IMPROVED REVISION OF RANKING FUNCTIONS FOR THE GENERALIZATION OF BELIEF IN THE CONTEXT OF UNOBSERVED VARIABLES

Klaus Häming and Gabriele Peters
University of Hagen, Universitätsstr. 1, 58097 Hagen, Germany

Keywords: Ranking functions, Machine learning, Reinforcement learning, Belief revision, Hybrid learning system.

Abstract: To enable a reinforcement learning agent to acquire symbolical knowledge, we augment it with a high-level knowledge representation. This representation consists of ordinal conditional functions (OCF) which allow it to rank world models. By this means the agent is enabled to complement the self-organizing capabilities of the low-level reinforcement learning sub-system by reasoning capabilities of a high-level learning component. We briefly summarize the state-of-the-art method how new information is included into the OCF. To improve the emergence of plausible behavior, we then introduce a modification of this method. The viability of this modification is examined first, for the inclusion of conditional information with negated consequents and second, for the generalization of belief in the context of unobserved variables. Besides providing a theoretical justification for this modification, we also show the advantages of our approach in comparison to the state-of-the-art method of revision in a reinforcement learning application.

1 INTRODUCTION

The creation of a system with autonomous learning capabilities creates a variety of challenges. Such a system (or "agent") has to figure out which actions are beneficial and which have to be avoided. Starting with three system requirements we developed the work described in this paper. These requirements are the following. First, an autonomous learning system should be able to learn from experience. A widely adopted approach to incorporate such a property is given by reinforcement learning (RL) (Sutton and Barto, 1998). We will use a basic Q-learning scheme to model this. Since RL is not the primary topic of this work, we describe the basic idea in a nutshell only. Second, the system should generate a representation of its knowledge that allows further reasoning. In this area belief revision (BR) techniques can be found. We will examine the usefulness of ordinal conditional functions (OCF) (Kern-Isberner, 2001; Spohn, 2009) in this work. Third, and most important, we want both mentioned approaches to benefit from each other. A mixture of low-level learning-by-doing and high-level deduction abilities is called a two-level learning approach. Psychological findings (Anderson, 1983; Gombert, 2003; Reber, 1989; Sun et al., 2005) indicate that such two-level learning principles can explain some of the human learning abilities. While humans are able to learn top-down or bottom-up (Sun et al., 2006), we will focus on the bottom-up part only. A combination of RL and BR has been proposed before (Leopold et al., 2008), influenced by (Sun et al., 2001) and (Ye et al., 2003). While we have already described the general idea of our approach in (Häming and Peters, 2010), we present here the detailed formalism and give a theoretical justification. This work is also related to the topic of relational reinforcement learning (RRL) (Dzeroski et al., 2001). However, RLL does not distinguish between high-level and low-level knowledge. It represents the Q-function directly in the form of propositional clauses.

2 NOTATION AND TERMINOLOGY

A variable a can represent a value from its domain $\mathfrak{D}_a$. Such a domain consists of discrete values. One such realization of a variable is called a *literal*. We write literals by denoting the variable as a subscript of its value (e.g., 3_a or $\mathtt{t}_a$). A *formula* consists of literals and logical operators such as $\wedge$, $\vee$, $\Rightarrow$, etc. It is

Häming K. and Peters G..
IMPROVED REVISION OF RANKING FUNCTIONS FOR THE GENERALIZATION OF BELIEF IN THE CONTEXT OF UNOBSERVED VARIABLES.
DOI: 10.5220/0003669501180123
In *Proceedings of the International Conference on Neural Computation Theory and Applications* (NCTA-2011), pages 118-123
ISBN: 978-989-8425-84-3

referred to by an uppercase letter, e.g., $A := 0_a \vee 1_b$. A *negation* of a literal refers to a formula. For example, if $\mathfrak{D}_a := \{1,2,3\}$, then

$$\overline{2_a} = (1_a \vee 3_a)\,. \tag{1}$$

The set of all variables is $\mathfrak{V}$, while the set of variables that are realized in a formula A is denoted by $\mathfrak{V}_A$.

