Software Effort Estimation using Bio-inspired Rao algorithm

Mid Sem Presentation

K Sai Vivek - 197139
S Sai Vamsi - 197171
K Vasantha Eswari Devi - 197140
Under the supervision of
Dr. Manjubala Bisi

National Institute of Technology Warangal, India
({kadali_851938,sahuka_911924,eswari_811954}@student.nitw.ac.in)

March 2, 2023

Table of Contents

Introduction

- **Software Effort Estimation:**
  Software effort estimation is an essential feature of software engineering for effective planning, controlling and delivering successful software projects.
  The failure in Effort estimation accuracy leads to customer disappointment and poor software development process.
- **Feature Selection:**
  It is a process of isolating the most consistent, non-redundant, and relevant features to use in model construction.
- **Objective:**
  Analyze various feature selection algorithms (Bio and Non Bio Inspired Algorithms).Using Binary Rao Algorithm for selecting best subset of features.

Background

- Effort Estimation techniques like Linear Regression, Support Vector Regression (SVR), Random Forest (for Decision Tree) are used and applied to COCOMO, CHINA, KEMERER, ALBRECHT datasets
- Non-Bio Inspired Feature Selection:
  - Information Gain Feature Selection: This feature selection technique evaluates the gain of each variable in the context of the target variable.
  - Correlation Based Feature Selection: Features with high correlation are more linearly dependent and hence have almost the same effect on the dependent variable
- Bio Inspired Feature Selection [1]:
  - Binary Rao Algorithm [2] : It uses both best and worst solutions in each iteration and random interactions among candidate solutions to quickly find an optimum solution.

Related Work

- According to the chaos report (2015) of The Standish Group International, 60% of IT projects were not on their scheduled time and 56% were not on the budget [3].
- From the Literature we observe that Software Effort Estimation is an important task without which it leads to inaccurate budget and scheduling problems.
- Bio-inspired feature selection algorithms can further improve the accuracy of existing estimation techniques.
- Multi Objective Binary Rao Algorithm can further improve the accuracy by considering more than one estimate parameter like : number of features,error. [4] .

Proposed Approach

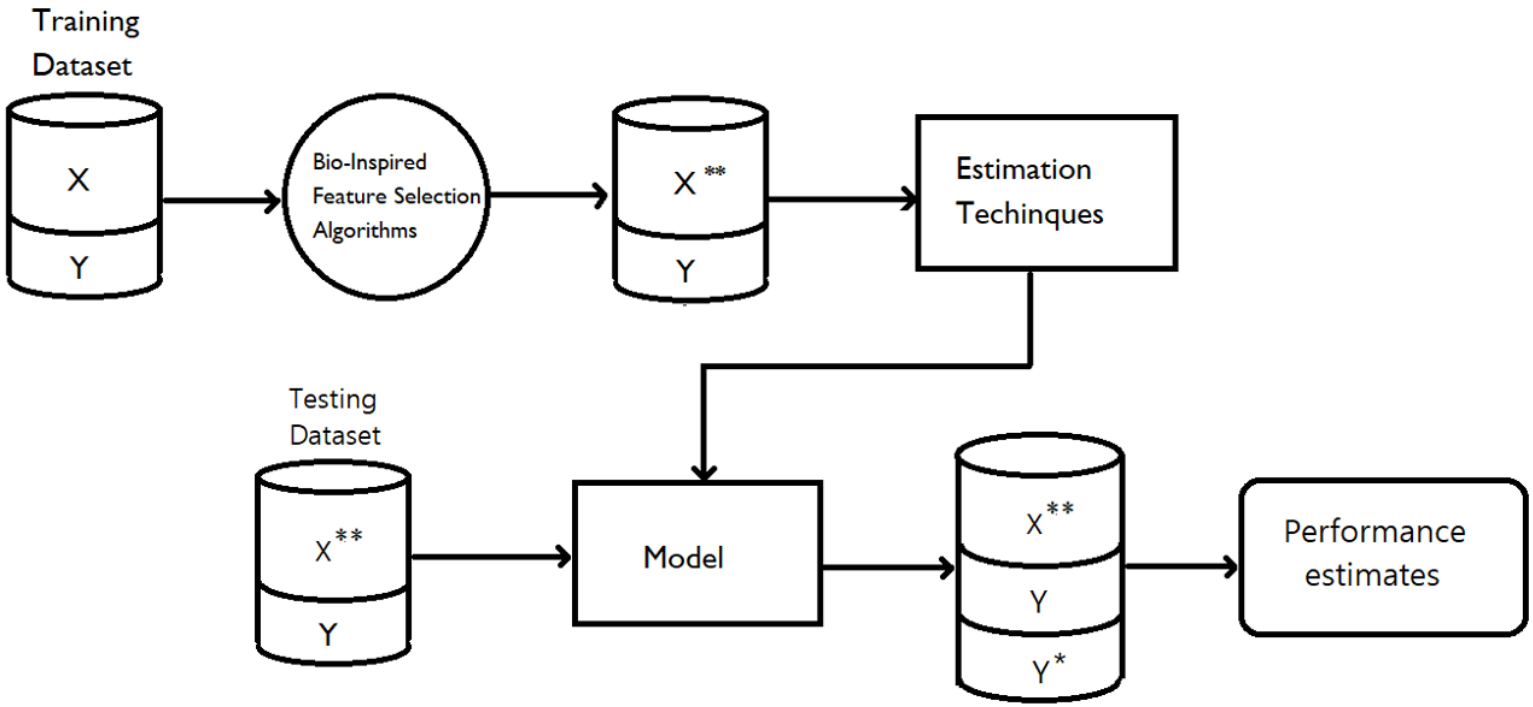

Figure 1: Frame Work for proposed Software Estimation Model

X is representing the original dataset features,X** is representing the reduced features,Y is the dependent variable,Y* is the predicted output

