

Universidade de São Paulo
Instituto de Matemática e Estatística
MAC 5778 - Sistemas Baseados em Conhecimento

PR-OWL:
Probabilistic Web
Ontology Language

Autor:
Walter Perez Urcia

São Paulo
Outubro 2015

1 Resumo

O conceito da web semântica é uma ideia de Tim Berners-Lee: “Web semântica é uma extensão da web comum em que a informação tem um melhor definido significado para os computadores e a cooperação das pessoas” [6]. Em outras palavras, mudar a interconexão da web comum a um gigantesco banco de dados relacionado.

O problema com aquela ideia de web semântica é que na web atual existem as seguintes situações:

- Amplidão: Na web existem milhões de páginas
- Não definição única de termos: Existem termos como alto e jovem que são subjetivos
- Incerteza: Conceitos que tem incerteza, por exemplo sintomas de uma doença podem ser de alguma outra com outra probabilidade
- Inconsistencia: Existem contradições lógicas
- Engano: A informação obtida não é sempre confiável

No ano 2004, World Wide Web Consortium estabeleceu como padrão a linguagem OWL (Web Ontology Language) para representar ontologias que eram a principal representação para conseguir uma web semântica. Uma ontologia é uma representação formal de conhecimento baseada na lógica de descrição que inclui tipos de entidades (e.g. Pessoa, Companhia), propriedades de aquelas entidades (e.g. nome, sobrenome), relações entre entidades (e.g. paiDe) e eventos que acontecem com aquelas entidades. Devido a que está baseada em lógica, existe a possibilidade de fazer inferências sobre os fatos estabelecidos. Além disso, também pode ser usada para ajudar aos motores de busca responder a perguntas mais complexas.

O principal problema de usar as ontologias comuns e OWL como medios para obter a web semântica é que a web atual tem incerteza em muitos conceitos. A incerteza poderia estar não só em tipos de entidades (e.g. Python é uma linguagem de programação ou um animal), se não também em propriedades ou relações que tem diferentes comportamentos em diferentes domínios (e.g. limpo pode significar que não tem registros policiais, ou que não está sujo). Tendo isto em consideração, existe uma necessidade de encontrar uma representação que use tanto lógica como teoria de probabilidades para representar o conhecimento e também de estender a linguagem OWL para ter suporte de probabilidades.

Aquela representação precisada pode ser encontrada extendendo o conceito de ontologias a ontologias probabilísticas. As ontologias probabilísticas estendem o conceito das ontologias comuns para lidar com incerteza sobre conceitos ambíguos, não confiáveis ou

incompletos, sejam estes relacionados com tipos, propriedades ou relações. Existem varias abordagens feitas para encontrar um jeito de representar um modelo de conhecimento baseado na lógica que lide com incerteza ao mesmo tempo. Entre as principais estão Hidden Markov Models (HMM), Bayesian Networks (BNs), Probabilistic Relational Models (PRM), entre outras [2], mas todas elas tem desvantagens ou não conseguem combinar ambas coisas satisfatoriamente.

Mas no ano 2008, Laskey conseguiu desenvolver a primeira representação que combina lógica com incerteza de forma satisfatoria. Aquela representação está baseada em um tipo de rede bayesiana chamado Multi-Entity Bayesian Network (MEBN) onde cada sentença em lógica de primeira ordem pode ser representada como um fragmento ou MFrag. Cada um desses fragmentos tem distribuições de probabilidades locais e o conjunto de MFraags construi toda a base de conhecimento [3].

Usando MEBNs foi desenvolvida a extensão da linguagem OWL com incerteza, chamada PR-OWL (Probabilistic Web Ontology Language) que conseguia adicionar probabilidades a conjuntos de variáveis que dependem de outras. Dito de outra forma, para cada sentença lógica representada em um MEBN e suas instâncias possíveis, podiam ser adicionadas probabilidades.

Por último, com aqueles dois conceitos: probabilistic ontologies e PR-OWL já definidos formalmente foram desenvolvidos alguns estudos como a implementação de uma ferramenta gráfica para modelar ontologias probabilísticas baseadas em MEBNs chamada UnBBayes [5]. Da mesma forma, usando aquela nova ferramenta foram feitos mais estudos em situações do mundo real como o modelamento de uma ontologia marítima [4] e outra para o reconhecimento de fraudes em Brasil [1].

Referências

- [1] RommelN. Carvalho, Shou Matsumoto, KathrynB. Laskey, PauloC.G. Costa, Marcelo Ladeira, and LaécioL. Santos. Probabilistic Ontology and Knowledge Fusion for Procurement Fraud Detection in Brazil. In *Uncertainty Reasoning for the Semantic Web II*, volume 7123 of *Lecture Notes in Computer Science*, pages 19–40. Springer Berlin Heidelberg, 2013.
- [2] Paulo C. G. da Costa and Kathryn Blackmond Laskey. Multi-entity bayesian networks without multi-tears, 2010.
- [3] Kathryn B. Laskey. MEBN: A Language for First-ORder Bayesian Knowledge Bases. *Artificial Intelligence*, 172, 2008.

- [4] Kathryn Blackmond Laskey, Richard Haberlin, Paulo Costa, and Rommel Novaes Carvalho. PR-OWL 2 Case Study: A Maritime Domain Probabilistic Ontology, 2011.
- [5] Show Matsumoto Marcelo Ladeira Paulo Costa Rommel Carvalho, Laecio Santos. UnBBayes-MEBN: Comments on Implementing a Probabilistic Ontology Tool. *IADIS Applied Computing 2008 conference*, 2008.
- [6] James Hendler Tim Berners-Lee and Ora Lasilla. The Semantic Web. *Scientific American*, 2001.