A *model* is a conjunction in which exactly one literal exists for each variable. The set of all models is referred to as $\mathfrak{M}$. If we restrict the set of variables the models are derived from, we will write the variable set as a subscript, e.g. $\mathfrak{M}_{\mathfrak{V}}$. A model M is said to be *a model of a formula* F, if F is true for the literals in M. We denote this as $M \models F$. If an agent believes a formula A, which means A can be inferred from its knowledge base κ, we will write $\kappa \models A$.

A conditional is denoted by $A \Rightarrow B$, where A is the antecedent and B is the consequent. The set of conditionals we obtain when the antecedent A is replaced by a set of formulas $\mathfrak{F}$, is referred to as $\{\mathfrak{F} \Rightarrow B\} := \{F \Rightarrow B | F \in \mathfrak{F}\}$.

3 REINFORCEMENT LEARNING AND BELIEF REVISION

Let us assume an *environment* that is described by a set of states. State transitions are performed depending on the current state and the current *action* carried out by the agent. The transitions are rewarded. A goal of RL consists in the identification of beneficial actions, i.e., those actions that produce high rewards.

So, we have a set of states $\mathfrak{S}$, a set of actions $\mathfrak{A}$, a transition function $\delta : \mathfrak{S} \times \mathfrak{A} \to \mathfrak{S}$, and a reward function $r : \mathfrak{S} \times \mathfrak{A} \to \mathbb{R}$. Knowledge about good and poor actions is learned by applying a learning technique. In our approach we apply Q-learning. This technique has the convenient property of being *policy*-free. This means that the result does not depend on the chosen strategy with which the agent explores the environment.

The agent's experience is captured in the Q(uality)-function that assigns an expected reward to each state-action-pair. The Q-function is updated after each state transition in the following way:

$$Q(S,A) = r + \gamma \max_{A'} Q(S',A') \tag{2}$$

with

$$S' := \delta(S,A')\,. \tag{3}$$

One can interpret this formula in the way that the agent will *believe* an action A to be a best action, if it has the highest $Q(S,A)$ value for a given state S. This is the point where we establish a connection to the high-level knowledge using BR in the following. BR is a theory of maintaining a knowledge base in such a way that the current belief is represented in a consistent manner (Alchourron et al., 1985; Darwiche and Pearl, 1996). We model our knowledge base κ as an ordinal conditional function (OCF). This is a ranking function that maintains a list of all models. The models the agent believes in are set to rank 0, while all ranks greater than 0 represent an increasing disbelief. We denote the rank an OCF κ assigns to a model M as $\kappa(M)$. By convention, contradictions shall have the rank ∞. The operator "$\models$" of Section 2 is defined for an OCF as

$$\begin{aligned}\kappa \models A :\Leftrightarrow (&\exists M_1, M_1 \models A : \kappa(M_1) = 0)\\ &\wedge (\forall M_2, M_2 \models \overline{A} : \kappa(M_2) > 0)\end{aligned} \tag{4}$$

which requires a believed formula to have a model with rank 0 and its negation to have a rank greater than 0.

In this work, the states and actions are described as formulas. Therefore it is possible to store information on them in a suitable OCF. The interplay between the OCF and the Q-function is described in Section 8.

