



Tim Menzies <menzies.tim@gmail.com>

---

## Decision on your manuscript #AUSE117

---

Robert Hall <Bob.ASEJ@gmail.com>

Tue, Apr 28, 2009 at 2:19 PM

To: tim@menzies.us

Cc: Bob.ASEJ@gmail.com

Dear Mr. Tim Menzies:

I have received review reports on your manuscript, "Real-time Optimization of Requirements Models", which you submitted to Automated Software Engineering.

Based on the advice received, your manuscript could be reconsidered for publication should you be prepared to incorporate major revisions. When preparing your revised manuscript, you are asked to carefully consider the reviewer comments which are attached, and submit a list of responses to the comments. Your list of responses should be uploaded as a file in addition to your revised manuscript.

PLEASE NOTE: YOUR REVISED VERSION CANNOT BE SUBMITTED IN .PS OR .PDF. IN THE EVENT THAT YOUR REVISED VERSION IS ACCEPTED, YOUR PAPER CAN BE SENT TO PRODUCTION WITHOUT DELAY ONLY IF WE HAVE THE SOURCE FILES ON HAND. Submissions without source files will be returned prior to final acceptance.

In order to submit your revised manuscript electronically, please access the following web site:

<http://ause.edmgr.com/>

Your login is: timmenzies

Your password is: menzies323

Please click "Author Login" to submit your revision.

I look forward to receiving your revised manuscript.

Sincerely,

Robert J. Hall, PhD, FASE  
Editor-in-Chief  
Automated Software Engineering

### COMMENTS FOR THE AUTHOR:

Reviewer #1: The paper concerns real time adaption of ultra lightweight requirements models (within 5 second tolerance). The paper uses search based optimization methods associated with the rapidly developing field of SBSE to optimize models of requirements. The work is written in a compelling and direct style (which occasionally is too direct and a little grating) and the contributions are clearly motivated and explained and backed by some evidence in the form of empirical work. I feel that this is a paper that should definitely be worthy of publication in JASE. However, I do have some suggestions for improvement and also I noticed some missing work and weaknesses in presentation that I think should be included. I think that a moderate revision will easily be sufficient to satisfy these change requests.

The revisions concern several broad aspects which need attention

1. Claims and search based techniques (these need to be reconsidered in the light of the No Free Lunch Theorem)
2. Related work. Some very closely related work on BSSE for requirements is missing
3. Clear explanation of fitness function and representation is needed
4. Clear setting out of empirical Software Engineering aspects is needed: research questions and answers to them.

These changes do not require any new experimental work, but merely refactoring of the paper and the results and some more clarity in places, and inclusion of related work. I think it is a "major revision" (in terms of category of referee outcome), but I am sure it can be achieved with relative ease and that it will make the impact of the paper much greater.

Here are some details:-

The authors make available their code which is helpful. It is not clear whether real world requirements data from JPL is also included in the on line provision. I would like some clarification on this issue in the revised version. Such provision would be a great help to follow on research, though I realize that for reasons of commercial confidentiality this may not be possible.

The take on simulated annealing is an odd one. Clearly it is well known that no search technique can outperform all others for an arbitrary problem (consider the no free lunch theorem). This is, so far as I know, well understood within the SBSE community and its publications and this rather special rendition: SA cannot find the best solution for all problems, seems rather curious (why attack SA in this way, why not any technique, HC, SA, Tabu, GAs ... )?. The authors cite several applications that have used Simulated Annealing, but they are different problems in SBSE so obviously the mileage will differ for each problem and one cannot generalize in the way that the authors do: They include work on testing and modularization as examples where SA is sub optimal. OK. But these are very different problems and have no necessary connection to requirement models. In fact Mancoridis have found that Hill climbing can outperformed simulated annealing \*for\* \*modularization\*; this says \*nothing\* about how it would perform for your problem. It did not seem to make sense to cite other examples of other problems here in the context of showing that SA is not the best available techniques.

So the authors statement:-

"

This is an exciting result since it means that current results from simulated annealing(e.g.[5,8,10,62,73]) could be greatly improved, just by switching to an alternate search engine.

"

really needs to be re-written. This statement is naive (sorry, but it really is!). It shows that the authors did not really understand the No Free Lunch (NFL) Theorem, which is central to search based optimization. Just because technique X beats SA in problem Y does not mean that it will beat it for all problems. The problems cited in the list above by the authors are all very different. NFL means that we cannot generalize from them and certainly can't claim any carry over to the problem for requirements modeling to which the authors apply their techniques in this paper.

I am not saying that this invalidates the paper, but it does require a careful toning down of claims here. I would strongly recommend that the authors read about the NFL too in this context.

In general in SBSE I am not sure that SA is even the most widespread used technique (though even if it were, this is not relevant to the authors argument here because different SBSE problems have different formulations

and therefore results will differ from the requirement modeling problem). I believe that GAs are the most widely used SBSE technique (certainly seems so from the repository on SBSE).

