A Technique for Representing Course Knowledge Using Ontologies and Assessing Test Problems

Javed Khan and Manas Hardas

Kent State University, Kent, Ohio 44240, USA
{javed,mhardas}@kent.edu

Summary. In this paper we present a novel method for qualitative assessment of educational resources, specifically 'test problems'. For test problems, the basic elements of design are in the form of concepts arranged in a hierarchy. Course concept knowledge can be represented in the form of prerequisite relation based ontology using which, assessment and information extraction from test problems is possible. Using a schema based on Web Ontology Language (OWL), course ontologies can be represented in a standard and sharable way. Some synthetic parameters for the assessment of a test problem in its concept space are introduced. The assessment system can be further extended to analyze any type educational resource.

Keywords: ontology, assessment, complexity, knowledge, test problems.

1 Introduction

The web has greatly facilitated online sharing of course material. There have been many organized attempts to create large digital courseware libraries to promote sharing like NIST's Materials Digital Library Pathway, NSDL Digital Libraries, OhioLink, ACM Professional Development Center etc. MITŠs Open Course Ware (OCW) project has more than 1000 course materials freely available, Universia maintains translated versions of OCW courses in 11 languages, China Open Resources for Education (CORE) has a goal to include Chinese versions of the OCW. The amount of digital courseware content available online is huge. Surprisingly, the real sharing of the materials among the educators is still very low. In OCW it has been noted that only 16% of the users are educators out of which not more than 26% use it for planning their course or teach a class [1]. Most courseware today, on the web or otherwise is not accompanied with a conceptual design. There is no well formed encoding principle for capturing and sharing the schema associated with course materials. To make this digital content reusable, the associated meta data should be consistently represented. Traditionally, concepts maps are used to represent the concept space for the course knowledge like for biology courses [2] and many other areas however they are too expressive and consequently contain more semantic relationships than necessary for effective computation. Ontologies provide a means to effectively map this knowledge into concept hierarchies. Standardization of semantic

K.M. Węgrzyn-Wolska and P.S. Szczepaniak (Eds.): Adv. in Intel. Web, ASC 43, pp. 349–354, 2007.
springerlink.com

representation standards like RDF and OWL offers great technical platform to represent the concept knowledge space symbolized by ontologies and greatly improve its machine usability. In this paper we present an approach to course knowledge representation using ontology in an expressible and computable format using has-prerequisite relationships where concepts involved in teaching a course are arranged in an hierarchical order of learning. Another original approach for specifically pointing out areas in ontologies of maximum relevance called as *CSG extraction* is given. We investigate the properties of test problems by following a purely knowledge based approach for assessment using course ontologies. The system has the potential to make the already available test-ware resources on the web reusable.

2 Course Ontology

Most ontologies today are so extensive in the breadth of knowledge that processing of these ontologies becomes a gargantuan computation task. There needs to be a way to efficiently process the relevant information to give results in minimum time and complexity of computation. Ontologies are made up of individuals, classes, attributes and relationships. The has-prerequisite relationship in a course ontology refers to the prerequisite understanding of the child node needed to understand the parent node. On the whole the course ontology is constructed in such a hierarchical fashion that the children of node represent the knowledge required to understand the parent node. A node is characterized by two values namely, self-weight and prerequisite weight, Figure 1. The self-weight of a concept node is the amount of knowledge required to grasp the concept. To understand the concept entirely, knowledge of the prerequisite concepts is also required, which is given by the prerequisite weight of the node. Another value which characterizes the course ontology representation is the link weight. The link weight again is the value of the semantic importance of child concept to the parent concept. Child concepts imperative in the understanding of parent concepts will have a greater link weights than the others.

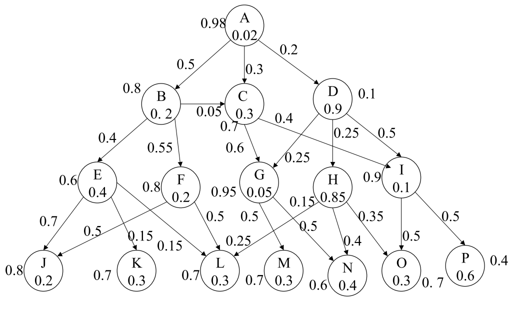

Fig. 1. Example Concept space graph, T(A)

2.1 Representing Course Ontologies

The course ontology is mathematically defined in the form of a concept space graph (CSG). A CSG is a view of the concepts space distribution in the domain of a particular course.

A concept space graph $T(C, L)$ is a projection of a semantic net with vertices C and links L where each vertex represents a concept and each link with weight $l(i, j)$ represents the semantics that concept c_j is a prerequisite for learning c_i, where $(c_i, c_j) \in C$ and the relative importance of learning c_j for learning c_i is given by the weight. Each vertex i in T is further labeled with self-weight value $W_s(i)$ and cumulative prerequisite set weight $W_p(i)$.

A CSG with root A is represented as T (A) in Figure 1. For any node in the CSG, the sum of self-weight and prerequisite weights and the sum of the link weights for all children is 1.

2.2 Prerequisite Effect of a Node

The notion of node path weight is introduced to compute the effect a prerequisite node has on a root node through a specific path. A single node can have different prerequisite effect on a root through different paths.

