Qualitative vs. quantitative contribution labels in goal models: setting an experimental agenda

Sotirios Liaskos Saeideh Hamidi Rina Jalman

School of Information Technology,
York University
Toronto, Canada
{liaskos,hamidi,rjalman}@yorku.ca

Abstract. One of the most useful features of goal models of the *i** family is their ability to represent and reason about satisfaction influence of one goal to another. This is done through contribution links, which represent how satisfaction or denial of the origin of the link constitutes evidence of satisfaction/denial of its destination. Typically in the *i** family, the nature and level of contribution is represented through qualitative labels ("+", "−", "++" etc.), with the possibility of alternatively using numeric values, as per various proposals in the literature. Obviously, our intuition seems to suggest, labels are easier to comprehend and to come up with, while the use of numbers raises the question of where they come from and what they mean, adds unwarranted precision and overwhelms readers. But are such claims fair? Based on some early experimental results, we make the case for more empirical work on the matter in order to better clarify the differences and understand how to use contribution representations more effectively.

Key words: requirements engineering, goal modeling, i-star

1 Introduction

Goal models of the *i** family [11, 2, 4] have been found to be useful for qualitatively representing and analyzing how stakeholder goals influence each other. The concept of the *contribution* link is used to show how satisfaction or denial of the goal which is origin to the contribution affects satisfaction/denial of the goal that is targeted by the contribution. The level of contribution, i.e. how strong the influence is, is represented with a label that decorates the link. Typically, this label is a symbol such as "+", "−−", "++" etc. denoting both whether the contribution is positive or negative and offering a coarse characterization of the strength of contribution. However, the use of numerical labelling has also been proposed [4, 8, 1]. Such labels may be simple numeric instantiations of the same modeling principles (e.g. [4]) or can have quite more distinct semantics from their qualitative counterparts [8, 1], implying also different ways of inference about satisfaction influence.

Which type of label should we then use? Our intuition suggests that qualitative labels are easier to comprehend and to come up with, while the use of numbers automatically raises the question of where they come from [7, 3], adds unjustified precision and discourages readers. Is that true? In this paper, we review two of our exploratory experimental studies that seem to suggest that use of numbers is not to be dismissed. In both studies, participants are asked to perform ad-hoc reasoning about optimal solutions by just looking at different goal graphs. The results do not offer any evidence that numerical representation obstructs success in that task; instead participants seem to be able to find the optimal solution using the numbers. Using these studies we make a case for more experimentation on the subject, in order to not only learn more about the visual properties of goal modeling languages but also force ourselves explicate what exactly we intend those visual languages to be used for.

The rest of the presentation is organized as follows. In Section 2 we discuss qualitative versus quantitative contribution in goal models. In Section 3 we describe our experiments. Then in Sections 4 and 5 we offer conclusions and our plans for the future.

2 Objectives of Research

Modeling and reasoning about contribution in $i*$ languages is generally understood as something to be done in a qualitative fashion. The use of qualitative labels is rooted on the idea that Non-Functional Requirements (NFRs), the matter modeled through e.g. $i*$ soft-goals, need to be dealt with in a way that does not necessitate availability of accurate empirical data and complex and hard calculations of global optima [10]. Through qualitative analysis, rough assessment of goal satisfaction is possible by examining whether there is evidence of support for some satisfaction of a goal combined with lack of evidence against such satisfaction. This is sufficient to know that important NFRs are satisfied to a good enough degree. Thus, as the framework opts for rough characterizations of satisfaction and influence thereof rather than precise analysis of quantitative data from the field, using qualitative labels is a logical choice.

However, Giogrini et al. show that numbers can be used as well [4]. These numbers are not measurements from the field but simply numeric representations of satisfaction and contribution levels, otherwise presented with qualitative symbols. Similar attempts, which depart from the label propagation semantics have also been proposed in the literature as surveyed by Horkoff and Yu [6]. But the utilization of numbers in place of qualitative labels raises two important issues: (a) the question how numbers (with all their precision) are elicited and (b) the suspicion that "+" and "−" are easier to comprehend as contribution labels than, say, 0.9 and 0.3.

