Towards Model Repair by Human Opinion–Guided Reinforcement Learning

Kyanna Dagenais
McMaster University, Hamilton, Canada
dagenaik@mcmaster.ca

ABSTRACT

Model repair often entails long sequences of model transformations. Finding the correct model repair sequence is challenging, and its complexity increases with the number of model transformations involved in the repair sequence. In realistic, longitudinally extensive modelling settings, the same model repair scenarios might be encountered repeatedly, providing an excellent opportunity to learn the most appropriate repair actions through reinforcement learning (RL). While such ideas have been explored before, the efficiency of RL-based methods in long repair sequences is still an open challenge. In this paper, we propose a method to improve learning performance by human opinions—cognitive constructs that are subject to uncertainty, but also emerge earlier than hard evidence. Our findings indicate that opinion-based guidance significantly improves the learning performance, even with moderately uncertain human opinions. To counter the uncertainty of individual human advisors, our method allows for collaborative guidance by experts of various expertise and skill levels.

CCS CONCEPTS

• **Computing methodologies** → **Reinforcement learning**.

KEYWORDS

human guidance, machine learning, subjective logic, uncertainty

ACM Reference Format:
Kyanna Dagenais. 2024. Towards Model Repair by Human Opinion–Guided Reinforcement Learning. In *ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems (MODELS Companion '24), September 22–27, 2024, Linz, Austria.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3652620.3676878

1 PROBLEM AND MOTIVATION

Model repair is a crucial activity that helps maintain models in a valid state. In real complex models, model repair frequently requires elaborate and long sequences of repair actions. Often, these repair sequences are beyond the humanly feasible limit, necessitating automation. One possible automation method is reinforcement learning (RL). Model repair is a longitudinally extensive endeavour; in real settings, modelling a complex system is a process that takes months or years. This provides plenty of opportunities to encounter similarly invalid models and learn the best course of action. RL is a particularly appropriate learning method in such settings because it does not require massive volumes of historical data. Instead, RL learns through trial and error. Such avenues have been explored in model-driven engineering (MDE) previously [3], but mainstream adoption is still lagging behind.

One of the key limitations of RL-based methods is their shallow learning curve. In early learning phases, RL is less performant in large and complex state spaces [2]. While performance improves in later phases, this improvement often comes too late, invalidating the purpose of using RL for complex tasks, e.g., model repair.

To improve the performance of RL, we propose leveraging the domain knowledge of modellers for guidance. In model repair by human-guided RL, modellers identify key states in the state space (i.e., instance models) and express whether the specific state is beneficial (likely converges to a valid model) or disadvantageous (likely will lead to other model errors). To accommodate the inherent uncertainty in these hints, we rely on human **opinions** – cognitive constructs that reflect uncertainty. They emerge earlier than hard evidence, but must be approached in a systematic fashion to form a sound base of reasoning. We rely on subjective logic (SL) [12] for an expressive and mathematically sound framework for quantifying human preferences. SL also defines compositional semantics for combining an arbitrary number of opinions into one joint opinion that reflects the beliefs of different individuals. The ability to include multiple human advisors, in turn, positions our method as a particularly good fit with real, industry-scale modelling settings.

Results. Our work demonstrates **significantly improved learning performance**, indicated by a much steeper learning curve compared to traditional RL, and higher cumulative reward.

Benefits. Our method fosters **faster model repair** thanks to the improved learning performance of the RL agent. Our method also allows for including the **human in the loop** as our results indicate that even fairly uncertain and sparse advice can improve learning performance. Finally, our method allows for **collaborative guidance of model repair**, rendering human advice more reliable as individuals with various expertise and skill levels can be involved.

Example We use the running example shown in Fig. 1. The domain model is rendered invalid by the ambiguity of the `name` feature in the `WebApp` class due to both defining and inheriting such a feature. In order to fix this model, an RL based automated model repair algorithm may take one of several actions, including removing classes or attributes. An engineer working with this model may have beliefs about these actions, as shown below.

