Grounding Possible Worlds Semantics in Experiential Semantics

Matthew Iklé
Adams State College
Ben Goertzel
Novamente LLC

Abstract

Probabilistic Logic Networks (PLN), a comprehensive framework for uncertain inference currently in use in the OpenCog and Novamente Cognition Engine AGI software architectures, has previously been described in terms of the "experiential semantics" of an intelligent agent embodied in a world. However, several aspects of PLN are more easily interpreted and formulated in terms of "possible worlds semantics"; here we use a formal model of intelligent agents to show how a form of possible worlds semantics can be derived from experiential semantics, and use this to provide new interpretations of several aspects of PLN (including uncertain quantifiers, intensional inheritance, and indefinite probabilities.) These new interpretations have practical as well as conceptual benefits, as they give a unified way of specifying parameters that in the previous interpretations of PLN were viewed as unrelated.

Introduction

The mind of an intelligent agent accumulates knowledge based on experience, yet also creates hypothetical knowledge about "the world as it might be," which is useful for guiding future actions. This dichotomy – between experience and hypothesis – occurs in regard to many types of knowledge; and in the context of declarative knowledge, it is related to the distinction between experiential and possible-worlds semantics. Here we discuss how these two forms of semantics may be related to each other in the context of a generally intelligent agent that interprets its experience (at least partially) using probabilistic logic. Our treatment pertains specifically to the Probabilistic Logic Networks (PLN) inference framework, which is currently in use in the OpenCog and Novamente Cognition Engine AGI software architectures and has been used for applications including natural language based reasoning (Gea06) and virtual agent reinforcement learning (Goe08); however, many of the points raised could be extended more generally to any probabilistic inference framework.

In much of our prior work on PLN, we have utilized "experiential semantics", according to which the meaning of each logical statement in an agent's memory is defined in terms of the agent's experiences. However we have also found that certain aspects of PLN are best interpreted in terms of "possible worlds semantics", in which the meaning of a statement is defined by reference to an ensemble of possible worlds including the one the agent interpreting the statement has experienced. The relation between these two semantic approaches in the PLN context has previously been left informal; the core goal of this paper is to specify it, via providing an *experiential grounding of possible worlds semantics.*

We study an agent whose experience constitutes one "actual world" drawn from an ensemble of possible worlds. We use the idea of bootstrapping from statistics to generate a set of "simulated possible worlds" from the actual world, and prove theorems regarding conditions under which, for a probabilistic predicate F, the truth value of F evaluated over these simulated possible worlds gives a good estimate of the truth value of F evaluated over the ensemble of possible worlds from which the agent's actual world is drawn.

The reader with a logic background should note that we are construing the notion of possible worlds semantics broadly here, in the philosophical sense (Lew86), rather than narrowly in the sense of Kripke semantics (Gol03) and its relatives. In fact there are interesting mathematical connections between the present formulation and Kripke semantics and epistemic logic, but we will leave these for sequel papers.

Then, we show how this apparatus of simulated possible worlds simplifies the interpretation of several aspects of PLN, providing a common foundation for setting various PLN system parameters that were previously viewed as distinct. We begin with indefinite probabilities (Iea07; Gea08), noting that the second-order distribution involved therein may be interpreted using possible worlds semantics. Then we turn to uncertain quantifiers, showing that the third-order distribution used to interpret these in (IG08) may be considered as a distribution over possible worlds. Finally, we consider intensional inference, suggesting that the complexity measure involved in the definition of PLN intension (Gea08) may be derived from a probability measure over possible worlds. By considering the space of possible worlds implicit in an agent's experience, one arrives at a simpler unified view of various aspects of the agent's uncertain reasoning, than if one grounds these aspects in the agent's experience directly. This is not an abandonment of experiential semantics but rather an acknowledgement that a simple variety of possible worlds semantics is derivable from experiential semantics, and usefully deployable in the development of uncertain inference systems for general intelligence.

We will not review the PLN inference framework here but will assume the reader has a basic familiarity with PLN terms and notation, as would be found by reading (Gea08) or (Iea07).