We have recently shown, however, that numeric representations of contributions have their merits [8]. With respect to question (a) above, i.e. where the numbers come from, making the assumption that the goal graph is acyclic and separated from a hard-goal AND/OR decomposition, we showed that the Analytic Hierarchy Process (AHP) can be used to elicit numeric contribution links – an idea also proposed earlier [9]. Simply, the soft-goal hierarchy is viewed as AHP criteria hierarchy and each OR-decomposition is treated as a separate decision problem. As such, the optimal solutions that the graph yields can be argued to be as valid as the AHP decisions they correspond to. This, in turn, supports the relevance of the numeric labels themselves when used for that specific purpose (i.e. deciding optimal solutions).

What is perplexing, though, is concern (b) above: the use of the numbers not to just make the AHP decisions but also as contribution labels on the goal model in order to convey information to readers. Moreover, the question seems to extend to both qualitative and quantitative approaches. The issue seems to be the difficulty to define what exactly this information is, i.e. what exactly the reader is supposed to learn or understand by looking at models such as those of Fig. 1. Moreover, if we explicate the objective of the representation, it is logical to subsequently ask whether there are ways to read it (i.e. ways to understand labels and how they combine) that are more natural and effective than others. Below we describe our attempts to understand this problem better via conducting two small experimental studies.

3 Scientific Contributions

The information contained in the contribution structure (i.e. a portion of a goal model containing contribution links) amounts to a set of binary relations between goals showing how one contributes to the other. Of course, this information may as well be expressed in text or a catalogue of separate logical formulae. But we choose to put all these individual pieces in a graph, because we apparently aim at presenting a whole that emerges by combining them. In particular, if we consider contribution structures to be visual representations of a decision-making problem, we seem to be hoping that using a graph facilitates ad-hoc detection of good decisions. In other words, by just looking at such diagrams some readers must be able to intuitively combine individual contribution links and detect optimal solutions. Moreover, it may be fair to even assume that the detection process is natural and does not assume prior training to the language.

We investigated whether the current representations of contribution indeed have such properties and whether the choice of contribution labels (qualitative vs. numbers) has any influence [8]. We presented to ten (10) experimental participants small goal models (5 soft-goals, 11-15 contributions) with either quantitative or qualitative labels. We first constructed the quantitative ones using random AHP priority numbers. Then, to construct the qualitative ones, we replaced the numbers in the quantitative models with qualitative labels by discretizing the continuous interval [0,1]. So a value in [0,0.2) is replaced by "$--$", a value in [0.2,0.4) is replaced by "$-$" etc. We presented the models to the participants in a within-subjects counterbalanced design and asked them to find for each model, without using any other aids, the optimal out of two/three options (children of OR-decompositions). We used the AHP definition of optimality in both cases. The participants are graduate students of Information Technology, and are not told anything about how to reason with goal models beforehand.

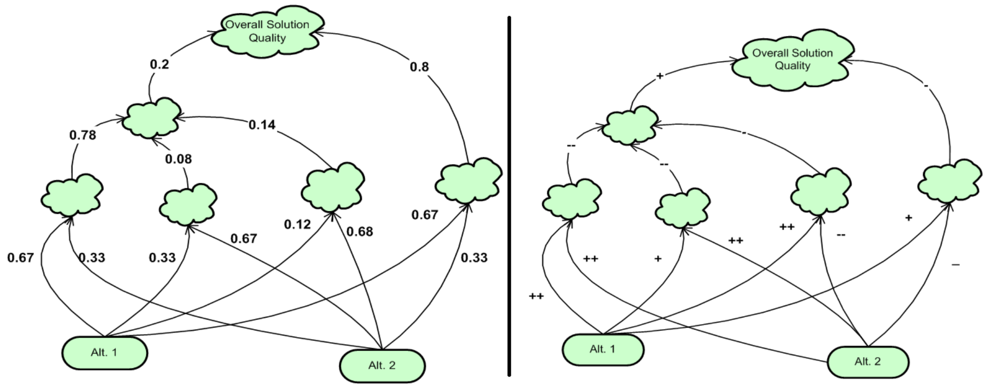

Fig. 1. Which solution has the best overall quality? (you have 60 seconds!)

