Forgetting Revisited

Yan Zhang and **Yi Zhou**
Artificial Intelligence Research Group
University of Western Sydney
yan,yzhou@scm.uws.edu.au

Abstract

In this paper, we propose an alternative notion, called *weak forgetting*, of forgetting a set of predicates in a first-order theory. One important feature of this new notion is that the result of weak forgetting is always first-order expressible. In contrast, this is not the case for the traditional notion of forgetting, called *strong forgetting*, introduced by Lin and Reiter. As a consequence, these two notions are not exactly the same. Interestingly, we prove that they coincide when the result of strong forgetting is first-order expressible. We also present a representation theorem to characterize weak forgetting from different aspects.

Introduction

In this paper, we reconsider the notion of forgetting in first-order logic. The reason of focusing on first-order logic is threefold. Firstly, first-order logic itself is an important, powerful, and widely used knowledge representation formalism. Secondly, first-order logic plays a central role in the logic family in the sense that many other logics can be viewed as restrictions, extensions and variations of it. Hence, studying forgetting in first-order logic should enlighten us on forgetting in other logics as well. Finally, the technical issues raised in first-order logic are also representative and rather complicated so that forgetting in first-order logic behaves relatively intricate and interesting.

Lin and Reiter (1994) introduced a notion of forgetting (we call it *strong forgetting*) a set of predicates in a first-order theory based on a semantical definition. However, the result of strong forgetting does not always exist. That is, there exists a first-order theory T and a set Ω of predicates such that there is no first-order theory T', which is the result of strongly forgetting Ω in T.

Motivated by this observation, we propose an alternative notion of forgetting (we call it *weak forgetting*) instead. Recently, Zhang and Zhou (2009) proposed four postulates to characterize forgetting in the modal logic S5, and showed that their postulates exactly capture forgetting in S5. In this paper, we define a similar set of postulates in first-order logic and apply them as the notion of weakly forgetting a set of predicates in a first-order theory. We also present a representation theorem to characterize weak forgetting alternatively, from which one can conclude that the result of weak forgetting is always first-order expressible.

As a consequence, weak forgetting and strong forgetting are not exactly the same. Interestingly, they coincide when the result of strong forgetting exists. That is, if the result of strong forgetting is first-order expressible, then a theory is the result of strong forgetting if and only if it is the result of weak forgetting. The similarities and differences between weak forgetting and strong forgetting can also be understood from a model theoretical point of view. As we will show, the models of the result of weak forgetting are exactly the models of the result of strong forgetting closed under elementary equivalence.

Strong Forgetting

The strong form of forgetting a set of predicates in a first-order theory is introduced by Lin and Reiter (1994). Let $\mathcal{M}$ and $\mathcal{M}'$ be two structures of the same signature σ and Ω a set of predicates in σ. We say that two structures $\mathcal{M}$ and $\mathcal{M}'$ are *identical with exception* on Ω, denoted by $\mathcal{M} \sim_\Omega \mathcal{M}'$, iff they agree on everything except the interpretations of the predicates in Ω.

Definition 1 (Strong forgetting (Lin and Reiter 1994)) *Let T be a theory of a signature σ and Ω a set of predicates in σ. A theory T' of σ is the result of* strongly forgetting Ω in T *iff for any σ-structure $\mathcal{M}'$, $\mathcal{M}' \models T'$ iff there exists a model M of T such that $\mathcal{M} \sim_\Omega \mathcal{M}'$. That is, $Mod(T') = \{\mathcal{M}' \mid \exists \mathcal{M} \in Mod(T), \mathcal{M} \sim_\Omega \mathcal{M}'\}$.*

We say that the result of strongly forgetting Ω in T *exists*, or is *first-order expressible*, iff there exists such a T'. Sometimes we use $SForget(T, \Omega)$ to denote the result of strongly forgetting Ω in T if clear from the context.

Proposition 1 *Let T be a theory of a signature σ and $\Omega = \{P_1, \dots, P_n\}$ a set of predicates in σ. Then, the result of strongly forgetting Ω in T is equivalent to $\{(\exists R_1 \dots R_n)\phi(P_1/R_1)\dots(P_n/R_n) \mid \phi \in T\}$, where $R_i, 1 \leq i \leq n$ is a predicate variable that has the same arity as P_i, and $\phi(P_i/R_i), 1 \leq i \leq n$ is the sentence obtained from ϕ by simultaneously replacing every occurrence of P_i in ϕ with R_i.*

