

Abordagens Estocásticas para Raciocinadores aplicáveis em Web Semântica

Juliano T. Brignoli¹, Denilson Sell², Fernando O. Gauthier³

¹IFC - Instituto Federal Catarinense
Rio do Sul - SC – Brasil

²UFSC – Universidade Federal de Santa Catarina
Pós-graduação em Engenharia e Gestão do Conhecimento
Florianópolis – SC – Brasil

³UFSC – Universidade Federal de Santa Catarina
Pós-graduação em Engenharia e Gestão do Conhecimento
Florianópolis – SC - Brasil

juliano.brignoli@gmail.com, denilson@stela.org.br, gauthier@inf.ufsc.br

Abstract. *We propose a treatment of uncertainty in areas of Semantic Web ontologies for modifying the schema PR-OWL. The scheme is limited to considering the uncertainties of randomness, but certain areas of representation of the real world can also have imprecise variables. To qualify for the process examined plausible inference is a hybrid model of reasoning that can theoretically be incorporated into the scheme of syntactic and functional PR-OWL. The article explores this conceptual integration, assuming that PR-OWL will increase the accuracy of the results of inference.*

Resumo. *Propõe-se um tratamento de incertezas em domínios de ontologias para Web Semântica modificando-se o esquema PR-OWL. O esquema limita-se a considerar as incertezas por aleatoriedade, mas, certos domínios de representação do mundo real também podem apresentar variáveis imprecisas. Para qualificar o processo de inferência plausível analisou-se um modelo de raciocínio híbrido que, teoricamente, pode ser incorporado no esquema sintático e funcional da PR-OWL. O artigo explora conceitualmente esta integração, pressupondo que PR-OWL aumente a acurácia dos resultados da inferência.*

1. Introdução

A inferência é um processo inerente à operacionalização das ontologias em Web Semântica, amplamente explorado em aplicações focadas na determinação de respostas a partir de uma busca analítica sobre um ambiente com variáveis interrelacionadas.

A teorização acerca da utilização de inferências perpassa fundamentalmente pelos postulados da Lógica de Primeira Ordem e todo o seu desenvolvimento aprimorado por séculos, dadas as contribuições de matemáticos e filósofos da antiguidade e da contemporaneidade.

No desenvolvimento da Web Semântica e, mais especificamente, nas construções de ontologias, os processos de inferência são executados por motores de raciocínio (*reasoners*). Existem diversas especificações ou formas de atuação dos raciocinadores aplicáveis em Web Semântica, existindo desde aqueles formulados por princípios da Lógica, por processos otimizados de busca com conotação semântica e ainda, aqueles com funcionalidades inerentes ao tipo de tratamento estocástico de informações.

Este artigo objetiva apresentar uma discussão acerca da utilização do raciocínio plausível em Web Semântica, destacando a Probabilistic-OWL como um dos formalismos para aplicações em domínios de ontologias com incertezas. Pretende abordar de maneira teórica a possibilidade da contribuição de um modelo híbrido de raciocínio, combinando o processamento da imprecisão e da aleatoriedade.

2. Os raciocinadores para Web Semântica

Racoinadores computacionais realizam operações lógicas aplicadas sobre bases de conhecimentos compostas por fatos e são amplamente utilizados nas ontologias, agindo de modo a executar regras que buscam alcançar respostas a um processo de inferência.