In the quantitative case, the participants answered correctly in the majority of the times; the binomial test confirmed that this is not the result of randomness (0.95 significance). We might then be allowed to hypothesize that, for small models, the visual result of quantitatively labelling goal models allows readers to correctly guess the optimal (according to AHP) decision, by just looking at the model. Did use of qualitative labels ("+", "−" etc.) have an even greater effect? Our result was that the participants were not more or (significantly) less able to actually identify the correct (according to AHP again) alternative, compared to the quantitative case.

Inspired be these early indications, we recently performed a second study. This time we presented 10 models, five (5) quantitative and five (5) qualitative to another eight (8) participants. Each goal model of one group matches a model of the other group in all aspects except for the contribution labels, as seen in Fig. 1. For the labels, we randomly assign contribution symbols and values, to produce qualitative and quantitative models respectively. Overall, the models contain 4 to 7 soft-goals and 6 to 14 contribution links in various organizations and two alternatives to choose from. We present the models separately in a random order to the participants and we give them sixty (60) seconds to select the optimal alternative for each. The optimal solution in the qualitative models is this time based on the standard label propagation algorithm. Of the forty (40) answers received for each type of model, qualitative and quantitative, sixteen (16) and thirty (30) agree respectively with our calculation of the optimal. The binomial test shows that in the latter case (the quantitative responses), the result is significantly unlikely to be random. Further, while in the qualitative case the successful responses fluctuate with respect to model size (in a non-characterizable fashion), the successful responses in the quantitative case are unaffected by model size (6 out of 8 persons get it right for all models).

To sum up, considering contribution structures as visual instruments for ad-hoc detection of optimal solutions, these small studies aim at understanding whether there is an a-priori way by which uninitiated readers expect that such instruments work. The preliminary evidence appears to support that some participants may have some prior inclination to deal with numbers in a specific way. This, we conjecture, might be due to the fact that dealing with weights, percentages and proportions is much more common in education and daily life than the use of custom symbols like qualitative labels. Regardless, we believe that until we have conclusive results we must resist the temptation to dismiss numerical representations of contribution as confusing or difficult to comprehend.

4 Conclusions

The possibility of qualitative satisfaction analysis is one of the major advantages of using goal modeling notations of the i^* family. However, when one considers goal models as visual instruments for making quick assessments about optimal decisions, the question of naturalness and efficiency of the representation arises. Our preliminary trials on readers without prior training seem to suggest that numeric representations evoke a way of visual reasoning that is consistent with simple aggregation arithmetic over contribution labels.

If further study confirms this or other similar results, the emerging research problem is how we incorporate them in the language itself, also in a way that both the good properties of qualitative reasoning are preserved and a systematic elicitation approach remains available. For example, designs that depart from static visual representations, such as interactive evaluation procedures [5, 3] seem to offer a possible answer to this problem by combining the elicitation and representation aspects. In any case, the more we focus on the role of the goal model as a communication and comprehension aid, the more relevant empirical work with ordinary readers becomes.

5 On-going and Future Work

The specific empirical goal we have set, i.e. that of understanding the comprehensibility of different labelling and aggregation strategies for contributions, requires us to consider many more experimental trials before getting a clear picture. Sizes, structures of models and domains they represent, numerical precision levels or even font sizes and shapes are examples of simple variables that need to be controlled for.

