Search Based Requirements Selection and Optimisation

Yuanyuan Zhang
CREST Centre
University College London
Agenda

Background
Problem
Solution
Empirical Study
Conclusion
Requirements Selection & Optimisation

Task

Using prioritisation, visualisation, and optimisation techniques helps decision maker to select the optimal or near optimal subset from all possible requirements to be implemented.

Background

Problem

Solution

Empirical Study

Conclusion

University College London  yuanyuan.zhang@cs.ucl.ac.uk
Requirements Interaction Management

“the set of activities directed towards the discovery, management, and disposition of critical relationships among sets of requirements.”
Requirements Change

| Unpredictable change | Predictable change |

- **Background**
- **Problem**
- **Solution**
- **Empirical Study**
- **Conclusion**

University College London  yuanyuan.zhang@cs.ucl.ac.uk
Problem

Requirements Selection
Problem

Requirements Selection

Background

Problem

Solution

Empirical Study

Conclusion

University College London

yuanyuan.zhang@cs.ucl.ac.uk
<table>
<thead>
<tr>
<th>Background</th>
<th>Problem</th>
<th>Solution</th>
<th>Empirical Study</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

University College London  yuanyuan.zhang@cs.ucl.ac.uk
Goals

Requirements Change

Value

Cost

Background | Problem | Solution | Empirical Study | Conclusion
---|---|---|---|---

University College London  yuanyuan.zhang@cs.ucl.ac.uk
Search based Requirements Optimisation

Use of meta-heuristic algorithms to automate and optimise requirements selection process

- Choose appropriate representation of problem
- Define problem specific fitness function (to evaluate potential solutions)
- Use search based techniques to lead the search towards optimal points in the solution space
Why Search Based Approach?

Robustness
Scalability
Sensitivity analysis
Insight
Feedback & Explanation of results

...
Model

Stakeholder:
\[ C = \{c_1, \ldots, c_j, \ldots, c_m\} \]

Weight:
\[ \text{Weight} = \{w_1, \ldots, w_j, \ldots, w_m\} \]

Requirements:
\[ R = \{r_1, \ldots, r_i, \ldots, r_n\} \]

Cost:
\[ \text{Cost} = \{\text{cost}_1, \ldots, \text{cost}_n\} \]
Model

• Each stakeholder $c_j$ assigns a value to requirements $r_i$:

$$value(r_i, c_j)$$

• Each stakeholder $c_j$ has a subset of requirements that expect to be fulfilled denoted by $R_j$

$$R_j \subseteq R, \quad \forall r \in R_j \quad value(r, c_j) > 0$$

• The overall score of a given requirement $r_i$ can be calculated by:

$$score_i = \sum_{j=1}^{m} w_j \cdot value(r_i, c_j)$$
Data Set Collection & Initialisation

- **Real World Data Sets**
- **27 Combination Random Data Sets**
- **Other Random Data Sets**

Collect -> Generate

- **Format**
- **Initialise**

- **Matlab Files**
  - `.mat`
  - `.dat`

**Requirement**
- Number
- Value
- Cost
- Dependency

**Stakeholder**
- Number
- Weight
- Subset

**Requirement-Stakeholder Matrix**
- Density
- Random Distribution

---

**Background**

**Problem**

<table>
<thead>
<tr>
<th>Solution</th>
<th>Empirical Study</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

University College London  yuanyuan.zhang@cs.ucl.ac.uk
Requirements Selection Process

Search Algorithms
- NSGA-II
- Archive-based NSGA-II
- Two-Achieve
- Pareto GA
- Single Objective GA
- Greedy*
- Random*

Data Sets Regeneration

Multi-Stakeholder Analysis
- Basic Value/Cost
- Today/Future Importance
- Value/Cost Trade-off Analysis
- Fitness-Invariant Dependency
- Fitness-Affecting Dependency
- Interaction Management
- Tensioning Analysis
- Fairness Analysis

Statistic Analysis
- ANOVA Analysis
- Spearman’s Rank Correlation

Iterate? Yes

Change? No

* Strictly speaking, these are not search algorithms.
Result Representation and Visualisation

- Requirements Subsets for Release Planning
- Insight Characteristic of Data Sets
- Performance of the Algorithms