4 STATE-OF-THE-ART REVISION OF ORDINAL CONDITIONAL FUNCTIONS

The current belief represented by the OCF consists of models, i.e., propositional information in the form of conjunctions. However, during exploration the information gathered and the information needed is in the form of conditionals. To check, whether an OCF believes a conditional the agent can temporarily believe its antecedent (known as *conditioning*) and check if the conjunction of the antecedent and the consequent is also believed. At the same time, the conjunction of the antecedent and the negation of the consequent must not be believed (that is, $\kappa(S\overline{A}) > 0$). Generally, we do not have to condition κ to find out whether a conditional is believed. It is sufficient to compute the belief ranks $r_1 = \kappa(SA)$ and $r_2 = \kappa(S\overline{A})$. If $r_1 < r_2$, the conditional will be believed. More difficult than querying the knowledge base is its update, called *revision*. The revision operator is "$*$". Conditionals in BR are usually denoted by $(A|S)$, where S is the antecedent and A the consequent. The meaning of $(A|S)$ is not exactly the same as $S \Rightarrow A$ (Kern-Isberner, 2001). The latter means that S implicates A irrespective of the values other variables. In contrast, $(A|S)$ expresses that A will be believed if κ is conditioned

with S and S alone, therefore a revision $(\kappa*(ST))$ may not result in A being believed. In our context of RL, if S is a complete state description, it will capture all the available information. Then, an expression such as $ST, T \neq S$ is necessarily a contradiction and therefore not believed. In this case, the meaning of $S \Rightarrow A$ and $(A|S)$ is the same. Therefore, on a first attempt, we use $(\kappa*(A|S))$ to revise κ with a conditional analogous to (Leopold et al., 2008). Then, we will examine the consequences of such a decision.

After a revision of κ with the conditional $S \Rightarrow A$, we want

$$(\kappa*(A|S))(SA) < (\kappa*(A|S))(S\overline{A}) \tag{5}$$

to hold. If this is already the case, nothing has to be done. Otherwise the following holds:

Theorem 1. *If $\kappa(SA) \geq \kappa(S\overline{A})$, then the OCF κ' derived from κ by rearranging the models using*

$$\forall M \in \mathfrak{M} : \kappa'(M) := (\kappa*(A|S))(M) = \begin{cases} \kappa(M) - \kappa(S \Rightarrow A) & : M \models S \Rightarrow A \\ a + b & : M \models S\overline{A} \end{cases} \tag{6}$$

with

$$a = \kappa(SA) - \kappa(S \Rightarrow A) + 1$$
$$b = \kappa(M) - \kappa(S\overline{A})$$

will result in $\kappa'(SA) < \kappa'(S\overline{A})$. Consequently, κ' expresses the belief in $S \Rightarrow A$.

Proof. Let us partition the models in κ into three disjoint sets:

$$\mathfrak{M}_1 = \{M | M \models \overline{S}\},$$
$$\mathfrak{M}_2 = \{M | M \models SA\}, \text{ and}$$
$$\mathfrak{M}_3 = \{M | M \models S\overline{A}\}.$$

We address the first rule of Equation 6 first. The purpose of it is to let $\kappa'(S \Rightarrow A) = 0$. The models in $\mathfrak{M}_1 \cup \mathfrak{M}_2$ are those that model $S \Rightarrow A$. Therefore we reduce in rank all models in $\mathfrak{M}_1 \cup \mathfrak{M}_2$ by $\kappa(S \Rightarrow A)$ which is the rank of the lowest ranked model in $\mathfrak{M}_1 \cup \mathfrak{M}_2$. Hence, $\kappa'(S \Rightarrow A) = 0$. We now consider term a of the second rule. We want $\kappa'(SA) < \kappa'(S\overline{A})$ to hold. That means, after revision, the lowest rank of the models in $\mathfrak{M}_3$ needs to be at least $\kappa'(SA) + 1$. Since the models of SA are found in $\mathfrak{M}_2$ and are therefore shifted by the first rule, $\kappa'(SA) = \kappa(SA) - \kappa(S \Rightarrow A)$. Adding 1 is arbitrary but sufficient to meet the requirements. Term a alone would shift the ranks of all models of $\mathfrak{M}_3$ to the rank $\kappa'(SA) + 1$. To preserve the relative ranking of the models, we need to add term b to the second rule. Since $\kappa(S\overline{A})$ is the rank of the lowest ranked model of $\mathfrak{M}_3$, this very model is still shifted to the rank $\kappa'(SA) + 1$. The other models, however, now keep their distance. □

5 NEGATED CONSEQUENTS

What will happen if κ is revised with $S \Rightarrow \overline{A}$? Then, an application of Equation 6 will result in $(\kappa*(\overline{A}|S))(S\overline{A}) < (\kappa*(\overline{A}|S))(SA)$.

