Preferences, Links, and Probabilities for Ranking Objects in Ontologies

Thomas Lukasiewicz [1, 2] and Jörg Schellhase [3]

[1] DIS, Università di Roma "La Sapienza"
Via Salaria 113, I-00198 Rome, Italy
lukasiewicz@dis.uniroma1.it
[3] Fachgebiet Wirtschaftsinformatik, Universität Kassel
Nora-Platiel-Straße 4, D-34127 Kassel, Germany
schellhase@wirtschaft.uni-kassel.de

Abstract. In previous work, we have introduced variable-strength conditional preferences for ranking objects in ontologies. In this paper, we continue this line of research. We propose a new ranking of objects, which integrates this user-defined preference ranking of objects with Google's importance ranking (called *PageRank*) based on the link structure between the objects. We also propose to use probabilistic description logics based on Bayesian networks and the description logic *DL-Lite* to compute the ranking of incompletely specified objects.

1 Overview

In their seminal work [5], Smyth and Poole deal with the problem of matching instances against models of instances, which are both described at different levels of abstraction and at different levels of detail, using qualitative probability theory. Informally, such problems can be described as follows. Given an instance I and a model of instances M, compute the qualitative probability that the instance I is matching the model M (that is, of I given M). For example, in web (resp., literature, product) search, an instance I may be given by a web page (resp., piece of literature, product), while a model M may be given by a web (resp., literature, product) search query.

In the work [4], we continue this important line of research. We present a formalism for ranking objects in description logics that allows for expressing conditional preferences, which are sentences of the form "generally, in the context ϕ, property α is preferred over property $\neg\alpha$ with strength s", in models of instances.

An orthogonal way of ranking objects is based on the analysis of the link structure between the objects. For example, web pages generally contain links to other web pages, and pieces of literature generally cite other pieces of literature. The PageRank technique, which stands behind the web search engine Google [1], is one of the most prominent ways of ranking objects based on the link structure between the objects. The PageRank of a web page u is defined as $R(u) = c \cdot (\sum_{v \in B_u} R(v) \,/\, N_v + E(u))$, where (i) B_u is the set of pages that point to u, (ii) N_v is the number of links from v, (iii) c is a normalization factor, and (iv) $E(u)$ is a vector over web pages representing a source of rank. Informally, the more web pages with high rank point to a web page, the higher is the rank of this web page. The PageRank ranking thus extracts the importance of

[2] Alternate address: Institut für Informationssysteme, Technische Universität Wien, Favoritenstraße 9-11, A-1040 Vienna, Austria; e-mail: lukasiewicz@kr.tuwien.ac.at.

a web page from the link structure between the web pages. In this paper, we propose to combine the user-defined preference ranking of objects based on conditional preferences with the importance ranking of objects specified by PageRank. This allows for influencing the PageRank ranking by user-defined conditional preferences (e.g., for a better web search or for personalization purposes), and to refine the ranking of objects based on user-defined conditional preferences by PageRank's importance ranking.

Furthermore, to compute the (combined) ranking of incompletely specified objects (which abound on the web), we propose to use probabilistic description logics that are based on Bayesian networks (similar to the ones in [3]) and the description logic *DL-Lite* [2]. They allow to specify terminological probabilistic knowledge between concepts, which can be used to compute the expected concept memberships of objects.

2 Example

Consider the following query Q in literature search: We are looking for publications with the word "matching" in the title. In case of a conference paper, we prefer papers of international conferences to papers of national conferences:

$$\begin{aligned} Q = {} & \mathit{Publication} \sqcap \mathit{in\_title}(\text{“matching”}) \sqcap \\ & (\mathit{type}(\text{“international”})|\mathit{ConfPublication})[70] \sqcap (\mathit{ConfPublication})[80]\,. \end{aligned}$$

Query Q contains two conditional preferences. Intuitively, an object that fulfills query Q has to be a publication with the word "matching" in the title and it should possibly satisfy the two conditional preferences. Publications that satisfy the conditional preferences have a lower rank than publications that falsify them. Query Q therefore divides the publications in the query result into three groups as follows: first international conference publications (lowest rank), second national conference publications (second lowest rank), and third non-conference publications (highest rank).

There are now two ways of combining this preference ranking with the importance ranking of PageRank. The first one is dominated by the preference ranking and simply uses the PageRank ranking to order the publications of the same preference rank according to their importance, while the second one is dominated by the PageRank ranking, and it uses the preference ranking as input $E(u)$ to the PageRank computation.

In order to rank incomplete objects, we can then additionally exploit the information encoded in probabilistic description logics. For example, suppose that "every publication is a conference publication with probability 0.9". Thus, if we know that an object o is a publication, then we can conclude that it is a conference publication with probability 0.9, which can then be exploited to compute the (expected) rank of o.

References

1. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings WWW-1998*, pp. 107–117.
2. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. *DL-Lite*: Tractable description logics for ontologies. In *Proceedings AAAI-2005*, pp. 602–607.
3. D. Koller, A. Levy, and A. Pfeffer. P-CLASSIC: A tractable probabilistic description logic. In *Proceedings AAAI-1997*, pp. 390–397.
4. T. Lukasiewicz and J. Schellhase. Variable-strength conditional preferences for ranking objects in ontologies. In *Proceedings ESWC-2006*, pp. 288–302.
5. C. Smyth and D. Poole. Qualitative probabilistic matching with hierarchical descriptions. In *Proceedings KR-2004*, pp. 479–487.