Results

- 2D and 3D Pareto Fronts
- Kiviat Diagrams
- Marked Line Charts
- Convergence
- Diversity

communicate & feedback

represent & visualise

Result Representation and Visualisation

- Requirements Subsets for Release Planning
- Insight Characteristic of Data Sets
- Performance of the Algorithms

Results

- 2D and 3D Pareto Fronts
- Kiviat Diagrams
- Marked Line Charts
- Convergence
- Diversity

communicate & feedback

represent & visualise

Background | Problem | Solution | Empirical Study | Conclusion
--- | --- | --- | --- | ---
University College London | yuanyuan.zhang@kcl.ac.uk
Visualisation

Pareto Optimal Front

Solution Space
Visualisation

Pareto Optimal Front

Objective 1

Objective 2

f1

f2

S1

S2

S3

S4

S5

Pareto-Front

University College London  yuanyuan.zhang@cs.ucl.ac.uk
1. Basic Value/Cost Trade-off

The problem is to select a set of requirements that maximise customers’ satisfaction (total value) and minimise required cost.

The model of fitness functions represented as:

Maximise

\[ f_1(x) = \sum_{i=1}^{n} \text{score}_i \cdot x_i \]

Minimise

\[ f_2(x) = \sum_{i=1}^{n} \text{cost}_i \cdot x_i \]
1. Basic Value/Cost Trade-off

Scale Problem

• consider three typical ‘scales’ cases of problem, with the number of customers ranging from 15 to 100 and the number of requirements ranging from 40 to 140.

• Investigate the relative performance of the approaches for cases.
1. Basic Value/Cost Trade-off

Synthetic data set: 15 stakeholders; 40 requirements
1. Basic Value/Cost Trade-off

<table>
<thead>
<tr>
<th>Stakeholders</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>100</td>
<td>140</td>
</tr>
</tbody>
</table>

**Empirical Study**

University College London  yuanyuan.zhang@cs.ucl.ac.uk
1. Basic Value/Cost Trade-off

Motorola Data set

4 Stakeholders

35 Requirements

<table>
<thead>
<tr>
<th>Background</th>
<th>Problem</th>
<th>Solution</th>
<th>Empirical Study</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>University College London</td>
<td><a href="mailto:yuanyuan.zhang@cs.ucl.ac.uk">yuanyuan.zhang@cs.ucl.ac.uk</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results comparison

Synthetic

Motorola

<table>
<thead>
<tr>
<th>Background</th>
<th>Problem</th>
<th>Solution</th>
<th>Empirical Study</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

University College London  yuanyuan.zhang@cs.ucl.ac.uk
Search vs. Greedy

Synthetic

Motorola

Background     Problem     Solution     Empirical Study     Future Work

University College London  yuanyuan.zhang@cs.ucl.ac.uk
2. Today/Future Importance Analysis

To provide robust solutions not only in the context of present conditions but also in response to those future changes that can be anticipated.

Maximise

\[ f_1(x) = \sum_{i=1}^{n} \text{score}_{i,\text{today}} \cdot x_i \]

Maximise

\[ f_1(x) = \sum_{i=1}^{n} \text{score}_{i,\text{future}} \cdot x_i \]

Minimise

\[ f_3(x) = \sum_{i=1}^{n} \text{cost}_i \cdot x_i \]

Background | Problem | Solution | Empirical Study | Conclusion
---|---|---|---|---

University College London | yuanyuan.zhang@cs.ucl.ac.uk
2. Today/Future Importance Analysis

Results from Ericsson Data Sets: 124 Requirements, 14 Customers

<table>
<thead>
<tr>
<th>Background</th>
<th>Problem</th>
<th>Solution</th>
<th>Empirical Study</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>University College London</td>
<td><a href="mailto:yuanyuan.zhang@cs.ucl.ac.uk">yuanyuan.zhang@cs.ucl.ac.uk</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Today/Future Importance Analysis

Projection onto the X-Y Plane

Background | Problem | Solution  | Empirical Study | Conclusion
---|---|---|---|---

University College London  yuanyuan.zhang@cs.ucl.ac.uk
2. Today/Future Importance Analysis

Spearman’s Rank Correlation Coefficient

This indicates a positive correlation between the value for today and value for the future.
3. Multi-Stakeholder Tensioning Analysis

The problem is to select a set of requirements that maximise the total value to each stakeholder, which is expressed as a percentage.