A Formal Model of Intelligent Agents

We very briefly review a simple formal model of intelligent agents: Simple Realistic Agent Model (SRAM). Following Legg and Hutter's framework (LH07), we consider a class of active agents which observe and explore their environment and take actions in it. The agent sends information to the environment by sending symbols from some finite alphabet called the *action space* Σ; and the environment sends signals to the agent with symbols from an alphabet called the *perception space*, denoted $\mathcal{P}$. Agents can also experience rewards, which lie in the *reward space*, denoted $\mathcal{R}$, which for each agent is a subset of the rational unit interval.

To Legg and Hutter's framework, we add a set $\mathcal{M}$ of memory actions, allowing agents to maintain memories (of finite size), and at each time step to carry out internal actions on their memories as well as external actions in the environment. Further extending the Legg and Hutter framework, we also introduce the notions of *goals* associated with symbols, drawn from the alphabet $\mathcal{G}$, and *goal-seeking agents*; and we consider the environment as sending goal-symbols to the agent along with regular observation-symbols. We also introduce a conditional distribution $\gamma(g, \mu)$ that gives the weight of a goal g in the context of a particular environment μ.

In this extended framework, an interaction sequence looks like

$$m_1 a_1 o_1 g_1 r_1 m_2 a_2 o_2 g_2 r_2 ...$$

where the m_i's represent memory actions, the a_i's represent external actions, the o_i's represent observations, the g_i's represent agent goals, and the r_i's represent rewards. The reward r_i provided to an agent at time i is determined by the goal function g_i. Introducing w as a single symbol denoting the combination of a memory action and an external action, and y as a single symbol denoting the combination of an observation, a goal and a reward, we can simplify this interaction sequence as

$$w_1 y_1 w_2 y_2 ...$$

Each goal function maps each finite interaction sequence $I_{g,s,t} = wy_{s:t}$ into a value $r_g(I_{g,s,t}) \in [0, 1]$ indicating the value or "raw reward" of achieving the goal during that interaction sequence. The total reward r_t obtained by the agent is the sum of the raw rewards obtained at time t from all goals whose symbols occur in the agent's history before t.

The agent is represented as a function π which takes the current history as input, and produces an action as output. Agents need not be deterministic and may induce a probability distribution over the space of possible actions, conditioned on the current history. In this case we may characterize the agent by a probability distribution $\pi(w_t|wy_{<t})$. Similarly, the environment may be characterized by a probability distribution $\mu(y_k|wy_{<k})$. The distributions π and μ define a probability measure over the space of interaction sequences.

Following Legg and Hutter, we will consider the class of environments that are *reward-summable*, meaning that the total amount of reward they return to any agent is bounded by 1. We will also use the term "context" to denote the combination of an environment, a goal function and a reward function. If the agent is acting in environment μ, and is provided with $g_t = g$ for the time-interval $T = t \in \{t_1, ..., t_2\}$, then the *expected goal-achievement* of the agent during the interval is

$$V^{\pi}_{\mu,g,T} \equiv \sum_{t_1}^{t_2} r_i$$

where E is the space of computable, reward-summable environments.

Next, we introduce a second-order probability distribution ν over the space of environments μ. One such distribution is the Solomonoff-Levin universal distribution in which one sets $\nu = 2^{-K(\mu)}$; but this is not the only distribution of interest. A great deal of real-world general intelligence consists of the adaptation of intelligent systems to other particular distributions ν over environment-space (Goe10; Goe09).

Inducing a Distribution over Predicates and Concepts

Given a distribution over environments as defined above, and a collection of predicates evaluated on subsets of environments, we will find it useful to define distributions (induced by the distribution over environments) defining the probabilities of these predicates.

Suppose we have a pair (F, T) where F is a function mapping sequences of perceptions into fuzzy truth values, and T is an integer connoting a length of time. We can define the prior probability of (F, T) as the average degree to which F is true, over a random interval of perceptions of length T drawn from a random environment drawn from the distribution over environments. More generally, if one has a pair (F, f), where f is a distribution over the integers, one can define the prior probability of (F, f) as the weighted average of the prior probability of (F, T) where T is drawn from f.

