SECOND-GUESSING IN TRACING TASKS CONSIDERED HARMFUL?

20 MARCH 2018

BHUSHAN CHITRE, JANE HUFFMAN HAYES UNIVERSITY OF KENTUCKY
ALEXANDER DEKHTYAR, CALIFORNIA POLYTECHNIC STATE UNIVERSITY
OUTLINE

BACKGROUND AND MOTIVATION
PROPOSED APPROACH
PRELIMINARY STUDY
CONCLUSIONS
FUTURE WORK
**BACKGROUND**

Requirements tracing – “ability to describe and follow life of requirement in both forward and backward directions”*

**Trace matrix** - collection of trace links, “specified association between pair of artifacts, one comprising source and one comprising target.”+

**Tracing between artifacts:**
- Requirements to design
- Test cases to requirements
- Code to requirements

---


PROBLEM

- Automated methods/tools for candidate trace matrix (TM)
  - Information retrieval based and other techniques
  - Not 100% accurate
  - Often retrieve unrelated items (false links)

SOLUTION

- Candidate TM verified by *human analysts*

*But certain analyst behaviors ---＞ decreased accuracy*
PROBLEM

SOLUTION

• Automated methods/tools for candidate trace matrix (TM)
  • Information retrieval based and other techniques
  • Not 100% accurate
  • Often retrieve unrelated items (false links)

Candidate TM verified by human analysts

But certain analyst behaviors --- decreased accuracy

---

r12.txt

Change Task Begin/End Times automatically with dependency changes. The start or end date should be changed automatically if links among tasks are changed

---

d13_4.txt

Every time Start/End time for a task changes, the start/end time of one or more its subtasks need to change as well.
<table>
<thead>
<tr>
<th>ID</th>
<th>Status</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>r12d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r13d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r14d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r15d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r15d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r16d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r17d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r2d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r3d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r4d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r5d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r6d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r7d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r8d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r9d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r10d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r11d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
<tr>
<td>r12d</td>
<td>Not Satisfied</td>
<td>View</td>
</tr>
</tbody>
</table>

### Requirement Test

**r12.txt**

Change Task Begin/End Times automatically with dependency changes. The start or end date should be changed automatically if links among tasks are changed.

---

**d13_4.txt**

Every time Start/End time for a task changes, the start/end time of one or more its subtasks need to change as well.
PROBLEM

- Automated methods/tools for candidate trace matrix (TM)
  - Information retrieval based and other techniques
  - Not 100% accurate
  - Often retrieve unrelated items (false links)

SOLUTION

- Candidate TM verified by human analysts

But certain analyst behaviors ---> decreased accuracy
MOTIVATION

Prior work [1, 2] shows these lead to errors of judgement

- Long time to decide
- Revisiting a link (backtracking)

Could be tied to human decision making systems – System 1 (S1) – fast, instinctive thinking and System 2 (S2) – slow, deliberate, logical thinking – above behaviors belong to S2


PROPOSED APPROACH/RESEARCH QUESTIONS

RQ1: Analyst behaviors that reliably lead to making errors, and where fall on Kahneman’s thinking system dichotomy (S1, S2)? (Phase 1 – discover)

RQ2: What enhancements for automated tracing tools can be designed to curb unwanted behaviors? (Phase 2 – enhance)

RQ3: Improvement in accuracy of final TM constructed by analysts using enhanced software? (Phase 3 – evaluate)
DISCOVERY OF ANALYST BEHAVIORS

- Replicate experiment of Kong et al. (RETRO-LOGGING) – more data
- Classify data per Kahneman dichotomy
- Is TM analysis performed best within System 1 decision-making?
DEVELOPMENT OF SOFTWARE ENHANCEMENTS

• For each behavior discovered, design feature(s) to enhance RETRO.NET
  • Warnings
  • Prohibitions
  • Restructuring
STUDY OF THE IMPACT

- Second replication of Kong et al. but use experimental and control groups
  - Do software enhancements actually curb behaviors?
  - Is decrease in unwanted behaviors accompanied by decrease in number of errors analysts make?
PRELIMINARY STUDY

Unwanted behavior/Software enhancements
- Long time to decide analyst more than average time on link decision, prompt with warning
- Backtracking analyst re-visit previous link decision then prompt with warning

Fourteen subjects in two groups
- RETRO.NET control (non-enhanced) – five participants finished
- RETRO.NET experimental (enhanced) – nine participants finished

“Changestyle” – 32 reqts to 17 tests
**RESULTS**

Measured precision, recall, f2 - measure, lag of final TM and time it took to complete task (minutes) – experimental better on most measures *not* time