I recommend not to use the phrase "search engine" when referring to search based optimization algorithms. There could easily be a confusion with Google, which would not help!

In related work, some recent work by Finkelstein on multi objective requirements models for fairness is missing (this was published in RE 08) and is clearly relevant since it involves a search based optimization model of requirement engineering. Also there was a survey by Finkelstein et al. on requirements optimization at REFSQ 2008. These two citations should be included in the reviewed version of the paper and worked onto the related work section. The authors should check for other related work on Search based optimization for requirements. They cite work on SBSE, which is fine, but they need to get the work specifically on SBSE for requirements into their related work section.

For instance, other work on SBSE for requirements that is clearly missing and needs to be included is:

1. The paper by Gunther Ruhr at FSE 08
2. The Zhang et al (GECCO 2007) and
3. the work by Bagnall (I&ST 2001).

The paper already has a lot of references, but these are clearly relevant and should be worked into the related work section.

Some of the citations are messed up too. For example citation [5] "S M B and S. Mancoridis". S M B is Brian Mitchell I believe.

Clearly some more scholarly care is required here, both to chase up relevant work and to ensure that the citations are present correctly.

Does Fig 4 add sufficient value to justify a whole page?

The colourful writing style comes up again on p17; Uribe and Stickel "struggled valiantly". I don't think that this style is suitable for an academic journal. Maybe I am old fashioned. I think it should be toned down and the style more dispassionate. Did they really struggle "valiantly" anyway? We are talking about research work not a major international conflict. I hope the authors understand why I raise this. It may well put readers off and that would be a pity.

Fig 10 is too small to read properly.

I missed a clear and unequivocal explanation of what the fitness function and representation were for the Simulated Annealing. This is standard for SBSE work; the representation and fitness function should be very easy to locate in the paper and should be defined clearly formally and with some informal explanation so that other authors can find these and replicate with their own pet search based optimization technique.

I also felt that the preparatory sections built up lots of background, but the meat in the results section was skimmed over by comparison. I would like to see this section extended to explain clearly the research questions, how the associated hypotheses were tested and the finding, mapped back to the research

questions. This is fairly standard for empirical software engineering research and makes it much easier to understand the results.

I liked the novel style opening "The room is crowded and everyone is talking at once.". However this may put off some readers. I expect that this is rather cramping the style of the authors but they should consider boxing out the scenario as a figure. At present the paper is nice as a compelling read, but starting off like this really could distract the more traditional SE reader who have views about how scientific work should be presented. I would not make this a requirement for major revisions, so if the authors want to overrule me on this then that's fine. I expect they knew when they opened with this provocative style that it would draw comment. If other referees comment on it then I think they definitely should add this to the list of "must fixes"

Reviewer #2: This paper presents a comparative performance study of several optimization algorithms that can play roles in automated requirements engineering. The results suggest that the authors' preferred algorithms have the potential to outperform some competing approaches.

The novelty of the work remains unclear. It appears to be little beyond an implementation and very limited performance study of known techniques. The paper presents little other (of real substance) than a small performance measurement experiment, results of which are interpreted as supporting the appropriateness of the algorithm for use in automated requirements engineering.

The title and abstract and indeed the overall narrative of this paper are misleading. This is a paper on a small-scale runtime performance study \*masquerading\* as a paper on requirements analysis. The link between the algorithms and any real advance in requirements engineering is extremely tenuous, and, in any case, not tested or supported by evidence.

The writing in this paper is somewhere between terrible and just plain poor: at the level of sentences, paragraphs, and overall narrative structure. The paper is rife with errors in grammar and usage. It's far too long for the actual new material presented. It takes far too long to get to the point. It never really explains the model at issue. The paper doesn't adequately state the problem it's addressing, the specific technical approach it's proposing to address the problem, the novel claims being made for the approach, the experimental or analytical support for these claims, or the overall conclusions one can draw from this work. To the extent that the paper does explicitly state its contributions, the forms of the contributions are not appropriate for a research paper. The paper is really dominated by tutorial background. The authors lose a good deal of credibility by making a whole slew of writing errors "right out of the block," on pages one and two.

The claims made explicitly or tacitly in this work are largely unsupported by evidence. The scenarios the paper describes are not credible. With due respect to the authors, this paper is nowhere near being suitable for journal publication. It's really a disservice to the reviewing community to submit such work for peer review. Authors should at least proofread their own works before asking others to do so.

Some detailed comments:

Opening scenario isn't credible. A decision as substantial as a redesign of the file system would not be made without documenting it in a way that would make everyone with an interest in that aspect of the aware of the decision. If major architectural decisions were made in this manner, well, the project has bigger problems to worry about than tooling. I suggest that the scenario

There's a grammatical error at the very start of the intro, on page 1: experts on one side the room can make decisions that one hinder decisions and the group is unaware of this conflict.