(1) Removing class `NameElement` → Not beneficial
(2) Removing attribute `name` from `NameElement` class → Neither

MODELS Companion '24, September 22–27, 2024, Linz, Austria

ACM ISBN 979-8-4007-0622-6/24/09
https://doi.org/10.1145/3652620.3676878

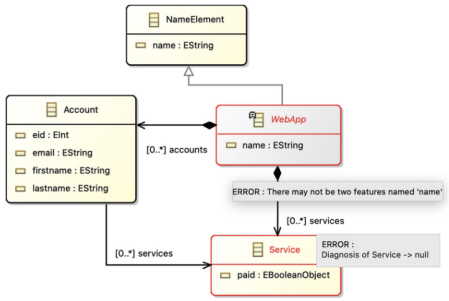


Figure 1: Running example. (Adopted from [2].)

(3) Removing attribute `name` from `WebApp` class → Beneficial

It would be ideal to guide the RL algorithm by these opinions, because they reflect the ideas of modeling experts. However, opinions expressed in natural language lack clear semantics necessary for a formal approach. Thus, we develop an approach in which opinions and RL policies are expressed and fused through SL [12].

2 BACKGROUND AND RELATED WORK

Reinforcement learning. RL [15] is a subset of machine learning formalized by Markov decision processes, in which an agent acts sequentially to learn optimal control of an environment. The environment is composed of states observable by the agent, and the agent chooses its actions in accordance with the prevalent state. The environment transitions to a new state and produces a reward based on the agent's action. The agent acts according to a policy, defined as a mapping from states to actions. This policy, $\pi(\text{action}|\text{state})$, gives the conditional probability of an agent choosing a particular action when in a particular state. By balancing exploration and exploitation, the agent aims to learn the optimal policy, i.e., the one that maximizes the sum of future rewards.

RL has been gaining popularity in MDE, with recent applications in model repair [2], model transformations [9], AI simulation by digital twins [13], and inference of simulators [8]. Our work includes MDE experts' intelligence into RL to improve its performance and to narrow the gap between human cognition and automation.

Opinions and subjective logic (SL). Relying on opinions to guide otherwise autonomous computer machinery (such as RL agents) requires epistemic uncertainty to be approached formally.

SL [12] is a formalism that defines the construct of an *opinion*. A (binomial) opinion is defined as a tuple $\omega_x = (b_x, d_x, u_x, a_x)$, about the truth of a boolean predicate x, where b_x is belief in x, d_x disbelief in x, u_x vacuity of evidence of x, and a_x prior probability of x. The parameters are subject to the constraints that $b_x, d_x, u_x, a_x \in [0, 1]$, and $b_x + d_x + u_x = 1$. The transformation of a binomial opinion to the domain of probability is defined as $P(x) = b_x + a_x u_x$, while a probability $p(x)$ transformed to a binomial opinion is defined as $\omega = (p, 1 - p, 0, p)$. SL also defines fusion operators, functions that map two binomial opinions into a new joint binomial opinion.

Example In the running example, opinions with complete certainty ($u = 0$) correspond to the following binomial opinions:

(1) Removing class `NameElement` → Not beneficial
$\omega = (0.2, 0.8, 0.0, 0.33)$
(2) Removing attribute `name` from `NameElement` class → Neither
$\omega = (0.5, 0.5, 0.0, 0.33)$
(3) Removing attribute `name` from `WebApp` class → Beneficial
$\omega = (0.8, 0.2, 0.0, 0.33)$

SL has gained attention within MDE as a method to address uncertainty in models. Troya et al. [16] present a survey of uncertainty in software models, identifying SL as a method to model uncertainty. To drive rule-based inconsistency management, Jongeling and Vallecillo [11] annotate models with uncertainty information translated into SL, allowing for alignment of several stakeholder opinions. Burgueño et al. [5] represent human confidence in model elements and infer confidence in generated artifacts. Bagheri and Ghorbani [1] use SL to combine viewpoints in collaborative modeling. These works conclude that opinions from SL offer high utility to MDE. Our work corroborates this observation in guided RL.