<table>
<thead>
<tr>
<th>Group</th>
<th>Aggregation</th>
<th>Prec.</th>
<th>Recall</th>
<th>F2</th>
<th>Lag</th>
<th>Time</th>
<th>Delta (TP)</th>
<th>Delta (FP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETRO</td>
<td>actual</td>
<td>0.063</td>
<td>1</td>
<td>0.251</td>
<td>1.1</td>
<td>NA</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Control</td>
<td>Mean</td>
<td>0.083</td>
<td>0.776</td>
<td>0.262</td>
<td>2.552</td>
<td>75</td>
<td>1.6</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.068</td>
<td>0.971</td>
<td>0.254</td>
<td>1.96</td>
<td>60</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Experimental</td>
<td>Mean</td>
<td>0.156</td>
<td>0.961</td>
<td>0.329</td>
<td>1.85</td>
<td>82</td>
<td>1.222</td>
<td>118.7</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.069</td>
<td>0.971</td>
<td>0.283</td>
<td>1.765</td>
<td>86</td>
<td>1</td>
<td>59.5</td>
</tr>
</tbody>
</table>
DISCUSSION/CONCLUSIONS

• Basic prompts might avert analysts from undesired behaviors – at expense of time

• Identified items for future study:
  • Collect number of times prompts appear
  • Collect amount of time analyst takes when dismissing, reacting to prompt
  • Track action taken by analyst after prompt
  • Track number of false positives (etc.) added and removed
  • Potentially track each individual true positive link displayed by RETRO.NET to learn its final disposition
FUTURE WORK

- Phase 1: Discover analyst behavior
- Phase 2: Enhance software to curtail/validate curtailment of unwanted behavior
- Phase 3
  - Undertake wider scope similar study
  - Collect richer data from larger groups
  - Undertake statistical analysis
ACKNOWLEDGMENT

- We thank participants from software engineering classes who participated in study
- We thank NASA and NSF as prior grants funded the development of RETRO.NET
- We thank Jody Larsen, the developer of RETRO.NET
- We thank NSF for partially funding this work under grants CCF-1511117 and CNS-1642134
REFERENCES

THANK YOU!

QUESTIONS?
Analysis and Tracing Process

Credit: Jody Larsen, “High Performance automated traceability.”
INTRODUCTION:

- SAFETY CRITICAL SOFTWARE SYSTEMS – IMPORTANCE OF REQUIREMENTS
  - HIGH-LEVEL DOCUMENT
  - LOW-LEVEL DOCUMENTS
- AUTOMATED METHODS GENERATE CANDIDATE TMS USING INFORMATION RETRIEVAL METHODS
DEPENDENT AND INDEPENDENT VARIABLES

- The independent variables: different version of RETRO.NET “control” and “experimental.”
- The dependent variables: precision, recall, f2-measure, lag and time to perform the experiment.
- Controlled variable: Answer set RTM of “ChangeStyle” dataset and “Retro.NET” tool.
IR MEASURES DEFINITIONS

Precision = \frac{\text{# of Correct Links Returned}}{\text{# of Returned Links}}

Recall = \frac{\text{# of Correct Links Returned}}{\text{Total # of Correct Links}}

f-measure: is the harmonic mean of recall

The $f_2$-measure, i.e., $f$-measure for $a = 2$.

Lag: Lag is a measure of the separation between true and false links. For a requirement $q$, $(q, d)$ for true link. lag$(q, d)$, the lag of an individual link $(q, d)$, is the number of false links that have higher relevance scores than $(q, d)$.

\[ lag = \frac{\sum_{(a,d) \in \tau} \text{lag}(q, d)}{|	au|} \]
HOW TRACING WORKS?

Tracing Task

- Initial Requirement Documents
- Retro.NET Tracing Software
- Candidate TM
- Human Analyst performing Vetting on Candidate TM
- Final TM
THREATS TO VALIDITY

Internal validity:
- Tracing tool
- Human error,
- Hypothesis guessing,
- Personal bias in constructing of the answer set

Construct validity: There were minimal threats to construct validity as standard IR measures (precision, recall, f2 and etc.)

External validity: Experimental dataset

Conclusion validity: statistical analysis

Reliability validity: The study process is defined and easily repeatable.
• We recruited Upper division software engineering computer science students.

• They signed the Informed consent and filled pre-study survey as a form of agreement to participate in our study.

• Held demo/training session to let users get familiar with tool and tracing process.

• Then they worked with testing dataset called “Moonlander” on their own time out the class with provided instructions.
RESULTS AND ANALYSIS

Total of 14 subjects participated in a preliminary study conducted in Spring 2017 at University of Kentucky.

We collected:
- Pre- and post-study survey
- Time logs (time to perform tracing)
- Final TM results (XML)

Out of 14 results
- 5 analysts were in control group (worked on non-enhanced RETRO.NET)
- 9 analysts were in experimental group (worked on enhanced RETRO.NET)
PROPOSED APPROACH/RESEARCH QUESTIONS

We propose three-step experimental study to:

1) Determine if there really are behaviors that lead to errors of judgement for analysts
2) Enhance the requirements tracing software to curtail such behaviors, and
3) Determine if curtailing such behaviors results in increased accuracy
THE STUDY

• Both groups used “changestyle” dataset - 32 requirements traced to 17 system tests

• Collected:
  • Pre- and post-study survey
  • Time logs (time to perform tracing)
  • Final TM results (XML)
SECOND-GUESSING IN TRACING TASKS CONSIDERED HARMFUL?

