTELLIGENT INFORMATION AND ENGINEERING SYSTEMS, PT II: 15TH INTERNATIONAL CONFERENCE, KES 2011*, 2011, vol. 6882, pp. 315–324.

[16] P. Sharma and J. Singh, “Machine Learning Based Effort Estimation Using Standardization,” in *2018 INTERNATIONAL CONFERENCE ON COMPUTING, POWER AND COMMUNICATION TECHNOLOGIES (GUCON)*, 2018, pp. 716–720.

[17] T. R. Benala and R. Bandarupalli, “Least Square Support Vector Machine in Analogy-based Software Development Effort Estimation,” in *2016 INTERNATIONAL CONFERENCE ON RECENT ADVANCES AND INNOVATIONS IN ENGINEERING (ICRAIE)*, 2016.

[18] A. B. Nassif, M. Azzeh, A. Idri, and A. Abran, “Software Development Effort Estimation Using Regression Fuzzy Models,” *Comput. Intell. Neurosci.*, 2019.

[19] J.-S. Chou, M.-Y. Cheng, Y.-W. Wu, and C.-C. Wu, “Forecasting enterprise resource planning software effort using evolutionary support vector machine inference model,” *Int. J. Proj. Manag.*, vol. 30, no. 8, pp. 967–977, Nov. 2012.

[20] R. K. Sachan *et al.*, “Optimizing Basic COCOMO Model using Simplified Genetic Algorithm,” in *TWELFTH INTERNATIONAL CONFERENCE ON COMMUNICATION NETWORKS, ICCN 2016 / TWELFTH INTERNATIONAL CONFERENCE ON DATA MINING AND WAREHOUSING, ICDMW 2016 / TWELFTH INTERNATIONAL CONFERENCE ON IMAGE AND SIGNAL PROCESSING, ICISP 2016*, 2016, vol. 89, pp. 492–498.

[21] C. Lopez-Martin, “Predictive accuracy comparison between neural networks and statistical regression for development effort of software projects,” *Appl. Soft Comput.*, vol. 27, pp. 434–449, Feb. 2015.

[22] A. B. Nassif; L. F. Capretz; D. Ho,"Estimating Software Effort Based on Use Case Points model using sugeno fuzzy inference system,"in 2011 23RD INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS,VOL 2,2011,pp.42-47

[23] A. BaniMustafa,"Predicting Software Effort Estimation Using Machine Learning Techniques," in 2018, 8th International Conference on Computer Science and Information Technology (CSIT), 2018, pp. 249-256.

[24] K. Iwata; T. Nakashima; Y. Anan; N. Ishii,"Effort Estimation for Embedded Software Development Projects by Combining Machine Learning with Classification," in 2016, 4th Intl Conf on Applied Computing and Information Technology/3rd Intl Conf on Computational Science/Intelligence and Applied Informatics/1st Intl Conf on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD)," pp. 265-270.

[25] P. L. Braga, A. L. I. Oliveira, and S. R. L. Meira, “Software effort estimation using machine learning techniques with robust confidence intervals,” in *19TH IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, VOL I, PROCEEDINGS*, 2007, p. 181+.

[26] B. Baskeles, B. Turhan, and A. Bener, “Software effort estimation using machine learning methods,” in *2007 22ND INTERNATIONAL SYMPOSIUM ON COMPUTER AND INFORMATION SCIENCES*, 2007, pp. 208–213.

[27] E. Kocaguneli; A. Tosun; A. Bener,"AI-Based Models for Software Effort Estimation," in 2010, 36th EUROMICRO Conference on Software Engineering and Advanced Applications, 2010, pp. 323-326.

[28] T. R. Benala; R. Bandarupalli,"Least Square Support Vector Machine in Analogy-Based software development effort estimation," in 2016, International Conference on Recent Advances and Innovations in Engineering (ICRAIE),2016, pp.1-6.

[29] P. C. Pendharkar, G. H. Subramanian, and J. A. Rodger,"A probabilistic model for predicting software development effort," in 2005, IEEE Trans. Softw. Eng., vol. 31, no. 7,2005, pp. 615–624.

[30] P. L. Braga, A. L. I. Oliveira, G. H. T. Ribeiro, and S. R. L. Meira,"Bagging predictors for estimation of software project effort,"in 2007, IEEE Int. Conf. Neural Networks - Conf. Proc.,

January-February 2020
ISSN: 0193-4120 Page No. 16003 - 16009

vol. 5,2007, pp. 1595–1600.

[31] J. Shivhare and S. K. Rath, "Software Effort Estimation using Machine Learning Techniques," in *PROCEEDINGS OF THE 7TH INDIA SOFTWARE ENGINEERING CONFERENCE 2014, ISEC `14*, 2014.

[32] L. Song, L. L. Minku, and X. Yao, "Software Effort Interval Prediction via Bayesian Inference and Synthetic Bootstrap Resampling," *ACM Trans. Softw. Eng. Methodol.*, vol. 28, no. 1, Feb. 2019.

[33] A. Mittal, K. Parkash, and H. Mittal, "Software cost estimation using fuzzy logic," ACM SIGSOFT Softw. Eng. Notes, vol. 35, no. 1, p. 1, 2010.

[34] A. Idri, F. azzahra Amazal, and A. Abran, "Accuracy Comparison of Analogy-Based Software Development Effort Estimation Techniques," Int. J. Intell. Syst., vol. 31, no. 2, pp. 128–152, 2016.