While expressed in terms of predicates, the above formulation can also be useful for dealing with concepts, e.g. by interpreting the concept *cat* in terms of the predicate *isCat*. So we can use this formulation in inferences where one needs a concept probability like $P(cat)$ or a relationship probability like $P(eat(cat, mouse))$.

Grounding Possible Worlds Semantics in Experiential Semantics

Now we explain how to ground a form of possible worlds semantics in experiential semantics. We explain how an agent, experiencing a single stream of perceptions, may use this to construct an ensemble of "simulated" possible worlds, which may then be used in various sorts of inferences. This idea is closely related to a commonplace idea in the field of statistics: "subsampling," a form of "bootstrapping."

In subsampling, if one has a single dataset D which one wishes to interpret as coming from a larger population of possible datasets, and one wishes to approximately understand the distribution of this larger population, then one generates a set of additional datasets via removing various portions of D. By removing a portion of D, one obtains another dataset. One can then look at the distribution of these auxiliary datasets, considering it as a model of the population D.

This notion ties in closely with SRAM, which considers a probability distribution over a space of environments which are themselves probability distributions. A real agent has a single series of remembered observations. It can induce an approximation of this distribution over environments by subsampling its memory and asking: what would it imply about the world if the items in this subsample were the only things I'd seen?

It may be conceptually useful to observe that a related notion to subsampling is found in the literary methodology of science fiction. Many SF authors have followed the methodology of changing one significant aspect of our everyday world, and depicting the world as they think it might exist if this one aspect were changed (or, a similar methodology may be followed via changing a small number of aspects). This is a way of generating a large variety of alternate possible worlds from the raw material of our own world.

The subsampling and SF analogies suggest two methods of creating a possible world within SRAM (and by repetition, an ensemble of possible worlds) from the agents experience. An agent's interaction sequence with its environment forms a sample from which it wishes to infer its environment. To better assess this environment, the agent may, for example,

1. create a possible world by removing a randomly selected collection of interactions from the agents memory. In this case, the agent's interaction sequence would be of the form $I_{g,\ s,\ t,(n_t)} = wy_{(n_t)}$ where (n_t) is some subsequence of $1 : t-1$.
2. create a possible world via assuming a counterfactual hypothesis (i.e. assigning a statement a truth value that contradicts the agents experience), and using inference to construct a set of observations that is as similar to its memory as possible, subject to the constraint of being consistent with the hypothesis.
3. create a possible world by reorganizing portions of the interaction sequence.
4. create a possible world by some combination of the above.

Here we will focus on the first option, leaving the others for future work. We denote an alteration of an iteration sequence $I^a_{g,s,t}$ for an agent a by $\tilde{I}^a_{g,s,t}$, and the set of all such altered interaction sequences for agent a by $\mathcal{I}^a$.

An agent's interaction sequence will presumably be some reasonably likely sequence. We would therefore be most interested in those cases for which $d_I(I^a_{g,s,t}, \tilde{I}^a_{g,s,t})$ is small, where $d_I(\cdot,\cdot)$ is some measure of sequence similarity. The probability distribution ν over environments μ will then tend to give larger probabilities to nearby sequences, than to ones that are far away. An agent would typically be interested in considering only minor hypothetical changes to its interaction sequences, and would have little basis for understanding the consequences of drastic alterations.

Any of the above methods for altering interaction sequences would alter an agent's perception sequence causing changes to the fuzzy truth values mapped by the function F. This in turn would yield new probability distributions over the space of possible worlds, and thereby yielding altered average probability values for the pair (F, T). This change, constructed from the perspective of the agent based on its experience, could then cause the agent to reassess its action w. This is what we mean by "experiential possible worlds" or EPW.

The creation of altered interaction sequences may, under appropriate assumptions, provide a basis for creating better estimates for the predicate F than we would otherwise have from a single real-world data point. More specifically we have the following results, which discuss the estimates of F made by either a single agent or a population of agents, based on each agent in the population subsampling their experience.