Related work. Model repair has been of particular interest in MDE. Closest to our work are model repair techniques that either rely on machine learning, human input, or both. The PARMOREL framework [2] relies on a unique combination of RL and user preferences to repair models. However, it does not accommodate guidance and uncertainty of human input, which limits the approach's performance. Barriga et al. [4] report on artificial intelligence driven approaches for model repair. These approaches include neural networks, genetic algorithms, tree planning, RL, and more. While the approaches are diverse, the paper recognizes that RL drives the majority of current work that uses artificial intelligence for model repair. To that end, our work focuses on improving the performance of RL agents for model repair. Eisenberg et al. [10] use multi-objective optimization to explore model transformation chains by user-defined optimization objectives. These objectives include number of transformation steps, transformation coverage, and model coverage. This approach uses meta-heuristic algorithms. In contrast we use RL.

3 APPROACH

We now present the approach to model repair by human-guided RL and evaluate it on a synthetic case. First, the advice is provided (Sec. 3.1). This step includes calibrating the uncertainty of the advice and subsequently, compiling the advice into a subjective opinion. Second, the RL agent's policy is shaped (Sec. 3.2). This step includes translating the policy, which consists of probabilistic values, to subjective opinions, fusing the previously obtained advice (as a subjective opinion) with the policy (as a subjective opinion), and finally, translating the fused policy into the probability domain.

3.1 Providing advice

In the initial step of our approach, *advice* is provided by the advisor once before the agent begins training. This advice denotes the perceived benefits of the RL agent performing specific actions in particular states. It is a subjective construct of the advisor's beliefs and is subject to uncertainty on the end of the advisor. As it is difficult to meaningfully express oneself numerically, we provide mechanisms to calibrate base rate and uncertainty and a DSL to elicit an attitude that aids in shaping advice into an opinion.

Calibrating a and u. To calibrate base rate a, the likelihood of an action being beneficial with no prior information, we use the structural properties of the problem. For example, given a set of actions A, the base rate is defined as $a = 1/|A|$. Uncertainty u is calibrated by an appropriate distance metric, e.g., using the quantification of domain expertise. For example, a mechanical engineer working with a mechanical model should be assigned a lower uncertainty than a mechanical engineer working with an electrical model.

Computing opinions. After setting a and u, belief b and disbelief d are computed. Given the constraints from Sec. 2, once a value for u is calibrated, the remaining weight $(1-u)$ is distributed between b and d. To elicit attitude towards options, we express advice values in an n-point scale. Thus, the calculation for the jth item in this scale in ascending order of confidence from least to most is defined as $b = \frac{j-1}{n-1} \times (1-u) \mid j \in \{1..n\}$, and $d = (1-u) - b$.

Example In the running example, since there are 3 actions, the base rate is calibrated as $a = 0.33$. We will assume the uncertainty has been calculated using an appropriate metric to $u = 0.2$. For eliciting attitude, we assume there is a 3-point scale, with levels (i) not beneficial, (ii) neither, and (iii) beneficial. Thus, the engineer's advice may be expressed as an opinion, as follows.

(1) Removing class `NameElement` → Not beneficial
$\omega = (0.15, 0.65, 0.2, 0.33)$
(2) Removing attribute `name` from `NameElement` class → Neither
$\omega = (0.4, 0.4, 0.2, 0.33)$
(3) Removing attribute `name` from `WebApp` class → Beneficial
$\omega = (0.65, 0.15, 0.2, 0.33)$

3.2 Policy shaping

After transforming advice into an opinion, it can guide the agent via policy shaping. Policy shaping involves guiding the agent by biasing its exploration strategy. In our approach, this means infusing the agent's policy with advice in the form of opinions.