Proposition 2 (Non-existence of strong forgetting) *There exists a theory T of a signature σ and a set Ω of predicates* *in σ such that there is no theory of σ, which is the result of strongly forgetting Ω in T.*

Proposition 3 *Let T be a theory of a signature σ and Ω a set of predicates in σ. If there is a theory T' that is the result of strongly forgetting Ω in T, then there exists a theory T_0 of the signature $\sigma \backslash \Omega$ such that T_0 is equivalent to T', i.e., $Mod(T_0) = Mod(T')$, thus is also the result of strongly forgetting Ω in T.*

Weak Forgetting

Sometimes it needs to be very cautious to use the notion of strong forgetting since the result of strong forgetting may not exist (see Proposition 2). For instance, suppose that we intend to apply forgetting to solving inconsistency between two theories. A simple solution, extended from the work in propositional case (Lang and Marquis 2002), is to take the consistent conjunction of forgetting some predicates in both theories as the result. However, strong forgetting cannot be directly applied here because the results of strong forgetting may not exist. Hence, the significance of strong forgetting in first-order logic is severely restricted.

In this paper, we propose a weaker notion of forgetting instead, and we call it *weak forgetting*. As we will show later, the result of weak forgetting is always first-order expressible. Hence, weak forgetting can be used arbitrarily without worrying about its existence, thus is more applicable than strong forgetting.

Definition 2 (Irrelevance) *Let T be a theory of signature σ and Ω a set of predicates in σ. We say that T is* irrelevant *to Ω, denoted by $IR(T, \Omega)$, iff there exists a theory T_0 equivalent to T such that every sentence in T_0 does not mention any predicate from Ω.*

Definition 3 (Weak forgetting) *Let T and T' be two theories of the same signature σ, and Ω a set of predicates in σ. We say that T' is the result of* weakly forgetting Ω in T *iff the following four postulates hold:*

(W) *Weakening: $T \models T'$.*

(PP) *Positive Persistence: for any theory T_0, if $IR(T_0, \Omega)$ and $T \models T_0$, then $T' \models T_0$.*

(NP) *Negative Persistence: for any theory T_0, if $IR(T_0, \Omega)$ and $T \not\models T_0$, then $T' \not\models T_0$.*

(IR) *Irrelevance: $IR(T', \Omega)$.*

We say that the result of weakly forgetting Ω in T *exists*, or is *first-order expressible*, iff there exists such a T'. Sometimes, we use $WForget(T, \Omega)$ to denote the result of weakly forgetting Ω in T if clear from the context.

Similar to knowledge forgetting, (W) means that after forgetting, the resulting theory should be weaker than the original one; (PP) and (NP) mean that forgetting should not affect those positive and negative irrelevant information, respectively; (IR) means that after forgetting, the resulting theory should be irrelevant to those predicates already forgotten. One can observe that these postulates are not independent. For instance, (NP) is a consequence of (W). However, we list all postulates here in order to illustrate the basic intuitions of weak forgetting.

Example 1 Consider a signature with a binary predicate P, a unary predicate Q and the equality symbol $=$. Let T_1 be the theory $\{\forall x \exists y(P(x, y) \vee Q(y))\}$. The result of strongly forgetting P in T_1 can be captured by $\exists R \forall x \exists y(R(x, y) \vee Q(y))$, which is equivalent to $\top$. This shows that $SForget(T_1, P)$ exists. On the other hand, $\top$ is the result of weakly forgetting P in T_1 as well since it is the only theory, up to equivalence, which is irrelevant to P and entailed by T_1.

Let T_2 be $\{\forall x \exists y(P(x, y) \wedge Q(y))\}$. It can be checked that both $SForget(T_2, P)$ and $WForget(T_2, P)$ exist, and both are equivalent to $\exists y Q(y)$.

Example 2 Consider a signature with a single binary predicate P and $=$. Let T_3 be the theory $\{\forall xyz(P(x, y) \wedge P(x, z) \rightarrow y = z), \forall xyz(P(x, z) \wedge P(y, z) \rightarrow x = y), \forall x \exists y P(x, y), \exists y \forall x \neg P(x, y)\}$. Roughly speaking, P in T_3 associates each element in a domain to another element. The first two sentences in T_3 specify that this association is one-to-one. The third one means that each element associates with another one, while the fourth means that there exists an element not associated. A typical model of T_3 is in the infinite natural number domain, and P is interpreted as the successor relation, i.e., $P(x, y)$ iff $y = x + 1$. In fact, every model of T_3 has an infinite domain.