This does not mean that all models of $S\overline{A}$ have a rank lower than $\kappa(SA)$. We show this in the following example. Let us define two variables a and b with their domains $\mathfrak{D}_a := \{1,2\}$ and $\mathfrak{D}_b := \{1,2,3\}$. The current belief is represented by an OCF, where the first entry represents the current belief; that means its model has rank 0. Now, we want the following OCF κ_{neg} to belief $1_a \Rightarrow \overline{1_b}$:

$$\kappa_{neg} = \left\| \begin{matrix} 2\,1 \\ \overline{1\,1} \\ 2\,2 \\ 1\,2 \\ 2\,3 \\ 1\,3 \end{matrix} \right\| \xrightarrow{(\kappa_{neg}*(\overline{1_b}|1_a))} \kappa'_{neg} = \left\| \begin{matrix} 2\,1 \\ \overline{2\,2} \\ 1\,2 \\ 1\,1\;2\,3 \\ 1\,3 \end{matrix} \right\| \tag{7}$$

which beliefs $1_a \Rightarrow 2_b$, but not $1_a \Rightarrow 3_b$. This behavior is perfectly sane since $(1_a \wedge 2_b) \wedge (1_a \wedge 3_b)$ is a contradiction. But the belief in $(1_a \wedge 1_b)$ is stronger than the belief in $(1_a \wedge 3_b)$. If we revise κ with $1_a \Rightarrow \overline{2_b}$, then the result will be

$$\kappa''_{neg} = (\kappa'_{neg} * (\overline{2_b}|1_a)) = \left\| \begin{matrix} 2\,1 \\ \overline{2\,2} \\ 1\,1\;2\,3 \\ 1\,3\;1\,2 \end{matrix} \right\|. \tag{8}$$

This expresses a belief in $1_a \Rightarrow 1_b$ which is certainly not what we expect an agent to believe if it has just been exposed to the information $1_a \Rightarrow \overline{1_b}$ and $1_a \Rightarrow \overline{2_b}$. Instead, a belief in $1_a \Rightarrow 3_b$ seems reasonable.

6 GENERALIZATION

We examine a revision by Equation 6 in the context of generalization by examining what effect the omission of variables in a formula has. Let us partition the set of variables $\mathfrak{V}$ into three *non-empty* subsets:

$$\mathfrak{V} = \mathfrak{X} \cup \mathfrak{Y} \cup \mathfrak{Z}, \text{ with}$$
$$\mathfrak{X} \cap \mathfrak{Y} = \emptyset, \mathfrak{X} \cap \mathfrak{Z} = \emptyset, \text{ and } \mathfrak{Y} \cap \mathfrak{Z} = \emptyset \tag{9}$$

Next, take a model from each of the subsets, such as

$$X \in \mathfrak{M}_{\mathfrak{X}}, Y \in \mathfrak{M}_{\mathfrak{Y}}, \text{ and } Z \in \mathfrak{M}_{\mathfrak{Z}}. \tag{10}$$

The revision $\kappa*(Z|X)$ will lead to a knowledge base that believes a particular model M' of $\{\mathfrak{M}_{\mathfrak{X}} \Rightarrow Z\} \subset \{\mathfrak{M}_{\mathfrak{X}\cup\mathfrak{Y}} \Rightarrow Z\}$.