Transforming the policy. Since the agent's policy is expressed in terms of probabilities (Sec. 2), first, we transform the agent's policy from probabilities to opinions. This is achieved by taking the policy for each state-action pair $\pi(\text{action}|\text{state}) = p$ and writing it as an opinion $\omega_{\text{agent}} : p \mapsto (p, 1-p, 0, p)$.

Fusing advice and policy. The transformed policy is fused with the advisor's opinions. As explained in Sec. 2, SL provides fusion operators to combine opinions. While several operators exist, our approach uses Belief Constraint Fusion (BCF), described in detail in [12]. For each opinion provided by the advisor, the corresponding agent opinion about the same state-action pair is fused to create a new, joint opinion, such that $\omega_{\text{agent}} \otimes \omega_{\text{advisor}} \to \omega_{\text{fused}}$.

Transforming fused opinion from opinion to probability domain. Finally, the fused opinions ω_{fused} are translated back to probabilities. This step uses the relationship described in Sec. 2, such that $p : \omega_{\text{fused}} \mapsto b + au$. Since the actions for each state form a complete probability space, the condition that for all actions available in a particular state $\sum p(\text{action}|\text{state}) = 1$ must hold. Thus, we apply normalization to scale the sum of probabilities.

3.3 Evaluation

We evaluate the effect of human advice on RL-based model repair by simulation. We are interested in modelling scenarios in which long sequences of edit operations are learned. Such data sets are not publicly available, so we resort to a synthetic simulation scenario.

Setup. We model the long-running model repair endeavour in the simulation scenario as an RL problem with many *episodes*. The RL agent's goal is to *find a path* from the start state (broken model) to the goal state (valid model) while avoiding terminal states (introducing other errors or irreparably breaking the model). A terminal state ends the episode with no reward, while the goal ends the episode with a reward of 1. We mark 20% of the state space as terminal. Long repair sequences are modelled by placing the start and goal states sufficiently far from each other. The agent navigates through this space using a *predefined set of actions* and learns the most optimal sequence of actions. We simplify the number of possible actions to four CRUD operations. By this, our problem is semantically equivalent to a two-dimensional exploration problem for which Open AI's Gym toolkit offers readily available training environments. We use the Frozen Lake environment. We set the *grid size* and *number of episodes* by manual experimentation. We observe that the unadvised agent shows sufficient improvement after about 5 000 episodes in a 12×12 grid; thus, we use 10 000 episodes. This grid size requires particularly long edit sequences—at least the Manhattan distance between the start and goal state, i.e., 22 steps.

As is customary in RL benchmarks, we use cumulative reward as our evaluation metric, i.e., the sum of rewards accumulated throughout training. A well-performing agent will accumulate rewards in early episodes and continuously gain rewards throughout training.

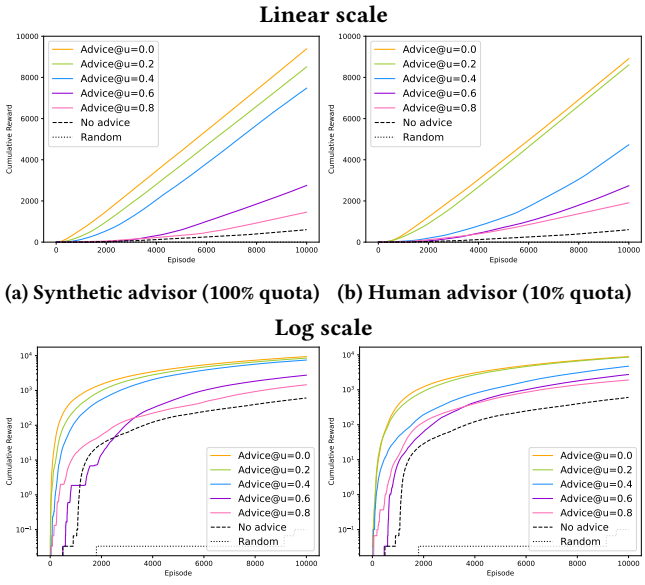
3.4 Results

Advised RL performs significantly better than unadvised RL. Fig. 2 shows a **steeper learning curve** in advised cases compared to the unadvised one. Even with limited human advice, learning performance improves substantially earlier than in the unadvised case. Both with synthetic and human advice, the **cumulative reward is higher** than without advice.