It can be checked that the results of weakly forgetting P in T_3 and strongly forgetting P in T_3 exist, and both are equivalent to the theory $T'_3 = \{\phi_2, \phi_3, \ldots, \phi_n, \ldots\}$, where ϕ_i is $\exists x_1 \ldots x_i \bigwedge_{k \neq m} x_k \neq x_m$, intuitively meaning that the domain has at least i elements. Hence, T_3 is an infinite theory whose models are all infinite structures. However, it is well-known that T'_3 cannot be defined by a single sentence.

Example 3 Consider a signature with a binary predicate P, a unary predicate Q and $=$. Let $T_4 = \{\forall xyz(P(x, y) \wedge P(x, z) \rightarrow y = z), \forall xyz(P(x, z) \wedge P(y, z) \rightarrow x = y), \forall x \exists y P(x, y), \forall y \exists x P(x, y), \forall xy(P(x, y) \rightarrow (Q(x) \leftrightarrow \neg Q(y)))\}$. Roughly speaking, P in T_4 represents a bijection between those elements that Q hold and the rest elements. In other words, the two subsets of elements (that Q hold and that Q do not hold, respectively) have the same cardinality and P is a witness.

The result of strongly forgetting P in T_4 does not exist because the class of all structures, in which the above two subsets have the same cardinality, is not closed under elementary equivalence. That is, there exist two $\{Q, =\}$-structures satisfying the same set of $\{Q, =\}$-sentences but only one of them is in this class of structures. However, the result of weakly forgetting P in T_4 exists, which is equivalent to the theory $T'_4 = \{\exists x Q(x), \exists x \neg Q(x), \phi_2, \phi_3, \ldots, \phi_n, \ldots\}$, where ϕ_i is $\exists x_1 \ldots x_i Q(x_1) \wedge \ldots \wedge Q(x_i) \wedge \bigwedge_{k \neq m}(x_k \neq x_m) \leftrightarrow \exists y_1 \ldots y_i \neg Q(y_1) \wedge \ldots \wedge \neg Q(y_i) \wedge \bigwedge_{k \neq m}(y_k \neq y_m)$, intuitively meaning that if there are at least i elements that Q hold, then there are at least i elements that Q do not hold, and vice versa.

A Representation Theorem

Theorem 4 (Representation theorem) *Let T and T' be two theories of a signature σ and Ω a set of predicates in σ. The following statements are equivalent:*

1. T' *satisfies the four postulates in Definition 3 with respect to* T *and* Ω*;*
2. $IR(T', \Omega)$, $T \models T'$ *and for any theory* T_0 *that satisfies these two conditions,* $T' \models T_0$*;*
3. T' *is equivalent to* $\{\phi \mid \phi \in \mathcal{S}_\sigma, IR(\phi, \Omega), T \models \phi\}$.
4. $Mod(T') = \{\mathcal{M}' \mid \exists \mathcal{M} \in Mod(T), \mathcal{M} \equiv_\Omega \mathcal{M}'\}$.

Corollary 5 (Existence of weak forgetting) *For any theory* T *of a signature* σ *and any set* Ω *of predicates in* σ*, there exists a theory* T'*, which is the result of weakly forgetting* Ω *in* T.

Proposition 6 *Let* T_1 *and* T_2 *be two theories of a signature* σ*, and* Ω_1 *and* Ω_2 *two sets of predicates in* σ.

1. *The result of weak forgetting represents a class of equivalent theories. That is, if* T_1 *and* T_2 *are both the results of weak forgetting, then* $T_1 \equiv T_2$*. Conversely, if* $T_1 \equiv T_2$ *and* T_1 *is the result of weak forgetting, then so is* T_2.
2. *For any sentence* ϕ *such that* $IR(\phi, \Omega_1)$, $T_1 \models \phi$ *iff* $WForget(T_1, \Omega_1) \models \phi$.
3. $WForget(WForget(T_1, \Omega_1), \Omega_2)$ *is equivalent to* $WForget(T_1, \Omega_1 \cup \Omega_2)$.
4. *If* $T_1 \models T_2$*, then* $WForget(T_1, \Omega_1) \models WForget(T_2, \Omega_1)$.
5. *If* $T_1 \equiv T_2$*, then* $WForget(T_1, \Omega_1) \equiv WForget(T_2, \Omega_1)$.
6. $WForget(T_1 \vee T_2, \Omega_1) \equiv WForget(T_1, \Omega_1) \vee WForget(T_2, \Omega_1)$*, where* $T_1 \vee T_2$ *is a shorthand for* $\{\phi \vee \psi \mid \phi \in T_1, \psi \in T_2\}$*, similar for* $WForget(T_1, \Omega_1) \vee WForget(T_2, \Omega_1)$.