Theorem 1. *Let $\mathcal{E}_n$ represent an arbitrary ensemble of n agents chosen from $\mathcal{A}$. Suppose that, on average over the set of agents $a \in \mathcal{E}_n$, the set of values $F(I)$ for mutated interaction sequences I is normal and unbiased, so that,*

$$E[F] = \frac{1}{n} \sum_{a\in\mathcal{E}_n} \sum_{I^a_{g,s,t}\in\mathcal{I}^a} F(I^a_{g,s,t}) P(I^a_{g,s,t}).$$

Suppose further that these agents explore their environments by creating hypothetical worlds via altered interaction sequences. Then an unbiased estimate for $E[F]$ is given by

$$\hat{F} = \frac{1}{n} \sum_{a\in\mathcal{E}_n} \sum_{\tilde{I}^a_{g,s,t}\in\mathcal{I}^a} F(\tilde{I}^a_{g,s,t}) P(\tilde{I}^a_{g,s,t})$$

$$= \frac{1}{n} \sum_{a\in\mathcal{E}_n} \sum_{\tilde{I}^a_{g,s,t}\in\mathcal{I}^a} F(\tilde{I}^a_{g,s,t}) \sum_{e\in E} [P(e|I^a_{g,s,t}) P(\tilde{I}^a_{g,s,t}|e)].$$

Proof. That $\hat{F}$ is an unbiased estimate for $E[F]$ follows as a direct application of standard statistical bootstrapping theorems. See, for example, (DE96). □

Theorem 2. *Suppose that in addition to the above assumptions, we assume that the predicate F is Lipschitz continuous as a function of the interaction sequences $I^a_{g,s,t}$. That is,*

$$d_F\left(F(\tilde{I}^a_{g,s,t}), F(I^a_{g,s,t})\right) \leq K d_I(\tilde{I}^a_{g,s,t}, I^a_{g,s,t}),$$

for some bound K and $d_F(\cdot,\cdot)$ is a distance measure in predicate space. Then, setting both the bias correction and acceleration parameters to zero, the bootstrap BC_α confidence interval for the mean of F satisfies

$$\hat{F}_{BC_\alpha}[\alpha] \subset [\hat{F} - Kz^{(\alpha)}\widehat{\sigma_I},\ \hat{F} + Kz^{(\alpha)}\widehat{\sigma_I}]$$

where $\hat{\sigma_I}$ is the standard deviation for the altered interaction sequences and, letting Φ denote the standard normal c.d.f., $z^{(\alpha)} = \Phi^{-1}(\alpha)$.

Proof. Note that the Lipschitz condition gives

$$\hat{\sigma}^2_F = \frac{1}{n|\mathcal{I}^a| - 1} \times \sum_{a\in\mathcal{E}_n} \sum_{\tilde{I}^a_{g,s,t}\in\mathcal{I}^a} d^2_F\left(F(\tilde{I}^a_{g,s,t}), F(I^a_{g,s,t})\right) P(\tilde{I}^a_{g,s,t})$$
$$\leq \frac{K^2}{n|\mathcal{I}^a| - 1} \sum_{a\in\mathcal{E}_n} \sum_{\tilde{I}^a_{g,s,t}\in\mathcal{I}^a} d^2_I(\tilde{I}^a_{g,s,t}, I^a_{g,s,t}) P(\tilde{I}^a_{g,s,t})$$
$$= K^2\hat{\sigma}^2_I.$$

Since the population is normal and the bias correction and acceleration parameters are both zero, the BC_α bootstrap confidence interval reduces to the standard confidence interval, and the result then follows (DE96). □

These two theorems together imply that, on average, through subsampling via altered interaction sequences, agents can obtain unbiased approximations to F; and, by keeping the deviations from their experienced interaction sequence small, the deviations of their approximations will also be small.