The model of fitness functions represented as:

Maximise

\[
\sum_{i=1}^{n} \frac{\text{value}(r_i, c_j) \cdot x_i}{\sum_{r \in R_j} \text{value}(r, c_j)}
\]

subject to

\[
\sum_{i=1}^{n} \text{cost}_i \leq B, \quad B > 0
\]
Data Sets Used

1. Motorola Data Set:
   35 Requirements and 4 Stakeholders

2. Greer and Ruhe Data Set:
   20 Requirements and 5 Stakeholders
Data Sets Used

3. 27 Combination Levels of Random Data Sets:

- the No. of requirements
- the No. of stakeholders
- the density of the stakeholder-requirement matrix

<table>
<thead>
<tr>
<th></th>
<th>$R_{\text{small}}$</th>
<th>$R_{\text{medium}}$</th>
<th>$R_{\text{large}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{\text{small}}$</td>
<td>$C_s R_s D_{\text{low}}$</td>
<td>$C_s R_s D_{m}$</td>
<td>$C_s R_s D_{h}$</td>
</tr>
<tr>
<td></td>
<td>$C_s R_s D_{m}$</td>
<td>$C_s R_s D_{m}$</td>
<td>$C_s R_s D_{h}$</td>
</tr>
<tr>
<td></td>
<td>$C_s R_s D_{h}$</td>
<td>$C_s R_s D_{h}$</td>
<td>$C_s R_s D_{h}$</td>
</tr>
<tr>
<td>$C_{\text{medium}}$</td>
<td>$C_m R_s D_{\text{low}}$</td>
<td>$C_m R_s D_{m}$</td>
<td>$C_m R_s D_{h}$</td>
</tr>
<tr>
<td></td>
<td>$C_m R_s D_{m}$</td>
<td>$C_m R_s D_{m}$</td>
<td>$C_m R_s D_{h}$</td>
</tr>
<tr>
<td></td>
<td>$C_m R_s D_{h}$</td>
<td>$C_m R_s D_{h}$</td>
<td>$C_m R_s D_{h}$</td>
</tr>
<tr>
<td>$C_{\text{large}}$</td>
<td>$C_l R_s D_{\text{low}}$</td>
<td>$C_l R_s D_{m}$</td>
<td>$C_l R_s D_{h}$</td>
</tr>
<tr>
<td></td>
<td>$C_l R_s D_{m}$</td>
<td>$C_l R_s D_{m}$</td>
<td>$C_l R_s D_{h}$</td>
</tr>
<tr>
<td></td>
<td>$C_l R_s D_{h}$</td>
<td>$C_l R_s D_{h}$</td>
<td>$C_l R_s D_{h}$</td>
</tr>
</tbody>
</table>
Multi-Stakeholder Tensioning Analysis

Motorola data set

30% Budgetary Resource Constraint 70%

<table>
<thead>
<tr>
<th>Background</th>
<th>Problem</th>
<th>Solution</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>University College London</td>
<td><a href="mailto:yuanyuan.zhang@cs.ucl.ac.uk">yuanyuan.zhang@cs.ucl.ac.uk</a></td>
<td>Empirical Study</td>
<td></td>
</tr>
</tbody>
</table>
Solutions on the Pareto Front
Average Solutions

![Average Solutions Diagram](image)
Tensions between the Stakeholders’ Satisfaction for Different Budgetary Resource Constraints
Multi-Stakeholder Tensioning Analysis

Greer and Ruhe data set

30% Budgetary Resource Constraint

70%

Background

Problem

Solution

Empirical Study

Conclusion

University College London yuanyuan.zhang@cs.ucl.ac.uk
Algorithms’ Performance

\[ C = \sum_{i=1}^{N} \frac{d_i}{N} \]

\[ P = \frac{num}{NUM} \]

---

**Rank Order for Convergence**

<table>
<thead>
<tr>
<th>Method</th>
<th>Winner</th>
<th>Runner Up</th>
<th>Loser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Search</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Two-Archive</td>
<td>95.19%</td>
<td>4.81%</td>
<td>0%</td>
</tr>
<tr>
<td>NSGA-II</td>
<td>7.04%</td>
<td>92.96%</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Solutions on the Reference Front**