Next, we consider the other models $\mathfrak{C} := \{\mathfrak{M}_{\mathfrak{X}\cup\mathfrak{Y}} \Rightarrow Z\} \setminus M'$. First, there is the obvious restriction that $C \in \mathfrak{C}$ is not allowed to contradict Z. We

already ruled this out in Equation 9. Let us look at the following sample OCFs:

$$\kappa_{gen} = \left\| \begin{matrix} 212 \\ \overline{211} \\ 221 \\ 222 \end{matrix} \right\| \text{ and } \quad \kappa_{\overline{gen}} = \left\| \begin{matrix} 211 \\ \overline{222} \\ 212 \\ 221 \end{matrix} \right\| \tag{11}$$

We can easily see that κ_{gen} believes $2_a \Rightarrow 2_c$ since $\kappa_{gen}(2_a \Rightarrow 2_c) = 0$, but at the same time $\kappa_{gen}((2_a \wedge 2_b) \Rightarrow 2_c) = 1 > 0$. A revision with $2_a \Rightarrow 2_c$ using Equation 6 would not change κ_{gen} at all.

The same issue occurs considering a revision with a conditional that has a negated consequent, such as $2_a \Rightarrow \overline{2_c}$. We show this for $\kappa_{\overline{gen}}$ which believes this conditional and would not be changed by a revision with $2_a \Rightarrow \overline{2_c}$ using Equation 6. Nevertheless is does not believe $(2_a \wedge 2_b) \Rightarrow \overline{2_c}$. We conclude that Equation 6 does not produce an OCF that is capable of generalization.

7 AN ALTERNATIVE REVISION

Because of the described drawbacks we suggest an alternative revision technique. The proposed revision introduced in this section, $(\kappa * (S \Rightarrow A))$, utilizes a new operator $\kappa[A]$ which returns the highest disbelief among all models of A. After a revision of κ with the conditional $S \Rightarrow A$, we still want the equivalent of Equation 5 to hold:

$$(\kappa * (S \Rightarrow A))(SA) < (\kappa * (S \Rightarrow A))(S\overline{A}) \tag{12}$$

This is investigated in the following

Theorem 2. *If $\kappa(SA) \geq \kappa(S\overline{A})$, then the OCF κ' derived from κ by rearranging the models using*

$$\forall M \in \mathfrak{M} : \kappa'(M) := (\kappa * (S \Rightarrow A))(M) = \begin{cases} \kappa(M) - \kappa(S \Rightarrow A) & : M \models S \Rightarrow A \\ a' + b' & : M \models S\overline{A} \end{cases} \tag{13}$$

with

$$a' = \kappa[SA] - \kappa(S \Rightarrow A) + 1$$
$$b' = \kappa(M) - \kappa(S\overline{A})$$

will result in $\kappa'(SA) < \kappa'(S\overline{A})$. Consequently, κ' expresses the belief in $S \Rightarrow A$.

Proof. Let $\kappa_1 := (\kappa * (A|S))$ and $\kappa_2 := (\kappa * (S \Rightarrow A))$. Since $\kappa[A] \geq \kappa(A)$, by application of Theorem 1 can be deduced that $\kappa_2(SA) = \kappa_1(SA) < \kappa_1(S\overline{A}) \leq \kappa_2(S\overline{A})$. □

So, concerning the preservation of current belief, this method works just as good as Equation 6, but introduces greater changes. In the following discussion of the properties of $(\kappa * (S \Rightarrow A))$ with respect to negation and generalization we justify these changes. First, we consider negation.

Theorem 3. *Let $\mathtt{t} \in \mathfrak{D}_b$ and $\kappa' := (\kappa * (A \Rightarrow \overline{\mathtt{t}_b}))$. Then*

$$\forall \mathtt{r} \in \mathfrak{D}_b \setminus \mathtt{t} : \kappa'(A \Rightarrow \mathtt{r}_b) \leq \kappa'(A \Rightarrow \mathtt{t}_b). \tag{14}$$

Proof. By applying Equation 13, we obtain $\kappa'(A\mathtt{t}_b) > \kappa'[A\overline{\mathtt{t}_b}]$. This is equivalent to

$$\forall \mathtt{r} \in \mathfrak{D}_b \setminus \mathtt{t} : \kappa'(A\mathtt{r}_b) < \kappa'(A\mathtt{t}_b).$$