While the two theorems above demonstrate the ability of the subsampling approach to generate probabilistic possible-worlds semantics from experiential semantics, they fall short of being relevant to practical AI inference systems, because the Lipschitz condition in Theorem 2 is an overly strong assumption. With this in mind we offer the following modification, that is more realistic and also in keeping with the flavor of PLN's indefinite probabilities approach. The following theorem basically says that: If one or more agents evaluate the truth value of a probabilistic predicate F via a series of subsampled possible worlds that are normally and unbiasedly distributed around the agent's actual experience, and if the predicate F is mostly smoothly dependent on changes in the world, then evaluating the truth value of F using subsampled possible worlds gives roughly the same results as would be gotten by evaluating the truth value of F across the overall ensemble of possible worlds from which the agent's experience is drawn.

Theorem 3. *Define the set*

$$I^{a;b} = \left\{\tilde{I}^a_{g,s,t} | d^2_F\left(F(\tilde{I}^a_{g,s,t}), F(I^a_{g,s,t}\right) = b\right\},$$

and assume that for every real number b the perceptions of the predicate F satisfy

$$\frac{1}{n}\sum_{a\in\mathcal{E}_n} P(I^{a;b}) \leq \frac{M(b)}{b^2}\sigma^2_I$$

for some $M(b) \in \mathbb{R}$. Further suppose that

$$\int_0^1 M(b)\,db = M^2 \in \mathbb{R}.$$

Then under the same assumptions as in Theorem 1, and again setting both the bias correction and acceleration parameters to zero, we have

$$\hat{F}_{BC_\alpha}[\alpha] \subset [\hat{F} - M\sqrt{n}z^{(\alpha)}\widehat{\sigma_I},\ \hat{F} + M\sqrt{n}z^{(\alpha)}\widehat{\sigma_I}]$$

Proof.

$$\hat{\sigma}^2_F = \frac{1}{n\cdot|\mathcal{I}^a| - 1} \times \sum_{a\in\mathcal{E}_n} \sum_{\tilde{I}^a_{g,s,t}\in\mathcal{I}^a} d^2_F\left(F(\tilde{I}^a_{g,s,t}), F(I^a_{g,s,t})\right) P(\tilde{I}^a_{g,s,t})$$
$$= \frac{1}{n\cdot|\mathcal{I}^a| - 1} \times \sum_{a\in\mathcal{E}_n} \int_0^1 \sum_{\tilde{I}^a_{g,s,t}\in\mathcal{I}^{a;b}} d^2_F\left(F(\tilde{I}^a_{g,s,t}), F(I^a_{g,s,t})\right) P(\tilde{I}^a_{g,s,t})\,db$$
$$\leq \frac{1}{n\cdot|\mathcal{I}^a| - 1} \times \sum_{a\in\mathcal{E}_n} \int_0^1 \sum_{\tilde{I}^a_{g,s,t}\in\mathcal{I}^{a;b}} d^2_F\left(F(\tilde{I}^a_{g,s,t}), F(I^a_{g,s,t})\right) P(\tilde{I}^a_{g,s,t})\,db$$
$$\leq b^2 n\frac{M^2}{b^2}\sigma^2_I = \left(M\sqrt{n}\right)^2\sigma^2_I.$$

□

In the following sections we show how this new formalization of possible worlds semantics can be used to clarify the conceptual and mathematical foundations of several aspects of PLN inference.

Reinterpreting Indefinite Probabilities

Indefinite probabilities (Iea07; Gea08) provide a natural fit with the experiential semantics of the SRAM model, as well as with the subsampling methodology articulated above. An indefinite probability truth-value takes the form of a quadruple $([L,\ U],\ b,\ k)$. The meaning of such a truth-value, attached to a statement S is, roughly: There is a probability b that, after k more observations, the truth value assigned to the statement S will lie in the interval $[L,\ U]$. We interpret an interval $[L,\ U]$ by assuming some particular family of distributions (usually Beta) whose means lie in $[L,U]$.

To execute inferences using indefinite probabilities, we make heuristic distributional assumptions, assuming a "first order distribution of means, with $[L, U\ ]$ as a $(100b)\%$ credible interval. Corresponding to each mean in this "first-order" distribution is a "second order distribution, providing for an "envelope" of distributions.

The resulting bivariate distribution can be viewed as an heuristic approximation intended to estimate unknown probability values existing in hypothetical future situations. Combined with additional parameters, each indefinite truth-value object essentially provides a compact representation of a single second-order probability distribution with a particular, complex structure.