Proposition 7 *Let* T *be a theory of a signature* σ *and* Ω *a set of predicates in* σ*. There exists a theory* T' *of the signature* $\sigma \backslash \Omega$ *such that* T' *is the result of weakly forgetting* Ω *in* T.

A fundamental problem that remains unclear is the relationships, including both similarities and differences, between weak forgetting and strong forgetting. In fact, some clues to the answer can be inspired from Theorem 4. We leave this topic to the next section.

Weak Forgetting vs Strong Forgetting

Corollary 8 (Weak forgetting $\neq$ strong forgetting) *Weak forgetting and strong forgetting do not coincide. More precisely, there exists a theory* T *of a signature* σ *and a set* Ω *of predicates in* σ *such that the result of weakly forgetting* Ω *in* T *exists, but the result of strongly forgetting* Ω *in* T *does not exist.*

Theorem 9 (From strong forgetting to weak forgetting) *Let* T *and* T' *be two theories of a signature* σ *and* Ω *a set of predicates in* σ*. If* T' *is the result of strongly forgetting* Ω *in* T*, then* T' *is the result of weakly forgetting* Ω *in* T.

Theorem 10 (From weak forgetting to strong forgetting) *Let* T *and* T' *be two theories of a signature* σ *and* Ω *a set of predicates in* σ*. If* T' *is the result of weakly forgetting* Ω *in* T *and there exists a theory which is the result of strongly forgetting* Ω *in* T*, then* T' *is also the result of strongly forgetting* Ω *in* T.

Corollary 11 (Weak forgetting $\approx$ strong forgetting) *Let* T *be a theory and* Ω *a set of predicates. If a theory* T_1 *is the result of weakly forgetting* Ω *in* T*, and a theory* T_2 *is the result of strongly forgetting* Ω *in* T*, then* $T_1 \equiv T_2$.

Theorem 12 (Model theoretical comparison) *The models of the result of weak forgetting are the models of the result of strong forgetting closed under elementary equivalence. Formally, let* T *be a theory and* Ω *a set of predicates. Then,* $Mod(WForget(T, \Omega)) = \{\mathcal{M}' \mid \exists \mathcal{M} \in Mod(SForget(T, \Omega)), \mathcal{M} \equiv \mathcal{M}'\}$*, where* $Mod(SForget(T, \Omega))$ *is specified as* $\{\mathcal{M}' \mid \exists \mathcal{M} \in Mod(T), \mathcal{M}' \sim_\Omega \mathcal{M}\}$.

To conclude, the main difference between weak forgetting and strong forgetting is located in those cases where the result of strong forgetting is not first-order expressible, e.g. Example 3. On the other hand, according to the existence property of weak forgetting (i.e., Corollary 5), the result of weak forgetting is always first-order expressible. In fact, this difference is important so that weak forgetting is more useful than strong forgetting in some application scenarios.

Conclusion

In this paper, we proposed an alternative notion of forgetting, namely weak forgetting, in first-order logic. In contrast with the traditional notion of strong forgetting, a major advantage of this new notion is that the result of weak forgetting always exists, or equivalently, is always first-order expressible (see Corollary 5). We also presented a representation theorem (see Theorem 4) to characterize weak forgetting from different aspects, which provides us a better understanding of the notion of weak forgetting.

Finally, we consider the relationships between weak forgetting and strong forgetting. In general, they do not coincide (see Corollary 8). However, if the result of strong forgetting is first-order expressible, then they are equivalent (see Corollary 11). To conclude, weak forgetting and strong forgetting are almost the same except in those cases where the result of strong forgetting does not exists but the result of weak forgetting does. From a model theoretical point of view. We proved that the models of weak forgetting are the models of strong forgetting closed under elementary equivalence (see Theorem 12).

Acknowledgement

This research is supported in part by an Australian Research Council (ARC) Discovery Projects Grant (DP0988396). The authors would like to thank Yin Chen and YongMei Liu for their useful discussions on an early draft of this paper.

References

Lang, J., and Marquis, P. 2002. Resolving inconsistencies by variable forgetting. In *KR*, 239–250.

Lin, F., and Reiter, R. 1994. Forget it! In *Proceedings of the AAAI Fall Symposium on Relevance*, 154–159.

Zhang, Y., and Zhou, Y. 2009. Knowledge forgetting: Properties and applications. *Artif. Intell.* 173(16-17):1525–1537.