Hence, if A is believed, the inequality of Equation 14 will hold strictly. On the other hand, if $\overline{A}$ is believed, then $\kappa'(A \Rightarrow \mathtt{r}_b) = 0$ as well as $\kappa'(A \Rightarrow \mathtt{t}_b) = 0$. □

Theorem 3 induces that the observed inconsistency described in Section 5 does not appear after the repeated application of Equation 13. Indeed, a revision of κ_{neg} with $(1_a \Rightarrow \overline{1_b})$ now results in

$$\kappa'_{neg} = (\kappa_{neg} * (1_a \Rightarrow \overline{1_b})) = \left\| \begin{matrix} 21 \\ \overline{22} \\ 12 \\ 23 \\ 13 \\ 11 \end{matrix} \right\|. \tag{15}$$

Also,

$$\kappa''_{neg} = (\kappa'_{neg} * (1_a \Rightarrow \overline{2_b})) = \left\| \begin{matrix} 21 \\ \overline{22} \\ 23 \\ 13 \\ 11 \\ 12 \end{matrix} \right\| \tag{16}$$

which illustrates that $1_a \Rightarrow 3_b$ is now believed as expected.

We now consider generalization with our alternative revision technique. Again, Equation 9 and Equation 10 are given.

Theorem 4. *Let $X \in \mathfrak{M}_X$, $Z \in \mathfrak{M}_Z$, and $\kappa' := (\kappa * (X \Rightarrow Z))$. Then*

$$\forall Y \in \mathfrak{M}_Y : \kappa'(X \wedge Y \Rightarrow Z) \leq \kappa'(X \wedge Y \Rightarrow \overline{Z}). \tag{17}$$

Proof. The proof is an analog of the proof of Theorem 3. After applying Equation 13, we obtain $\kappa'(X\overline{Z}) > \kappa'[XZ]$. This is equivalent to

$$\forall Y \in \mathfrak{M}_Y : \kappa'(X \wedge Y \wedge Z) < \kappa'(X \wedge Y \wedge \overline{Z})$$

Hence, if $X \wedge Y$ is believed, the inequality of Equation 17 will hold strictly. On the other hand, if $\overline{X \wedge Y}$ is believed, then $\kappa'(X \wedge Y \Rightarrow Z) = 0$ as well as $\kappa'(X \wedge Y \Rightarrow \overline{Z}) = 0$. □

A similar theorem will hold, if the consequent is negated.

To complete this section, we show that the previous counter-examples can be resolved using Equation 13. A revision of κ_{gen} with $2_a \Rightarrow 2_c$ now yields

$$\kappa_{gen} = \left\| \begin{matrix} 212 \\ \overline{211} \\ 221 \\ 222 \end{matrix} \right\| \xrightarrow{(\kappa_{gen} * (2_a \Rightarrow 2_c))} \kappa'_{gen} = \left\| \begin{matrix} 212 \\ \overline{222} \\ 211 \\ 221 \end{matrix} \right\|. \quad (18)$$

This κ'_{gen} expresses a belief in $2_a \wedge 1_b \Rightarrow 2_c$ and $2_a \wedge 2_b \Rightarrow 2_c$.

A revision of $\kappa_{\overline{gen}}$ with $2_a \Rightarrow \overline{2_c}$ provides

$$\kappa_{\overline{gen}} = \left\| \begin{matrix} 211 \\ \overline{222} \\ 212 \\ 221 \end{matrix} \right\| \xrightarrow{(\kappa_{\overline{gen}} * (2_a \Rightarrow \overline{2_c}))} \kappa'_{\overline{gen}} = \left\| \begin{matrix} 211 \\ \overline{221} \\ 222 \\ 212 \end{matrix} \right\| \quad (19)$$

which now believes $2_a \wedge 1_b \Rightarrow \overline{2_c}$ as well as $2_a \wedge 2_b \Rightarrow \overline{2_c}$.

