Summary: This paper contains an explanation why costs in WA cannot be linked to probabilities. Furthermore it extracts useful features of WA and interprets them in terms of probabilities. Only a theoretical proposal, no implementation/testing given.

3 Estimating Hypothesis Likelihood

* no work on interpreting the cost propagation in WA in terms of probabilities; therefore, it is not clear whether the min cost hypothesis is the most probable one.
* all abductive approaches to interpretation suffer from the following problems:
  + cannot be proved or disproved because there’s no gold standard to measure correctness against (in the form of e.g. a correct logical structure of a sentence)
  + in cost-based abduction one limits oneself to thinking that the likelihood of a hypothesis depends on the joint likelihood of assumptions, and that these assumptions are independent
  + MLN abduction implies that the probability of a background axiom to hold does not depend on the observation (e.g. you observe that Anna talks to Bob, so you want to increase the probability of the axiom that Anna and Bob are friends, but you can’t)
    - is it true?
  + imply that unifications always hold

4 Abduction for Discourse Processing

* **Unification.** WA prioritises hypotheses with maximum number of unifications (other things being equal).
  + Unification is a principal method by which coreference is resolved. A naïve approach suggests merging predicates with equal names. Results in overmerging unless there is enough (logical) knowledge about disjoint predicates (e.g. dog and cat are different animals, so can’t merge animal(x=dog) and animal(y=cat)). But can’t assume there’s an exhaustive knowledge base for this. Inoue et al. 2012a proposes weighted unification.
  + MODEL PREFERENCE: Other things being equal, a model that contains more true unification nodes is favoured.
* **Observations costs.** Costs reflect demands for propositions to be proved. Propositions more likely to be linked referentially to other parts of discourse are expensive to assume (example: given *The smart man is tall*, we will want to prefer those hypotheses which prove *smart* and assume *tall* (to link *smart* referentially), so we put high cost on *smart* and low on *tall*. Given the following KB:
  + educated(x)1.2 -> smart(x)
  + big(x)1.2 -> tall(x)
  + educated(John), big(Bill)

WA should prefer to assume that x is John – we prove that he is smart and pay the cost of assuming he is tall (because *The smart man is tall*). We don’t want to assume that x is Bill because then we’d prove *tall* and assume *smart*, which goes against our initial assumption.

* + MODEL PREFERENCE: Other things being equal, a model that explains referential observables is favoured;
* **Weighted conjuncts in the antecedents.** Weights on antecedents correspond to the semantic contribution of that antecedent to explaining the consequent (the greater the weight, the greater the contribution).
  + MODEL PREFERENCE: Other things being equal, a model that results from application of more reliable axioms is favoured.

5 Graph Representation of Hypotheses

* idea: hypotheses represented as models of an AODAG such that:
  + AND nodes are operation nodes (inference and unification);
  + OR nodes are literal nodes;
  + u -> v means that u is an immediate parent of v;
  + a model is a truth assignment to set V in G and corresponds to a unique hypothesis;
* list of conditions ensured by the graph framework:
  + all observables are true in every model
  + an operation node is true if its result is true (otherwise it’s false)
  + literal node is true if one of its explanations is true
* notes about the model:
  + in a model, nodes that are assigned T and have no parents with T are called *assumptions;*
  + in a model, if u -> v and T(u) and T(v) then u explains v in the model;

6 Probabilities and Independence Assumptions

Estimate likelihood of a hypothesis by defining a set [Xi] of random variables associated to each node in AODAG. They take values of {T,F} and the joint probability distribution of set {X1…Xn} is the product of conditional probabilities, where Xi is conditioned on pii, which denotes all other variables from the set on which Xi depends.

Applying local Markov property to the Bayesian network (AODAG) we can determine that a variable is independent of its descendants given its parents.

In our case exception is unification because unifications anywhere in the model raise the probability of a model. For a model X we want to know how many unifications there are in the model, so we introduce another type of a random variable: numbUv for every axiom node v which stands for number of true unifications that are parents of v

Refv – a random variable which accounts for referentiality: if v is a referential observable or has such an observable as its child, then Refv = T, otherwise F.

# Note: only associated with literal nodes

numbUv – associated with axiom node; a random variable which indicates how many true unifications exist that are parents of v.

# Question: how do I indicate it when there is more than 1 unification?

Comment: The probabilities resulting from the above formulation interpret the following inequalities (which Hobbs 1992 interpreted as weights and costs):