Even uncertain advice is useful. As shown in Tab. 1, in the cases of moderate-high uncertainty ($u = 0.6, 0.8$), the advised agents collect **more cumulative reward** than the unadvised agent.

Human guidance is as effective as an idealized, fully-informed synthetic advisor. This trend is demonstrated by the **comparable cumulative rewards** in Tab. 1. The performance is within 5% in three of five evaluated cases ($u = 0.0, 0.2, 0.6$), with one case showing an improvement of the human over the machine ($u = 0.2$). In two of five evaluated cases, the difference is more pronounced, once in favor of the machine advisor ($u = 0.4$) and once in favor of the human advisor ($u = 0.8$). Improvements due to human advice are particularly apparent in highly uncertain situations.

The human advisor is more efficient than the fully-informed oracle. The human advisor achieves a comparable performance by **advising about only a fraction of the design space (10%)**. This alleviates the typical problem of human-in-the-loop RL in which the human becomes the bottleneck.



(a) Synthetic advisor (100% quota) **(b) Human advisor (10% quota)**

(c) Synthetic advisor (100% quota) **(d) Human advisor (10% quota)**

Figure 2: Cumulative rewards with different advisors

Table 1: Cumulative rewards. Bold is best at the given level of certainty. (Unadvised: 607.267. Random: 0.100.)

u	Synthetic	Human	Δ
0.0	**9 386.467**	8 907.733	-5.08%
0.2	8 511.367	**8 607.500**	+1.13%
0.4	**7 476.633**	4 727.433	-36.77%
0.6	**2 751.833**	2 737.600	-0.52%
0.8	1 454.400	**1 907.933**	+31.18%

4 DISCUSSION AND CONCLUSION

The main takeaway of the evaluation is that opinions, even with uncertainty, improve the performance of the RL agent. This suggests that advised agents can navigate from the broken model to the valid model more efficiently than unadvised agents. By testing synthetic and human sources of advice, we see that human guidance is as effective as an idealized advisor in most cases. We note that in some cases, the human advisor is actually more effective than the synthetic one. Thus, it is clear that human creativity is more sophisticated than an idealized machine. This highlights the need for mixed human-machine intelligence in the model repair process. A detailed analysis of results is available in our technical report [7].

As discussed in Sec. 3, we assume uncertainty is calibrated with an appropriate measure. By measure, we mean a set of values $\mathbb{M}$ with operations $0 : \to \mathbb{M}$, $+ : \mathbb{M} \times \mathbb{M} \to \mathbb{M}$ (0 neutral, + associative and commutative) and an order relation $\leq$ on $\mathbb{M}$. Measures to calibrate uncertainty have been explored by Dagenais and David [6]. A pertinent example is the seniority of modelling experts, with more experienced experts receiving lower uncertainty scores than less experienced ones. Another such measure is the relative expertise of single modelers w.r.t. to a field, e.g., mechanical engineers are assigned lower uncertainty when working with mechanical models and higher when working with electrical models.

We assumed advice was given once before training. However, given the appropriate interaction infrastructure, our approach can be applied at any point during the training process. We envision conversational modeling [14] being a critical enabler here. We use a simplistic DSL to provide advice, but see opportunities to generate DSLs for RL guidance in particular problem spaces.