In the EPW context, the second-order distribution in an indefinite probability is most naturally viewed as a distribution over possible worlds; whereas, each first-order distribution represents the distribution of values of the proposition within a given possible world.

As a specific example, consider the case of two virtual agents: one agent, with cat-like characteristics, called "Fluffy" and the second a creature, with dog-like characteristics, named "Muffin." Upon a meeting of the two agents, Fluffy might immediately consider three courses of action: Fluffy might decide to flee as quickly as possible, might hiss and threaten Muffin, or might decide to remain still. Fluffy might have a memory store of perception sequences from prior encounters with agents with similar characteristics to those of Muffin.

In this scenario, one can view the second-order distribution as a distribution over all three courses of action that Fluffy might take. Each first-order distribution would represent the probability distribution of the result from the corresponding action. By hypothetically considering all three possible courses of action and the probability distributions of the resulting action, Fluffy can make more rational decisions.

Reinterpreting Indefinite Quantifiers

EPW also allows PLN's universal, existential and fuzzy quantifiers to be expressed in terms of implications on fuzzy sets. For example, if we have

ForAll $\$X$
    Implication
        Evaluation F $\$X$
        Evaluation G $\$X$

then this is equivalent to

AverageQuantifier \$X
    Implication
        Evaluation F^* \$X
        Evaluation G^* \$X

where e.g. F^* is the fuzzy set of variations on F constructed by assuming possible errors in the historical evaluations of F. This formulation yields equivalent results to the one given in (Gea08), but also has the property of reducing quantifiers to FOPLN (over sets derived from special predicates).

To fully understand the equivalence of the above two expressions, first recall that in (Gea08), we handle quantifiers by introducing third-order probabilities. As discussed there, the three levels of distributions are roughly as follows. The first- and second-order levels play the role, with some modifications, of standard indefinite probabilities. The third-order distribution then plays the role of "perturbing the second-order distribution. The idea is that the second-order distribution represents the mean for the statement $F(x)$. The third-order distribution then gives various values for x, and the first-order distribution gives the sub-distributions for each of the second-order distributions. The final result is then found via an averaging process on all those second-order distributions that are "almost entirely" contained in some $ForAll\_proxy\_interval$.

Next, $AverageQuantifier\ F(\$X)$ is a weighted average of $F(\$X)$ over all relevant inputs $\$X$; and we define the fuzzy set F^* as the set of perturbations of a second-order distribution of hypotheses, and G^* as the corresponding set of perturbed implication results. With these definitions, not only does the above equivalence follow naturally, so do the "possible/perturbed worlds" semantics for the ForAll quantifier. Other quantifiers, including fuzzy quantifiers, can be similarly recast.

Specifying Complexity for Intensional Inference

A classical dichotomy in logic involves the distinction between extensional inference (involving sets with members) and intensional inference (involving entities with properties). In PLN this is handled by taking extension as the foundation (where, in accordance with experiential semantics, sets ultimately boil down to sets of elementary observations), and defining intension in terms of certain fuzzy sets involving observation-sets. This means that in PLN intension, like higher-order inference, ultimately emerges as a subcase of FOPLN (though a subcase with special mathematical properties and special interest for cognitive science and AI). The prior formulation of PLN intension contains a "free parameter" (a complexity measure) which is conceptually inelegant; EPW remedies this via providing this parameter with a foundation in possible worlds semantics.

To illustrate how, in PLN, higher-order intensional inference reduces to first-order inferences, consider the case of intensional inheritance. $IntensionalInheritance\ A\ B$ measures the extensional inheritance between the set of properties or patterns associated with A and the corresponding set associated with B. This concept is made precise via formally defining the concept of "pattern," founded on the concept of "association." We formally define the association operator ASSOC through:

ExtensionalEquivalence
    Member \$E (ExOut ASSOC \$C)
    ExOut
        Func

```
List
    lnheritance $E $C
    Inheritance
        NOT $E
        $C
```

where $\text{Func}(x, y) = [x - y]^+$ and $^+$ denotes the positive part.