Experimental Results

- COCOMO dataset is used which contains 17 features and 63 instances.Out of these 17 attributes 15 are effort multipliers, one for LOC (Lines of Code) and one for actual development effort.
- These effort multipliers are divided into two types.Some attributes (acap, pcap, tool etc.) are increased to decrease the actual development effort and some attributes (data, time,turn etc.) are decreased to decrease the actual development effort.
- CHINA dataset is used which contains 19 features and 499 instances.Out of these 18 are effort multipliers, one for actual development effort.

- KEMERER dataset is used which contains 8 features and 15 instances.Out of these 7 are effort multipliers, one for actual development effort.
- ALBRECHT dataset is used which contains 9 features and 24 instances.Out of these 8 are effort multipliers, one for actual development effort.
- For each model Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are calculated to predict best model suitable for effort estimation.

Single Objective Binary Rao Algorithm for Feature Selection:

- Randomly initialize the Matrix (size: Number of Samples X Number of features) with binary values.
- Calculate the fitness values of each sample using any estimate technique and find best and worst samples.
- Update the worst sample using the below objective function if fitness of new sample is more than the present.
- Objective Function :

$$S_{new} = S_{worst} + r(S_{best} - S_{worst})$$

Featuer Selection for CHINA		Mean Absolute Error	Root Mean square Error
No Feature Selection	Linear Regression	0.077	0.369
	SVM	0.123	0.633
	Random Forest	0.243	0.328
Information Gain (Non-Bio)	Linear Regression	0.075	0.304
	SVM	0.07	0.615
	Random Forest	0.313	0.368
Correlation (Non-Bio)	Linear Regression	0.084	0.353
	SVM	0.023	0.609
	Random Forest	0.241	0.371
Bio Inspired -Binary Rao	Linear Regression	0.026	0.148
	SVM	0.054	0.594
	Random Forest	0.22	0.31

Table 1: Errors for CHINA dataset

Featuer Selection for COCOMO		Mean Absolute Error	Root Mean square Error
No Feature Selection	Linear Regression	0.317	0.912
	SVM	0.472	0.681
	Random Forest	0.562	0.612
Information Gain (Non-Bio)	Linear Regression	0.478	0.878
	SVM	0.352	0.811
	Random Forest	0.552	0.673
Correlation (Non-Bio)	Linear Regression	0.41	0.849
	SVM	0.451	0.803
	Random Forest	0.568	0.615
Bio Inspired -Binary Rao	Linear Regression	0.366	0.727
	SVM	0.231	0.784
	Random Forest	0.502	0.597

Table 2: Errors for COCOMO dataset

Featuer Selection for albrecht		Mean Absolute Error	Root Mean square Error
No Feature Selection	Linear Regression	0.187	0.648
	SVM	0.303	0.549
	Random Forest	0.415	0.507
Information Gain (Non-Bio)	Linear Regression	0.3	0.866
	SVM	0.152	0.648
	Random Forest	0.423	0.552
Correlation (Non-Bio)	Linear Regression	0.186	0.512
	SVM	0.383	0.487
	Random Forest	0.237	0.37
Bio Inspired -Binary Rao	Linear Regression	0.072	0.459
	SVM	0.157	0.644
	Random Forest	0.328	0.359

Table 3: Errors for ALBRECHT dataset

Featuer Selection for kemerer		Mean Absolute Error	Root Mean square Error
No Feature Selection	Linear Regression	0.158	0.643
	SVM	0.299	0.524
	Random Forest	0.678	0.682
Information Gain (Non-Bio)	Linear Regression	0.17	0.586
	SVM	0.314	0.517
	Random Forest	0.529	0.669
Correlation (Non-Bio)	Linear Regression	0.186	0.512
	SVM	0.383	0.487
	Random Forest	0.64	0.671
Bio Inspired -Binary Rao	Linear Regression	0.097	0.349
	SVM	0.192	0.75
	Random Forest	0.564	0.617

Table 4: Errors for KEMERER dataset

Conclusion and Future Work

Conclusion:
From the above tables,we can conclude that Bio Inspired rao algorithm is optimal algorithm for software development effort estimation when compared with Non Bio Inspired algorithms like Information Gain and Correlation Based feature selection models.

Future Work:

- Multi Objective Binary Rao Algorithm should be implemented.
- In Multi Objective Binary Rao Algorithm :
  - Minimizing number of features.
  - Maximization of Adjusted R Squared.

References I

[1] Asad Ali and Carmine Gravino. “Improving software effort estimation using bio-inspired algorithms to select relevant features: An empirical study”. In: **Science of Computer Programming** 205 (2021), p. 102621.

[2] Karpagalingam Thirumoorthy et al. “A feature selection model for software defect prediction using binary Rao optimization algorithm”. In: **Applied Soft Computing** 131 (2022), p. 109737.

[3] Yasir Mahmood et al. “Improving estimation accuracy prediction of software development effort: A proposed ensemble model”. In: **2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)**. IEEE. 2020, pp. 1–6.

[4] Chao Ni et al. “An empirical study on pareto based multi-objective feature selection for software defect prediction”. In: **Journal of Systems and Software** 152 (2019), pp. 215–238.

Thank You !