Multi-Objective Software Effort Estimation

Federica Sarro
Senior Research Associate
Dept. of Computer Science, CREST
University College London

f.sarro@ucl.ac.uk
Multi-Objective Software Effort Estimation

F. Sarro*, A. Petrozziello**, M. Harman*

*CREST, Department of Computer Science, University College London, UK
** School of Computing, University of Portsmouth, UK
Software Development Effort Estimation

Process of predicting the most realistic amount of effort required to realise a software project

(effort usually quantified in person-hours/person-months)
Would you ever start producing anything without knowing the cost?
Why is it Important?

Project Scheduling /Staffing

Project Bidding
Why is it Difficult?

I need a budget estimate for my project, but I don't have a scope or a design for it yet.

Okay, my estimate is $3,583,729.

You don't know anything about my project. That makes two of us.
Options for Estimation

Experts tend to under-estimate

What is the margin of error?

Predictions of project effort lie within 30%-40% of the true value

30%  40%

K. Molkken and M. Jorgensen. A review of surveys on software effort estimation. ISESE'03.
Options for Estimation

Experts tend to under-estimate within 30%-40% of the true value.

Data Driven Methods

- Regression-based
- Analogy-based
- Search-based
Options for Estimation

After ~30 years of research...

Data Driven Methods

- Linear Regression
- Stepwise Regression
- Manual Stepwise Regression
- Support Vector Regression
- Classification and Regression Trees
- Case-based Reasoning
- K-Nearest Neighbours
- Genetic Algorithms
- Genetic Programming
- Tabu Search
- Simulated Annealing

... still unable to par human-estimates!
Confidence Guided Effort Estimation (CoGEE)

CoGEE is a multi-objective evolutionary approach that seeks to build robust estimation models.
Novelty of Our Approach

All previous evolutionary approaches sought to improve only point estimates
none of them was clearly better than the state-of-the-art
none of them parred human-expert estimates
Empirical Evaluation

CoGEE realised as a Non-dominated Sorting Genetic Algorithm-II (NSGAII)

Compared VS.

3 baselines

3 state-of-the-art effort estimators

3 alternative single/multi-objective formulations

5 industrial datasets from the PROMISE repository (724 projects)
RQ4. Comparison to Industrial Practices

How does our approach, CoGEE, compare to human-expert-based estimates?
RQ4. Comparison to Industrial Practices

Human-expert-based predictions of project effort lie within 30% and 40% of the true value.

The evidence for these thresholds comes from a survey of current industry practices by Molkken and Jørgensen.
RQ4. Comparison to Industrial Practices

The median error of CoGEE lies within both thresholds for all the datasets.
RQ4. Comparison to Industrial Practices

This is not always true for the state-of-the-art approaches.
RQ4. Comparison to Industrial Practices

CoGEE provides human-competitive results! CoGEE outperforms the state-of-the-art techniques!
Empirical Results

Our proposed bi-objective evolutionary algorithm

Creates a new state-of-the-art that pars currently claimed human-expert estimates (RQ4)

Outperforms scientific approaches previous published
(significantly and with medium and large effect size for all the datasets considered)

3 baselines (RQ1)

3 state-of-the-art methods (RQ2)

3 alternative single/multi-objective formulations (RQ3)
Criteria Satisfied by Our Work

✓ (G) The result solves a problem of indisputable difficulty in its field

✓ (E) The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions

✓ (D) The result is publishable in its own right as a new scientific result independent of the fact that the result was mechanically created

✓ (B) The result is better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal

✓ (F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered
Why our entry is the “best”

Human-competitive results to a long-standing and difficult problem

Advances the state of the art

Novelty

Thorough empirical study (724 real-word projects)

Breakthrough results published in ICSE’16 top tier SE conference

Sarro et al. “Multi-Objective Software Effort Estimation”, ICSE’16
Multi-Objective Software Effort Estimation

F. Sarro, A. Petrozziello, M. Harman

Software Effort Estimation (SEE)

Process of predicting the most realistic amount of effort required to realise a software project

(effort usually quantified in person-hours/person-months)

Confidence Guided Effort Estimation (CoGEE)

CoGEE is a multi-objective evolutionary approach that seeks to build robust estimation models

Criteria Satisfied by Our Work

✓ The result solves a problem of indisputable difficulty in its field
✓ The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions
✓ The result is publishable in its own right as a new scientific result independent of the fact that the result was mechanically created
✓ The result is better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal
✓ The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered

Why our entry is the “best”

http://www0.cs.ucl.ac.uk/staff/F.Sarro/projects/CoGEE/