We next define a pattern in an entity A as something that is associated with, but simpler than, A. Note that this definition presumes some measure $c()$ of complexity. One can then define the fuzzy-set membership function called the "pattern-intensity," via

$$IN(F, G) = [c(G) - c(F)]^+[P(F|G) - P(F|\not G)]^+.$$

The complexity measure c has been left unspecified in prior explications of PLN, but in the present context we may take it as the measure over concepts implied by the measure over possible worlds derived via subsampling as described above (or perhaps by counterfactuals).

Reinterpreting Implication between Inheritance Relationships

Finally, one more place where possible worlds semantics plays a role in PLN is with implications such as

```
Implication
    Inheritance  Ben  American
    Inheritance  Ben  obnoxious
```

We can interpret these by introducing predicates over possible worlds, so that e.g.

$$Z_{Inheritance\ Ben\ American}(W) < t >$$

denotes that t is the truth value of *Inheritance Ben American* in world W. A prerequisite for this is that *Ben* and *American* be defined in a way that spans the space of possible worlds in question. When defining possible worlds by differing subsets of the same observation-set, this is straightforward; in the case of possible worlds defined via counterfactuals it is subtler and we omit details here.

The above implication may then be interpreted as

AverageQuantifier $\$W$
    Implication
        Evaluation $Z_{Inheritance\ Ben\ obnoxious}$ $\$W$
        Evaluation $Z_{Inheritance\ Ben\ American}$ $\$W$

The weighting over possible worlds $\$W$ may be taken as the one obtained by the system through the subsampling or counterfactual methods as indicated above.

Conclusion

We began with the simple observation that the mind of an intelligent agent accumulates knowledge based on experience, yet also creates hypothetical knowledge about "the world as it might be," which is useful for guiding future actions. PLN handles this dichotomy via a foundation in experiential semantics, and it is possible to formulate all PLN inference rules and truth value formulas in this way. Some PLN truth value formulas are simplified by interpreting them using possible world semantics. With this in mind we used subsampling to define a form of experientially-grounded possible-worlds semantics, and showed its use for handling indefinite truth values, probabilistic quantifiers and intensional inference. These particular technical ideas illustrate the more general thesis that a combination of experiential and possible-worlds notions may be the best approach to comprehending the semantics of declarative knowledge in generally intelligent agents.

References

Thomas J. DiCiccio and Bradley Efron. Bootstrap confidence intervals. *Statistical Science*, Vol. 11, No. 3, 189-228, 1996.

Ben Goertzel and Hugo Pinto et al. Using dependency parsing and probabilistic inference to extract gene/protein interactions implicit in the combination of multiple biomedical research abstracts. In *Proceedings of BioNLP-2006 Workshop at ACL-2006, New York*, 2006.

Ben Goertzel and Matthew Iklé et al. *Probabilistic Logic Networks.* Springer, 2008.

Ben Goertzel. An integrative methodology for teaching embodied non-linguistic agents, applied to virtual animals in second life. In *Proceedings of the First AGI Conference, Memphis*, 2008.

Ben Goertzel. The embodied communication prior. *Proceedings of ICCI 2009, Hong Kong*, 2009.

Ben Goertzel. Toward a formalization of real-world general intelligence. *Submitted to AGI-10*, 2010.

Robert Goldblatt. Mathematical modal logic: a view of its evolution. *Journal of Applied Logic 1: 30992*, 2003.

Matthew Iklé and Ben Goertzel et al. Indefinite probabilities for general intelligence. In *Proceedings of 2006 AGIRI conference, Bethesda, MD: IOS Press*, 2007.

Matthew Iklé and Ben Goertzel. Probabilistic quantifier logic for general intelligence: An indefinite probabilities approach. In *Proceedings of AGI-08. Memphis*, 2008.

David Lewis. *On the Plurality of Worlds.* Basil Blackwell, 1986.

Shane Legg and Marcus Hutter. A formal measure of machine intelligence. In *Proceedings of Benelaam-2006, Ghent*, 2007.