BHUSHAN CHITRE, JANE HUFFMAN HAYES UNIVERSITY OF KENTUCKY

ALEXANDER DEKHTYAR, CALIFORNIA POLYTECHNIC STATE UNIVERSITY
SOFTWARE ENGINEERING ARTIFACTS

- Guide and inform development
- Support verification and validation
- Relate to each other
SOFTWARE ENGINEERING ARTIFACTS

- Guide and inform development
- Support verification and validation
- Relate to each other
SOFTWARE ENGINEERING ARTIFACTS

- Guide and inform development
- Support verification and validation
- Relate to each other
Req 1: When roll hold mode becomes the active mode, the roll hold reference shall be set to the actual roll attitude of the aircraft, except under the following conditions:

The roll hold reference shall be set to zero if the actual roll angle is less than 6 degrees in either direction, at the time of roll hold engagement.

The roll hold reference shall be set to 30 degrees in the same direction as the actual roll angle if the actual roll angle is greater than 30 degrees at the time of roll hold engagement.

The roll reference shall be set to the cockpit turn knob command, up to a 30 degree limit, if the turn knob is commanding 3 degrees or more in either direction.

- Support verification and validation
- Relate to each other
SOFTWARE ENGINEERING ARTIFACTS

- Guide and inform development
- Support verification and validation
- Relate to each other
SOFTWARE ENGINEERING ARTIFACTS

- Guide and inform development
- Support verification and validation
- Relate to each other
SOFTWARE ENGINEERING ARTIFACTS

• Guide and inform development
• Support verification and validation
• Relate to each other

Requirements-based Testing for Controller Development

This model is used to show how to perform requirements-based testing using test harnesses. Test Sequence Blocks, and the test manager. To view the items, visit www.RequirementsTestingAutopilot.pdf (MIT License).
SOFTWARE ENGINEERING ARTIFACTS

- Guide and inform development
- Support verification and validation
- Relate to each other
TRACE MATRIX

• Tracing to identify relationships

• Trace matrix supports
  • Change impact
  • Regression testing
  • Criticality assessment+
<table>
<thead>
<tr>
<th>Requirement</th>
<th>Functional design</th>
<th>Internal design</th>
<th>Code</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurant has two ordering stations</td>
<td>Mgmt screen #2</td>
<td>Page 45</td>
<td>Line 12485</td>
<td>34, 57, 63</td>
</tr>
<tr>
<td>A waiter may order from any station</td>
<td>Order screen</td>
<td>Page 19</td>
<td>Line 6215</td>
<td>12, 14, 34, 57, 92</td>
</tr>
<tr>
<td>Any customer at a table may request a separate check</td>
<td>Order screen</td>
<td>Page 39</td>
<td>Line 2391</td>
<td>113, 85</td>
</tr>
<tr>
<td>A customer may get checks from more than one station</td>
<td>Check printing</td>
<td>Page 138</td>
<td>Lines 49234, 61423</td>
<td>74, 104</td>
</tr>
</tbody>
</table>

- Tracing to identify relationships
- Trace matrix supports
  - Change impact
  - Regression testing
  - Criticality assessment
TRACE MATRIX

• Tracing to identify relationships
• Trace matrix supports
  • Change impact
  • Regression testing
  • Criticality assessment+
TRACE MATRIX

- Tracing to identify relationships
- Trace matrix supports:
  - Change impact
  - Regression testing
  - Criticality assessment

**Relationship Matrix**

<table>
<thead>
<tr>
<th>Source Requirements Model</th>
<th>Target Requirements Model</th>
<th>Type</th>
<th>Link Type</th>
<th>Realization</th>
<th>Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirement</td>
<td>UseCase</td>
<td>Both</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes**

The system must provide for secure access and user verification on login and password. The user is to be provided by system. The user may change their password according to a set of defined rules.
TRACE MATRIX

- Tracing to identify relationships
- Trace matrix supports
  - Change impact
  - Regression testing
  - Criticality assessment+
TRACE MATRIX

- Tracing to identify relationships
- Trace matrix supports
  - Change impact
  - Regression testing
  - Criticality assessment

**r12.txt**

Change Task Begin/End Times automatically with dependency changes. The start or end date should be changed automatically if links among tasks are changed.

**d13_4.txt**

Every time Start/End time for a task changes, the start/end time of one or more its subtasks need to change as well.
IR FOR TRACING

Design Document

representation

Requirements Document
IR FOR TRACING

Design Document

representation

Matching algorithm

1. Analyst
2. Requirements Document
3. Analyst
Design Document

1. Yes
2. Yes
3. No

Feedback

Final Traceability Matrix

Analyst