Advances in Cognitive Systems X (20XX) 1-6 Submitted X/20XX; published X/20XX

Signature Entrenchment and Counterfactual Reasoning in Automated Theory Change

Xue Li XUE.SHIRLEY.LI@ED.AC.UK
Alan Bundy A.BUNDY@ED.AC.UK
Eugene Philalithis E.PHILALITHIS@ED.AC.UK
School of Informatics, University of Edinburgh, UK

Abstract

This file contains the proofs of theorems in the main paper organised section by section.

1. Introduction

There is no theorems to prove in this section.

2. Conceptual change

3. The ABC Repair System

There is no theorems to prove in this section.

There is no theorems to prove in this section.

4. Measuring Signature Entrenchment

Theorem 4.1. *If there is no path from predicate* p *to predicate* q *in the theory's theory graph, then assertions of* p *cannot contribute to any proof of an assertion of* q*.*

Proof. In Theorem 4.1, no path from predicate p to predicate q in the theory's theory graph means that there is no rule connection between p and q. Therefore, assertions of p and q are independent from each other, so that assertions of p cannot contribute to any proof of an assertion of q. □

Therefore, theorem 4.1 decides whether adding a theorem of p has an impact of building the proofs of q's instances.

4.1 Predicate Entrenchment

Theorem 4.2. *Based on the preferred distance, the important properties that predicate entrenchment* $e(p)$ *has are as follows, where* p*,* p_1 *and* p_2 *are predicates, and* $\mathbb{S}_p$ *is the set of predicates which occur in* $\mathbb{PS}$ *while* $\mathbb{S}_t$ *is the set of predicates which occur in the theory but not in* $\mathbb{PS}$*.*

1. *$\forall p \in (\mathbb{S}_t \cup \mathbb{S}_p),\ e(p)$ has exactly one value.*
   The entrenchment of a predicate should be just one value.

   Proof. Based on Equation (3), the preferred distance of a predicate has only one value, accordingly, the entrenchment score of a predicate has one fixed value because its calculation function is monotonous. In summary, property 1 in Theorem4.2 is held by our measurement. □

2. *$0 < e(p) \leq 1$.*
   The range of an entrenchment should be [0,1], where 0 means that a predicate is not trusted at all and 1 represents that the predicate is most entrenched and fully trusted.

   Proof. Based on Equation (3) and (4), we can conclude that:

   $$\forall pd(p_1) \neq \infty,\ 0 \geq pd(p_1) \geq pd_{Max}$$

   $\because \forall pd(p_1) \neq \infty,\ e(p_1) = 1 - \frac{pd(p_1)}{pd_{Max}+1}$,
   $\therefore 1 - \frac{pd_{Max}}{pd_{Max}+1} \leq e(p_1) \leq 1 - \frac{0}{pd_{Max}+1}$
   Therefore, we can conclude that:

   $$0 < \frac{1}{pd_{Max}+1} \leq e(p_1) \leq 1,\ \forall pd(p_1) \neq \infty \tag{1}$$

   On the other hand, $\forall pd(p_2) = \infty$, we have $e(p_2) = \frac{1}{pd_{Max}+2}$.
   $\because 0 \leq pd_{Max}$
   $\therefore 2 < pd_{Max} + 2$. Thus, we have the follows.

   $$0 < e(p_2) \leq \frac{1}{2} \tag{2}$$

   □

3. *$\forall p_2 \in \mathbb{S}_p,\ e(p_2) = 1\ \wedge\ \forall p_1 \in \mathbb{S}_t,\ 0 < e(p_1) < 1$.*
   Because $\mathbb{PS}$ is more trusted than the theory, a preferred predicate is most entrenched whose entrenchment is 1. Any predicate appearing only in the theory is believed to some extent, but less than a preferred predicate. Meanwhile, any predicate that occurs in the theory is considered to convey some information. Therefore, its entrenchment is bigger than 0 but smaller than 1.

   Proof. $\because \forall p_1 \in \mathbb{S}_t,\ pd(p_1) = 0$.
   $\therefore e(p_1) = 1 - \frac{0}{pd_{Max}+1} = 1$

On the other hand, $\forall pd(p_2) \in \mathbb{S}_t,\ pd(p_2) > 0$, so $pd_{Max} > 0$, we have

$$e(p_2) = \begin{cases} 1 - \frac{pd(p_2)}{pd_{Max}+1} < 1 - \frac{0}{pd_{Max}+1} < 1,\ pd(p) \neq \infty \\ \frac{1}{pd_{Max}+2} < \frac{1}{2} < 1\ pd(p_2) = \infty \end{cases}$$

$\square$

4. $\forall p_1, p_2 \in \mathbb{S}_t,\ e(p_1) > e(p_2)$, *iff* $pd(p_1) < pd(p_2)$.
When neither predicate occurs in $\mathbb{PS}$, p_1 is more entrenched than p_2 if and only if p_1 is closer to preferred predicates in terms of its preferred distance. The smaller $pd(p_1)$ is, the more impact on $\mathbb{PS}$ changing p_1 will have.

Proof. Consider different combinations between p_1 and p_2, we prove this property as follows.

- When $pd(p_1) \neq \infty, pd(p_2) \neq \infty$, if $e(p_1) > e(p2)$, then $1 - \frac{pd(p_1)}{pd_{Max}+1} > 1 - \frac{pd(p_2)}{pd_{Max}+1}$.
$\therefore \frac{pd(p_1)}{pd_{Max}+1} < \frac{pd(p_2)}{pd_{Max}+1}$
$\therefore pd(p_1) < pd(p_2)$.
- When $pd(p_1) \neq \infty, pd(p_2) = \infty$, if $e(p_1) > e(p2)$, then $1 - \frac{pd(p_1)}{pd_{Max}+1} > 1 - \frac{1}{pd_{Max}+1}$.
$\therefore pd(p_1) < 1$, Then it can be concluded that $pd(p_1) < 1 < pd(p_2) = \infty$
- The case that $pd(p_1) = \infty, pd(p_2) \neq \infty$ and $e(p_1) > e(p_2)$ does not exist.
$\because\ e(p_2) = 1 - \frac{pd(p_1)}{pd_{Max}+1} = \frac{pd_{Max}+1-pd(p_1)}{pd_{Max}+1}$, and $pd_{Max} \geq pd(p_1)$, then $pd_{Max} - pd(p_1) \geq 0$.
$\therefore e(p_2) \geq \frac{1}{pd_{Max}+1} = e(p_1)$. Thus, $e(p_1) > e(p_2)$ cannot happen in this case.
- The case that $pd(p_1) = \infty, pd(p_2) = \infty$ and $e(p_1) > e(p_2)$ does not exist because $e(p_1) = e(p-2)$.

In conclusion, $e(p_1) > e(p_2)$ only happen when $pd(p_1) \neq \infty$. We have proved that: $e(p_1) > e(p_2) \implies pd(p_1) < pd(p_2)$.

On the other hand, if $pd(p_1) < pd(p_2)$, then neither of them is infinity.
$\because e(p_1) = 1 - \frac{pd(p_1)}{pd_{Max}+1}$, and $e(p_2) = 1 - \frac{pd(p_2)}{pd_{Max}+1}$.
$\therefore e(p_1) - e(p_2) = \frac{pd(p_2)-pd(p_1)}{pd_{Max}+1} > 0$
In summary, $pd(p_1) < pd(p_2) \implies e(p_1) > e(p_2)$

$\square$

4.2 Argument Entrenchment

There is no theorems to prove in this section.

5. Evaluation

There is no theorems to prove in this section.

6. Conclusion

There is no theorems to prove in this section.