There's another grammatical error on page 1: As our tools grow better, and they will be used by larger groups who will build more complex models.

The presence of two obvious grammatical errors on page one of a journal submission is not a sign of care in

preparation and not a sign of a promising situation.

There's a grammatical error at the top of page 2: The problem of co-ordinating group discussions is challenging in the 21st century net-enabled world where participants communicate via via multiple channels;

Numerous statements in this paper are speculative and without support, e.g., For example, suppose a requirements analyzer finds a major problem or a novel better solution. Such a result would command the attention of the whole group, in which case everyone in the room would interrupt their current deliberations to focus on the new finding.

The writing in this paper has many problems. One problem is that antecedents are not always clear, as in this sentence: A premise of this approach is that requirements analyzers offer feedback . To what does this approach refer?

References are made to important works without citations, e.g.,: Based on current projections from JPL .

This paper states that the main contribution is a problem definition. A problem definition is not really a contribution worth publishing without substantiation of the importance of the problem and a solution. "The specific contribution is to define the problem of real-time requirements engineering."

The paper is very unclear, at least in the first three pages, on the nature of the formalism being used to represent requirements. It's clearly some kind of logic-based approach, but what is it precisely? Moreover, the paper presents at most anecdotal evidence that the new method is promising: At least for the models explored in this paper, we can achieve optimizations in around 10<sup>2</sup> seconds.

Nor is commenting on a search engine a research contribution: Another important result from this paper is to comment on a standard search engine, used widely in the field of search-based software engineering (SBSE).

Spelling error: we are willing to trade off representational or constrain expressiveness for faster runtimes.

This paper makes strong claims for the proposed approach before the approach is explained and certainly before supporting evidence is offered, e.g.: It is trivial for our preferred method (KEYS and KEYS2) to offer robust information around partial solutions.

Ambiguous antecedent: DDP provides a succinct ontology for representing this design process.

Use of undefined term: What assertions? The DDP tool supports a graphical interface for the rapid entry of the assertions.

Unexplained and unsubstantiated claim: Cost savings in at least two sessions have exceeded \$1 million, while savings of over \$100,000 have resulted in others.

Critical but unexplained and unsupported assumption: Our proposed solution to real-time requirements optimization assumes that the behavior of a large system is determined by a very small number of key variables.

Reviewer#3:

I think there is a kernel of some interesting novel work here. Unfortunately, the paper in its current form is not publishable for several reasons.

1. It apparently has not been proofread even quickly. The errors here are obvious and quite frequent.

2. The whole discussion of the "real time requirements

optimization" problem, while a nice idea and interesting, falls flat by being bogus. It is interesting to try to calculate how fast things have to go today in order that five years from now the tools handle the load. That is good.

However, I cannot follow much of the actual quantitative reasoning. Here are some questions:

- How do you know X is likely between 1 and 4? Do you have measurements or an argument to back this up? How do you know it isn't 200?
- You claim that KEYS2 meets the 0.01 target, but \*none\* of the performance results in Figure 11 meet this, except for the small model. It seems like 0.038 now could balloon into  $(5\text{sec}) \times 3.8$  or even  $(5\text{ sec})^{(3.8)}$ .
- Why is there one number (0.01) for all current-day models? It seems like Model 5 is already 3 times bigger than Model 2, so why is it necessary that both meet the 0.01 requirement? Or are you claiming that 0.01 is the bound for the biggest model of today? (If so, then KEYS2 is off by a factor of nearly 4.)
- In the appendix, you say machines in 2013 will be "560%" faster, but you actually mean "5.6 times faster". And this is wrong, because you have the Moore's Law calculation wrong: the doubling is every 18 months, not two years. So this number should be a factor of 10.1 not 5.6. This presumably means we should shoot for 0.02 today, no?
- Why are models expected to grow by a factor of 8? Is it we cannot type fast enough today? We will be eight times smarter then? What is the basis for this claim?

3. It is weird in Figure 12 that the x axes go 1,10,20,30, etc. Why isn't this 0,10,20,30,...?

4. First paragraph of Sec 7: Why? why is it "unfair" to compare BDD-based against search-based? Aren't they solving the same problem?

5. Sec 7.2: what is the variable "C"?

6. I'm not sure it is valid to make claims about KEYS2 performance versus MaxWalkSAT when you are actually comparing to MaxFunWalk. You have to do a better job showing how they are related. (E.g. at what level are they the same algorithm?)

7. Sec 7.3 para 2: you claim the scoring function is "a Euclidean distance measure". I'm not sure what you mean by that. Do you mean an admissible metric? An L-p norm? What? Please say more and convince us of this and show why it is important.

8. Last para of Sec 9. "...some requirements models". This comes out of the blue. For which reqts models does KEYS approach fail?

9. In general, what are the limits of applying KEYS/KEYS2? When can we expect them to fail? To be worse than A\*?

---