Conclusion. In this paper, we identify opinion-guided reinforcement learning as a significant technique to augment traditional model repair methods. We assess the performance of opinion-guided RL under various uncertainty models to understand the effects of human opinion on the exploration processes, such as model repair with long repair sequences. We observe that opinions, even with moderate levels of uncertainty, result in a steeper learning curve and improve the RL agent's performance. This has clear benefits for model repair as (i) the advantages of RL manifest earlier in the modelling endeavour and (ii) identifying the correct repair sequences becomes tractable. As opinions emerge earlier than hard evidence can be produced, opinion-guided RL offers improvement to traditional methods in model repair with long repair sequences. We envision our work being employed by tool builders to augment model repair methods with mixed human-machine intelligence.

Ongoing work focuses on developing a prototype framework.

REFERENCES

[1] Ebrahim Bagheri and Ali A. Ghorbani. 2009. A belief-theoretic framework for the collaborative development and integration of para-consistent conceptual models. *J. Sys. & Soft.* 82, 4 (2009), 707–729. https://doi.org/10.1016/j.jss.2008.10.012
[2] Angela Barriga et al. 2022. PARMOREL: a framework for customizable model repair. *Soft. Sys. Mod.* 21, 5 (2022), 1739–1762.
[3] Angela Barriga, Adrian Rutle, and Rogardt Heldal. 2019. Personalized and Automatic Model Repairing using Reinforcement Learning. In *ACM/IEEE 22nd Intl. Conf. on Model Driven Engineering Languages and Systems Companion*. 175–181.
[4] Angela Barriga, Adrian Rutle, and Rogardt Heldal. 2022. AI-powered model repair: an experience report—lessons learned, challenges, and opportunities. *Soft. Sys. Mod.* 21, 3 (2022), 1135–1157. https://doi.org/10.1007/s10270-022-00983-5
[5] Loli Burgueño et al. 2018. Expressing confidence in models and in model transformation elements. In *Proceedings of the 21th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*. 57–66.
[6] Kyanna Dagenais and Istvan David. 2024. Driving Requirements Evolution by Engineers' Opinions. In *ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion, MODELS-C*. ACM.
[7] Kyanna Dagenais and Istvan David. 2024. *Opinion-Guided Reinforcement Learning*. Technical Report. arXiv:2405.17287 [cs.LG]
[8] Istvan David and Eugene Syriani. 2024. *Automated Inference of Simulators in Digital Twins*. CRC Press, 122–148. https://doi.org/10.1201/9781003425724-11
[9] Martin Eisenberg, Hans-Peter others Pichler, Antonio Garmendia, and Manuel Wimmer. 2021. Towards Reinforcement Learning for In-Place Model Transformations. In *2021 ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS)*. 82–88.
[10] Martin Eisenberg, Apurvanand Sahay, Davide Di Ruscio, Ludovico Iovino, Manuel Wimmer, and Alfonso Pierantonio. 2024. Multi-objective model transformation chain exploration with MOMoT. *Inf Softw Technol* 174 (2024), 107500.
[11] Robbert Jongeling and Antonio Vallecillo. 2023. Uncertainty-aware consistency checking in industrial settings. In *ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*. 73–83.
[12] Audun Jøsang. 2016. *Subjective logic*. Vol. 3. Springer.
[13] Xiaoran Liu and Istvan David. 2024. AI Simulation by Digital Twins: Systematic Survey of the State of the Art and a Reference Framework. In *ACM/IEEE Intl. Conf. on Model Driven Engineering Languages and Systems Companion*. ACM.
[14] Sara Pérez-Soler et al. 2018. Collaborative Modeling and Group Decision Making Using Chatbots in Social Networks. *IEEE Software* 35, 6 (2018), 48–54.
[15] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
[16] Javier Troya, Nathalie Moreno, Manuel F. Bertoa, and Antonio Vallecillo. 2021. Uncertainty representation in software models: a survey. *Software and Systems Modeling* 20, 4 (2021), 1183–1213. https://doi.org/10.1007/s10270-020-00842-1