Ontologias em Web Semântica podem apresentar domínios complexos e dinâmicos e dispõem de uma linguagem formal que descreve representações, é a OWL, normatizada pelo W3C. No âmbito da OWL encontra-se a OWL DL, como sendo uma derivação que suporta inferência a partir da Lógica Descritiva. Horrocks (2002) define a Lógica Descritiva como uma denominação geral para uma família de formalismos de representação do conhecimento. Ao referenciar ferramentas para representação de ontologias é evidente o uso intensivo do software Protégé (PROTÉGÉ, 2010). Para este framework existem vários plug-ins disponibilizados, tais como, o raciocinador Pellet, da Clark & Parsia LLC, sendo um mecanismo de inferência especificado em OWL DL, desenvolvido em Java e de código aberto. Conforme observado em (PELLET, 2010), Pellet é uma implementação baseada no algoritmo de Tableaux (tablô), um método formal oriundo da Lógica de Primeira Ordem. O Pellet dá suporte à especificação de tipos de dados que podem ser definidos pelo usuário, incorpora heurísticas para detecção de ontologias OWL Full na tentativa de expressá-las com OWL DL e identifica axiomas que causam inconsistências entre conceitos.

Em (SILVA, 2008) é possível conhecer diversos mecanismos de inferência para servirem às ontologias e entre vários observou-se o Fuzzy DL. Este sugere uma abordagem mais exploratória e uma reflexão acerca da utilização de métodos focados ao processamento de incertezas existentes em domínios nas ontologias para a Web Semântica. Na seqüência deste artigo, propõe-se discorrer sobre o contexto estocástico que pode ser manifestado na representação de ontologias.

3. Fuzzy e a incerteza pela imprecisão

Em conformidade com a abordagem de (ROSS, 1995), a Lógica Fuzzy é um formalismo matemático para modelar problemas que apresentam variáveis munidas de incertezas inerentes

à imprecisão de seus valores. Na Lógica Clássica (bivalente), dado um conjunto A, um elemento $x \in A$ ou $x \notin A$, em Fuzzy, um elemento possui uma medida de possibilidade num intervalo [0, 1]. A modelagem fuzzy sugere o uso de variáveis lingüísticas para descrever o domínio em representação. Estas variáveis estão associadas com objetos ou conceitos do mundo real possíveis de uma especificação quantitativa, tais como, velocidade, altura, temperatura, etc. Estas variáveis são segmentadas em intervalos denominados Conjuntos Difusos que podem, em circunstâncias, caracterizar certas sobreposições de conceitos, como, pouco, pouquíssimo, muito, muitíssimo, denotando imprecisão de conceitos na representação de domínios.

4. Raciocínio Plausível Bayesiano

O raciocínio bayesiano apresenta solução de inferência na modelagem de domínios cujas variáveis estão sujeitas ao acaso, onde o fator randomicidade é o tipo de incerteza manifestada. É propício enfatizar uma diferença fundamental entre a proposta da Lógica Fuzzy e do Raciocínio Bayesiano: em Fuzzy, uma variável é incerta pela sua imprecisão e a certeza é uma medida de possibilidade. No contexto da aplicação do Raciocínio Bayesiano, as variáveis são incertas devido à aleatoriedade intrínseca e a certeza é uma medida de probabilidade.

Pearl (1988) aborda a descrição formal deste processo de inferência bayesiana onde a relação entre variáveis acarreta em uma rede de interdependências, logo, a modificação de uma variável do ambiente pode afetar várias outras em termos de crenças probabilísticas. De maneira sucinta apresenta-se o princípio da formulação matemática do Teorema de Bayes: Seja o espaço de probabilidade (ϵ, P) e os eventos compostos $H_1, H_2, \dots, H_k \subseteq \epsilon$ desde que nenhum desses eventos tenha probabilidade nula, então:

$$P(H_i / e) = \frac{P(e / H_i) \cdot P(H_i)}{P(e)} \quad (1)$$

$$\text{onde } P(e) = P(H_1) \cdot P(e / H_1) + P(H_2) \cdot P(e / H_2) + \dots + P(H_k) \cdot P(e / H_k) \quad (2)$$