Furthermore, depending on what training and background assumptions one makes for the use of i^* representations in practice there are at least two ways to organize the investigation. One possibility is to keep looking for natural a-priori visual reasoning procedures, formed through other experiences of the reader or potentially utilized through analogies. In that case the educational, professional or other background of the reader becomes an important factor. We plan to run the same experimental procedures with student participants from non-mathematical disciplines such as humanities and from random samples from the general population. We plan to also add a qualitative component in order to elicit how participants exactly think whenever they try to reason about contribution structures. Are for example any analogies utilized, i.e. does the representation remind them of something familiar to which they refer in applying a reasoning strategy? A second possibility is to assume that contribution labelling and aggregation models are a subject to be trained at beforehand. In that case, the learnability of these models needs to be tested, such as how soon and with what accuracy trainees are able to comprehend a contribution structure.

Finally, in all cases, comprehensibility criteria can be defined in different ways. For us it is the ad-hoc detection of optimal solutions. Alternative criteria include how quickly and accurately the user can read and remember individual contributions, successfully explain a given satisfaction degree or, say, detect conflicting nodes. By thinking of concrete criteria we actually force ourselves to think, decide and propose what uses of goal models are important and why. This call for reflection is, we find, one of the greatest benefits of experimental work.

References

1. Daniel Amyot, Sepideh Ghanavati, Jennifer Horkoff, Gunter Mussbacher, Liam Peyton, and Eric S. K. Yu. Evaluating goal models within the goal-oriented requirement language. *International Journal of Intelligent Systems*, 25(8):841–877, 2010.
2. Daniel Amyot and Gunter Mussbacher. User requirements notation: The first ten years, the next ten years. *Journal of Software (JSW)*, 6(5):747–768, 2011.
3. Golnaz Elahi and Eric S. K. Yu. Requirements trade-offs analysis in the absence of quantitative measures: a heuristic method. In *Proceedings of the ACM Symposium on Applied Computing (SAC'11)*, pages 651–658, Taichung, Taiwan, 2011.
4. Paolo Giorgini, John Mylopoulos, Eleonora Nicchiarelli, and Roberto Sebastiani. Reasoning with goal models. In *Proceedings of the 21st International Conference on Conceptual Modeling (ER'02)*, pages 167–181, London, UK, 2002.
5. Jennifer Horkoff and Eric Yu. Finding solutions in goal models: an interactive backward reasoning approach. In *Proceedings of the 29th International Conference on Conceptual modeling (ER'10)*, ER'10, pages 59–75, Vancouver, Canada, 2010.
6. Jennifer Horkoff and Eric Yu. Comparison and evaluation of goal-oriented satisfaction analysis techniques. *Requirements Engineering Journal (REJ)*, 2012.
7. Emmanuel Letier and Axel van Lamsweerde. Reasoning about partial goal satisfaction for requirements and design engineering. In *Proceedings of the 12th International Symposium on the Foundation of Software Engineering FSE-04*, pages 53–62, Newport Beach, CA, 2004.
8. Sotirios Liaskos, Rina Jalman, and Jorge Aranda. On eliciting preference and contribution measures in goal models. In *Proceedings of the 20th International Requirements Engineering Conference (RE'12)*, pages 221–230, Chicago, IL, 2012.
9. N.A.M. Maiden, P. Pavan, A. Gizikis, O. Clause, H. Kim, and X. Zhu. Making decisions with requirements: Integrating i* goal modelling and the AHP. In *Proceedings of the 8th International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ'02)*, Essen, Germany, 2002.
10. John Mylopoulos, Lawrence Chung, and Brian Nixon. Representing and using nonfunctional requirements: A process-oriented approach. *IEEE Transactions on Software Engineering*, 18(6):483–497, 1992.
11. Eric S. K. Yu. Towards modelling and reasoning support for early-phase requirements engineering. In *Proceedings of the 3rd IEEE International Symposium on Requirements Engineering (RE'97)*, pages 226–235, Annapolis, MD, 1997.