8 APPLICATION

We examine the effect of the proposed algorithm in a cliff-walk gridworld (Sutton and Barto, 1998) (Figure 2). For this application six cases are examined: plain Q-learning, OCF-augmented Q-learning with application of Equation 6, OCF-augmented Q-learning with application of Equation 13, plain Q-learning with futile information, OCF-augmented Q-learning with application of Equation 6 and futile information, and OCF-augmented Q-learning with application of Equation 13 and futile information. An OCF-augmented Q-learner is a Q-learner that has conditionals extracted from its Q-table. These conditionals revise the learner's OCF and this OCF acts as a filter for the choice of actions afterwards. Figure 1 shows this architecture.

We add futile information to model the case where the agent perceives properties of its environment that are not helpful with regard to its goal. The OCF-augmented Q-learners are expected to be able to generalize and therefore identify the futile information. The generalization is performed in the same way as in (Leopold et al., 2008) by counting the pattern frequency. The general idea is to keep track of how often sub-patterns of antecedents are used in the context of particular consequents. If they occur frequently enough, we will revise the OCF with the sub-pattern instead of the complete state description. The state description is also adopted from (Leopold et al., 2008), where a qualitative description is used which consists of the relative position of the agent towards the goal (north, south, east, west) and a distance (near, middle, far) amended with information on adjacent obstacles. Reaching the goal triggers a reward of 100, getting closer towards the goal is rewarded by 0.5. Stepping into the cliff is punished by -10, every other step receives -1. After 100 steps the episode is forced to end.

The results are depicted in Figure 3. It is evident that a revision with Equation 13 clearly outperforms a revision with Equation 6. The latter is worse than a plain Q-learner and even seems to deteriorate over time. An explanation for this may lie in the fact that the OCF gets contaminated by harmful conditionals. However, this has not been examined in this work.

The computational cost of the described improvement depends on the representation of the OCF. If the OCF is implemented by creating every possible conjunction beforehand, then Equation 6 and Equation 13 will lead to roughly the same runtime, because $\kappa[A]$ is the rank of A in a reversed κ.

In a different approach we initialized the OCF without any conjunctions to be able to handle larger problems. Conjunctions not occurring in the OCF received a rank of infinity. Then the revision process generates conjunctions as needed. Clearly, this breaks the symmetry between $\kappa(A)$ and $\kappa[A]$. The runtime of this approach is about 1.5 times larger than the runtime of the previously described approach.

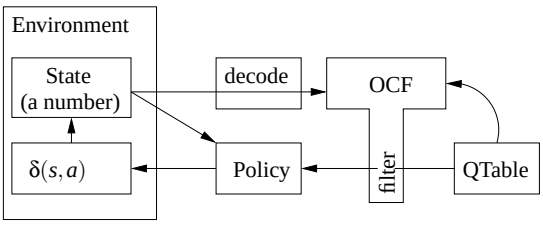

Figure 1: OCF-augmented RL system. The OCF acts as a filter that limits the choices of the policy.

9 CONCLUSIONS

The theoretical considerations presented in this work alleviate a severe disadvantage of the to-date revision of ordinal conditional functions with conditional information. The presented examples clearly indicate the removal of quite apparent implausibilities. The aptitude of this approach to create an agent which shows emergent understanding of its environment

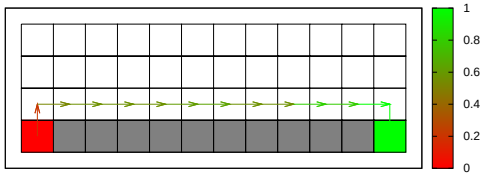

Figure 2: Cliff-walk gridworld. The goal of a moving agent is to reach the green square, starting from the red one. Entering the dark squares (representing a cliff) results in a high negative reward. Superimposed is the learned path after 100 episodes. The path color indicates the expected reward by displaying the value of $\min(1, \frac{\text{expected reward}}{\text{goal reward}})$ using the displayed color key.