As fórmulas (1) e (2) compõem o cerne do mecanismo de inferência das denominadas Redes Bayesianas. Contudo, conforme discorre Carvalho *et al.* (2008) estas redes apresentam limitações para representar diversas situações do mundo real e da Web como, dificuldades em lidar com elevado número de variáveis aleatórias e o uso de recursão, o que retrata problemas no campo da complexidade computacional. Logo, Carvalho *et al.* (2008) apresenta uma proposta acerca do raciocínio bayesiano em ontologias para Web Semântica utilizando Redes Bayesianas Multi-Entidades (MEBN). Estas redes integram a Lógica de Primeira Ordem com o raciocínio bayesiano, conforme proposta desenvolvida por Laskey (2008). No âmbito da especificação de ontologias, a linguagem OWL não suporta raciocínio plausível, assim, Carvalho *et al.* (2008) recorre à linguagem PR-OWL (probabilistic OWL) como sendo a primeira implementação de MEBN para ontologias de domínios com incertezas. (PR-OWL, 2011). Trata-se de uma contribuição para os trabalhos desenvolvidos pela URW3-XG (W3C Uncertainty Reasoning for the World Wide Web Incubator Group). As MEBN modelam conhecimento por meio de MFrag que constituem esquemas de entidades de domínio por meio de grafos. Segundo Carvalho *et al.* (2008), uma MFrag é uma distribuição de probabilidade das instâncias de suas variáveis aleatórias. A figura 1 ilustra por meio de uma MFrag, a modelagem que representa o nível de perigo a que uma nave é exposta.

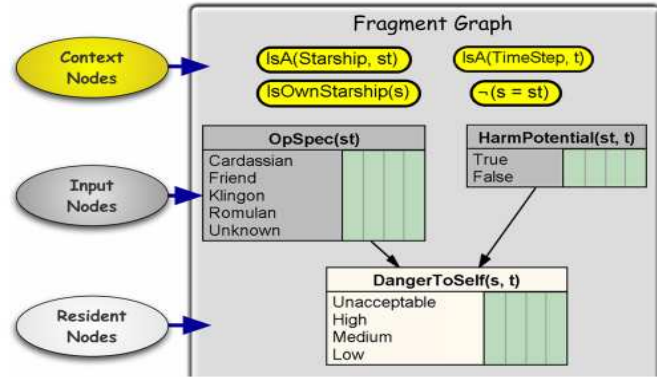


Figura 1 - exemplo de uma Mfrag (Carvalho *et al.* (2008))

O problema representado pela figura 1 é significativo para elucidar elementos de investigação no campo do processamento de incertezas. Observa-se a especificação de variáveis aleatórias que denotam subjetividade, ou seja, são vinculados a um domínio com valores imprecisos. Os estados *High*, *Medium*, *Low*, caracterizam conceitos difusos que incorporam mais incertezas à MEBN.

Em conformidade com esta observação objetiva-se convergir para uma proposição: utilizar Lógica Fuzzy na formulação híbrida do raciocínio plausível em ontologias para qualificar a inferência em função dos conceitos difusos existentes em nós residentes em uma Mfrag. Na seqüência será apresentada a proposta de formulação.

5. Possibilidade da representação de Raciocínio Híbrido para Web Semântica

Em suas obras, Ross (1995) e Kandel (1986) introduzem algumas abordagens e ensaios matemáticos de modo a nortear pesquisas que venham a caracterizar uma utilização combinada dos processos de inferência fuzzy e bayesiana.

Brignoli (2001) sugeriu em seu trabalho que uma Rede Bayesiana pode ter uma redistribuição de suas probabilidades de saída (resultados) quando inferirem sob variáveis de entrada que apresentam imprecisão como incerteza eminente. Para tal, o autor apresenta $\phi = f(\rho, \delta)$ como uma função que expressa o que denominou de qualificador para o modelo híbrido de raciocínio. Mostra por simulação que ϕ causará a redistribuição das probabilidades da Rede Bayesiana tendo ρ como uma probabilidade condicional sem imprecisão e, δ representando um ou mais eventos difusos. Estes eventos podem ser as variáveis de entrada da Rede Bayesiana.