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Search</td>
<td>2.68%</td>
</tr>
<tr>
<td>Two-Archive</td>
<td>94.57%</td>
</tr>
<tr>
<td>NSGA-II</td>
<td>38.25%</td>
</tr>
</tbody>
</table>
Algorithms’ Performance

1. The diversity of the Two-archive algorithm is significant in most cases

2. The Two-archive and NSGA-II algorithms always have a better convergence than the Random Search

3. The Two-Archive algorithm outperforms NSGA-II and Random Search in terms of convergence in some case
4. Multi-Stakeholder Fairness Analysis

Fairness on Absolute *number* of fulfilled requirements:

Maximise \( \overline{NA} \)

Minimise \( \sigma(NA) \)

Fairness on absolute *value* of fulfilled requirements:

Maximise \( \overline{VA} \)

Minimise \( \sigma(VA) \)

where \( VA_j = \sum_{i=1}^{n} \text{value}(r_i,c_j) \cdot x_i \)
4. Multi-Stakeholder Fairness Analysis

Fairness on the percentage of value and cost of fulfilled requirements:

- Minimise $\sigma(Cost - C)$
- Maximise $\overline{VP}$
- Minimise $\sigma(VP)$
- Minimise $\sum_{i=1}^{n} COST_i \cdot x_i$
4. Multi-Stakeholder Fairness Analysis

Background

Problem

Solution

Empirical Study

Future Work

University College London  yuanyuan.zhang@cs.ucl.ac.uk
4. Multi-Stakeholder Fairness Analysis

Motorola Data Set

Fairness on
Percentage of
Fulfilled Value

<table>
<thead>
<tr>
<th>Overall Cost</th>
<th>Std of Percentage of &quot;Fulfilled Value&quot;</th>
<th>Average Percentage of &quot;Fulfilled Value&quot;</th>
</tr>
</thead>
</table>

[Graph showing data distribution]

University College London  
yuanyuan.zhang@cs.ucl.ac.uk
5. Requirements Interaction Management (RIM)

- r1
- r2
- r3
- r4
- r5

Requirements

And
Or
Precedence
Value-related
Cost-related
And

Given requirement $r_i$ is selected, then requirement $r_j$ has to be chosen.

Define an equivalence relation $\xi$ on the requirements array $R$ such that $r(i, j) \in \xi$
Requirements $r_i$ and $r_j$ are conflicting to each other, only one of $r_i, r_j$ can be selected.

Define an equivalence relation $\varphi$ on the requirements array $R$ such that $r(i, j) \in \varphi$.
Precedence

Given requirement $r_i$ has to be implemented before requirement $r_j$

Define an partial order $\mathcal{X}$ on the requirements array $R$ such that $r(i, j) \in \mathcal{X}$
Cost-related

Given requirement $r_i$ is selected, then this selection affects the cost of implementing requirement $r_j$.

Define an partial order $\omega$ on the requirements array $R$ such that $r(i, j) \in \omega$.
Value-related

Given requirement $r_i$ is selected, then this selection affects the value of requirement $r_j$ for the stakeholder.

Define an partial order $\psi$ on the requirements array $R$ such that $r(i, j) \in \psi$
Empirical Study 5: RIM

And, Or and Precedence Dependencies

- Cost

Value

-300 -250 -200 -150 -100 -50 0

5 10 15 20 25 30 35 40 45

34 Customers, 50 Requirements
Empirical Study 5: RIM

And, Or and Precedence Dependencies

- Cost vs. Value

- Customers, 258 Requirements

4 Customers, 258 Requirements

<table>
<thead>
<tr>
<th>Background</th>
<th>Problem</th>
<th>Solution</th>
<th>Empirical Study</th>
<th>Conclusion</th>
</tr>
</thead>
</table>

University College London  yuanyuan.zhang@cs.ucl.ac.uk
Empirical Study 5: RIM

And, Or and Precedence Dependencies

- Cost
- Value
- Archive based NSGA-II with dependencies
- NSGA-II with dependencies
- NSGA-II without dependencies

21 Customers, 412 Requirements
Conclusion

- Basic Value/Cost Trade-off analysis
- Today/Future Importance Analysis
- Multi-Stakeholder tension and fairness analysis
- Requirements Interaction Management

http://crestweb.cs.ucl.ac.uk/resources/sbse_repository/

University College London  yuanyuan.zhang@cs.ucl.ac.uk