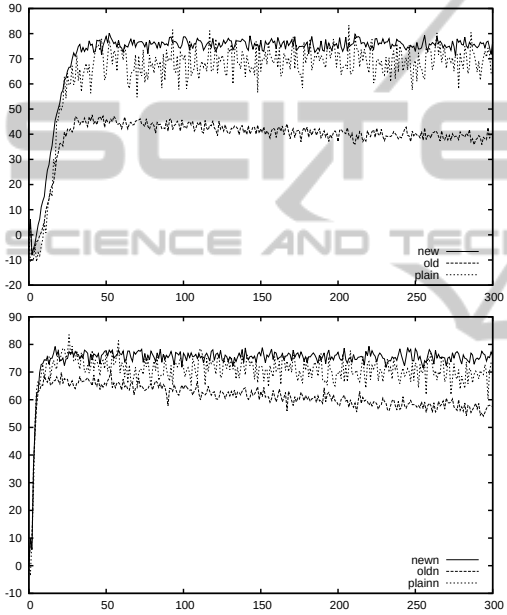

Figure 3: Results. The diagrams show the rewards over a series of 300 episodes. On the top the results with futile information are depticted, on the bottom the results without futile information. *plain*/*plainn* show results of a plain Q-learner, *old*/*oldn* show results of revisions with Equation 6, and *new*/*newn* show results of revisions with Equation 13. For the OCF-augmented learners the values are averages of 1000 runs. Since the plain Q-learner exhibits large variations, its values have been averaged over 2000 runs.

needs to be examined in more detail. Especially the analysis of the symbolic belief representation in different contexts is certainly on our agenda. First experiments indicate the accumulation of a proper symbolic description of favorable state-action-pairs.

The symbolic representation also allows for symbolic reasoning to incorporate a top-down path of learning. The combination of theses techniques is definitely of interest and needs to be addressed in future publications.

ACKNOWLEDGEMENTS

This research was funded by the German Research Association (DFG) under Grant PE 887/3-3.

REFERENCES

Alchourron, C. E., Gardenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *J. Symbolic Logic*, 50(2):510–530.

Anderson, J. R. (1983). *The architecture of cognition*. Hardvard University Press, Cambridge, MA.

Darwiche, A. and Pearl, J. (1996). On the logic of iterated belief revision. *Artificial intelligence*, 89:1–29.

Dzeroski, S., Raedt, L. D., and Driessens, K. (2001). Relational reinforcement learning. In *Machine Learning*, volume 43, pages 7–52.

Gombert, J.-E. (2003). Implicit and explicit learning to read: Implication as for subtypes of dyslexia. *Current psychology letters*, 1(10).

Häming, K. and Peters, G. (2010). An alternative approach to the revision of ordinal conditional functions in the context of multi-valued logic. In *Proceedings of the 20th International Conference on Artificial Neural Networks*, volume LNCS 6353, pages 200–203.

Kern-Isberner, G. (2001). *Conditionals in nonmonotonic reasoning and belief revision: considering conditionals as agents*. Springer-Verlag New York, Inc.

Leopold, T., Kern Isberner, G., and Peters, G. (2008). Combining reinforcement learning and belief revision: A learning system for active vision. In *Proceedings of the 19th British Machine Vision Conference*, pages 473–482.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psycology: General*, 3(118):219–235.

Spohn, W. (2009). A survey of ranking theory. In *Degrees of Belief*. Springer.

Sun, R., Merrill, E., and Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. In *Cognitive Science*, volume 25(2), pages 203–244.

Sun, R., Terry, C., and Slusarz, P. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, 112:159–192.

Sun, R., Zhang, X., Slusarz, P., and Mathews, R. (2006). The interaction of implicit learning, explicit hypothesis testing, and implicit-to-explicit knowledge extraction. *Neural Networks*, 1(20):34–47.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge.

Ye, C., Yung, N. H. C., and Wang, D. (2003). A fuzzy controller with supervised learning assisted reinforcement learning algorithm for obstacle avoidance. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 33(1):17–27.