Brignoli (2001) utilizou a fórmula do Teorema de Bayes (1) com modificações em sua forma original, conforme segue:

$$P(H_i / \delta) = \frac{P(\delta / H_i)}{P(\delta)} \quad (3)$$

onde,

$$P(\delta H_i) = P(H_i) \cdot [P(\delta H_i) \cdot \mu_{\mathcal{E}_1}(\delta) + (1 - P(\delta H_i)) \cdot \mu_{\mathcal{E}_2}(\delta)] \quad (4)$$

e,

$$P(\delta) = \sum_{i=1}^n P(H_i) \cdot [P(\delta H_i) \cdot \mu_{\mathcal{E}_1}(\delta) + (1 - P(\delta H_i)) \cdot \mu_{\mathcal{E}_2}(\delta)] \quad (5)$$

Na fórmula, ϵ_1 e ϵ_2 foram usados para representar conjuntos difusos de entrada que determinaram a incerteza por imprecisão existente nas variáveis de entrada da Rede Bayesiana. Em termos de confiabilidade do modelo é importante ressaltar que após ser aplicada a função ϕ de qualificação, o axioma estatístico $\sum P(H_i) = 1$ continua verdadeiro.

Uma conclusão notória obtida nesta modelagem é o que a utilização do modelo de inferência híbrido causou um espalhamento na distribuição de probabilidades da Rede.

A argumentação para este espalhamento está baseada na observação de que, quando as probabilidades estão condicionadas a eventos difusos, aquelas acima do ponto de máxima entropia, diminuem, enquanto as demais aumentam. Desta forma, o processo de qualificação provoca uma redistribuição das probabilidades e o referencial para esta mudança de informação é justamente o ponto em que está situada a maior incerteza no conhecimento da Rede. (BRIGNOLI, 2001, p. 110)

Considerando a validade do modelo descrito, voltamos a referenciar a PR-OWL e o esquema ilustrado na figura-1. Carvalho *et al.* (2008) propôs a implementação de uma ferramenta visual que servisse como plug-in ao Protégé para apoiar a representação de ontologias com domínio probabilístico e incorporou funcionalidades da PR-OWL nesta ferramenta. O autor argumenta a vantagem desta implementação visual dado o fato que arquivos PR-OWL são similares aos arquivos HTML, sendo exaustiva sua construção e manutenção.

Conhecida e disponibilizada a PR-OWL, propõe-se a incorporação de mecanismos de inferência híbridos que agreguem aos raciocinadores de ontologias em Web Semântica, a partir, do próprio uso da PR-OWL, as funcionalidades para operar sobre domínios com imprecisão e aleatoriedade. Torna-se oportuno explorar uma esquematização para uma Híbrido-OWL como sendo uma extensão da PR-OWL. A figura 2 sugere um macro-esquema:

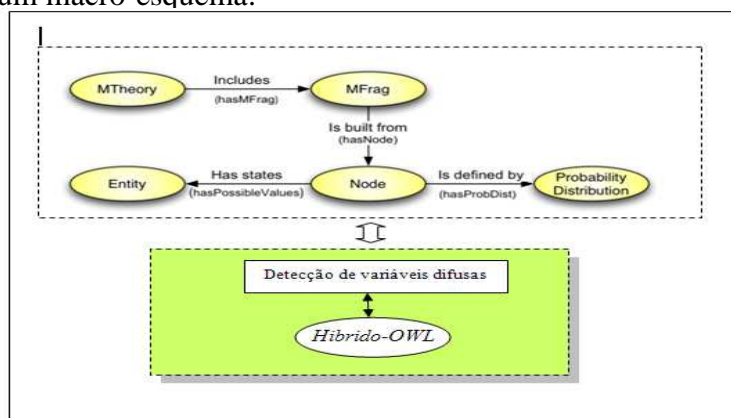


Figura 2 - esquema de uma PR-OWL híbrida (Adaptado de Carvalho et al. (2008))

É importante enfatizar que o objetivo é apresentar uma possibilidade de integração das funcionalidades de uma inferência híbrida fuzzy-bayesiana ao esquema da PR-OWL já em uso nas ontologias com domínios de incerteza.

Em termos operacionais e de implementação computacional, poder-se-ia agregar ao código PR-OWL a capacidade de inferir imprecisão em ontologias com a inserção adequada do modelo apresentado na fórmula (3).

6. Conclusões

A exploração teórica mostrou a possibilidade de uma investigação com maior profundidade no tema visando propostas implementáveis. A abordagem estocástica e o estudo das incertezas são convenientes no âmbito das ontologias em Web Semântica

devido à magnitude de variáveis que representam domínios. Notou-se que as ontologias no âmbito da Web Semântica podem acarretar conceitos e variáveis incertas, expressas na forma de imprecisão ou de plausibilidade.

O estudo das inferências foi fundamental para a proposta de construção de motores de raciocínio aplicáveis às ontologias em Web Semântica e a PR-OWL incrementou a qualidade da acurácia em ontologias que requerem, por sua natureza de descrição, uma forma de raciocínio plausível oriundo da aleatoriedade especificada nas relações das variáveis.

Na análise da funcionalidade da PR-OWL ficou perceptível sua limitação ao raciocínio sobre incertezas por imprecisão, ponto crucial que identificou uma lacuna no processo de raciocínio sobre incertezas nas ontologias em Web Semântica e oportunizou uma reflexão acerca da utilização de um raciocinador híbrido agregado às funcionalidades da PR-OWL. Por ter sido utilizado em outras aplicações, o raciocinador híbrido elucidado neste artigo sugeriu um caminho promissor na investigação sobre esquemas formais para lidar com cenários de incerteza em ontologias para Web Semântica. Esta verificação assegura, *a priori*, uma adequação ao esquema PR-OWL, com provável incremento da eficácia na inferência.

Como maior contribuição ao processo de inferência realizada em ontologias, o raciocinador híbrido argumentado neste trabalho tende em acarretar uma qualificação dos resultados no que tange ao aspecto da semântica produzida.

Referências

- Bravo, Carlos de Oliveira. (2010). Geração Automática de Ontologias para Web Semântica. Dissertação de Mestrado. UnB. Brasília.
- Brignoli, Juliano T. (2001). Modelo Híbrido Difuso-Probabilístico: uma alternativa para Sistemas Especialistas. Dissertação. UFSC. Florianópolis.
- Carvalho, Rommel Novaes et al. (2008). Raciocínio Plausível na Web Semântica através de Redes Bayesianas Multi-Entidades – MEBN. Dissertação. UnB. Brasília.
- Horrocks, I. (2002). DAML+OIL: a description logic for the semantic web. IEEE Data Engineering Bulletin, 25(1):4–9. URL <http://citeseer.ist.psu.edu/578691.html>.
- Kandel, Abraham. (1986). Fuzzy Mathematical Techniques with Applications. Florida: Addison-Wesley Publishing Company.
- Laskey, K. B. (2008). MEBN: A Language for First-Order Bayesian Knowledge Bases. Artificial Intelligence, 172(2-3).
- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. California: Morgan Kaufmann.
- Pellet. (2010). Disponível em: <http://pellet.owldl.org/>. Acesso em: <15/08/2010>.
- Protégé. (2010). Disponível em: <http://protege.stanford.edu>. Acesso em: <15/08/2010>.
- Pr-OWL. (2011). Disponível em: <http://www.pr-owl.org>. Acesso em: <22/08/2011>.
- Ross, Timothy J. (1995). Fuzzy Logic With Engineering Applications. McGraw-Hill.
- Silva, Marcel Ferrante. (2008). Semantic web reasoning. Seminários de disciplina. Organização e Tratamento da Informação. UEMG. Belo Horizonte.