When two concepts x_0 and x_t are connected through a path consisting of nodes given by the set $[x_0, x_1, ..., x_t]$ then the node path weight between these two nodes is given by:

$$\eta(x_0, x_t) = W_s(x_t) \prod_{m=t}^{1} l(x_{m-1}, x_m) * W_p(x_{m-1}) \quad (1)$$

In the Figure 2, concept L is connected to B through E and F. Therefore the prerequisite effect it has on B is dependent on the prerequisite effect both E and F have on B respectively. From the node path weight calculations we can see that L has a stronger prerequisite effect on B through F rather than E. This is because, L is more important to F (0.5) than E (0.15), prerequisite importance of L is more to F (0.8) than E (0.6) and subsequently F (0.55) is more important to B than E (0.4). Thus node path weights takes into consideration not only the singular effect a node has on its immediate parent but also the combined prerequisite effect a node would have to a root, B in this case, along a certain path.

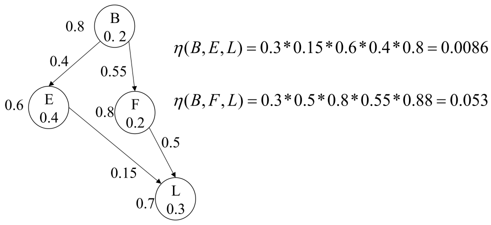

Fig. 2. Calculating prerequisite effect of a node along a path; Node path weight

3 Assessment Approach and Projection Graph

The assessment process is essentially a two step approach. The first main step is the extraction of the relevant concepts from the CSG and is called as *ŞCSG extractionŤ*. A generalized CSG can be vast and therefore it is irrational to process the whole ontology. We define a pruned sub-graph called as *projection graph* which cuts the computation based on a limit on propagated semantic significance. The pruning is achieved by introducing a variable called as the projection threshold coefficient (λ). By varying the threshold coefficient the size of the computable projection graph can be varied and thus the semantic significance. The threshold coefficient can be thought of as a parameter which can inversely set the depth to which the topic has been taught.

3.1 Projection Graph

Given a CSG, $T(C, L)$, with local root concept x_0, and projection threshold coefficient λ, a projection graph $P(x_0, \lambda)$ is defined as a sub graph of T with root x_0 and all nodes x_t where there is at least one path from x_0 to x_t in T such that node path weights $\eta(x_0, x_t)$ satisfies the condition: $\eta(x_0, x_t) \geq \lambda$.

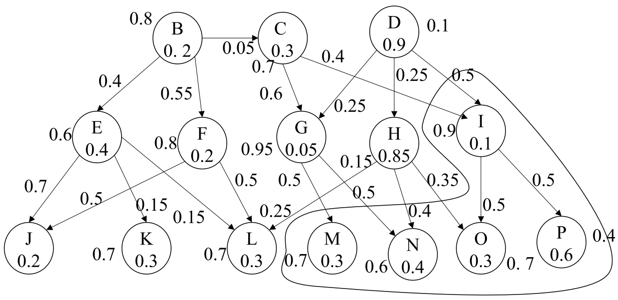

Fig. 3. Projection of concepts B, D and overlapping region

The projection set for x_0 is $[x_0, x_1, ..., x_t]$ represented as $P(x_0, \lambda) = [x_0^{x_0}, x_1^{x_0}, ..., x_t^{x_0}]$ where x_i^j represents the i^{th} element of the projection set of node j. The projections for B and D and the overlapping region of their projections is shown in Figure 3. All nodes that satisfy the condition of node path weights greater than threshold coefficient (through any path) are included in the projection.

4 Assessment Parameters

The second step in the assessment approach is to apply the parameters algorithms to the extracted projection graph to obtain the assessment values.

4.1 Coverage

The coverage of a question gives a cumulative prerequisite effect of the projection graph on the knowledge required to answer a particular question. Coverage of a concept is a direct indicator to the scope of the question in context of the concept space of the course. Formally,Ş*coverage of a node x_0 with respect to the root node r is defined as, the product of the sum of the node path weights of all nodes in the projection set $P(x_0, \lambda)$ for the concept x_0 , and the incident path weight $\gamma(r, x_0)$ from the root rŤ.* If the projection set for concept node x_0, $P(x_0, \lambda)$ is given by then the coverage for node x_0 about the ontology root r is defined as,

$$\alpha(x_0) = \gamma(r, x_0) * \sum_{m=0}^{n} \eta(x_0, x_m) \tag{2}$$

where $\gamma(r, x_0)$ is called as the *Incident Path Weight* and

$$\gamma(r, x_0) = \frac{\eta(x_0, x_n)}{W_s(x_n)} = \frac{\eta(x_0, x_n)}{\eta(x_n, x_n)} \tag{3}$$

Total coverage of multiple concepts in a problem given by set $[C_0, C_1, ..., C_n]$ is, $\alpha(T) = \alpha(C_0) + \alpha(C_1) + ... + \alpha(C_n)$.

The node path weight defines the prerequisite effect of a node to its designated root. Therefore the summation of the node path weights of all the nodes in the projection set gives the cumulative prerequisite effect of the nodes in the projection graph on their respective mapped concept roots. The concepts in the projection graph in turn are the concepts which are required to understand a particular concept, controlled by the threshold coefficient. The *coverage* is thus, the amount of knowledge required to answer or rather understand a particular concept.

4.2 Diversity

Diversity is calculated by measuring the effect of common and uncommon prerequisite concepts from the projections of the mapped concepts. Diversity is formally defined as Ş*the ratio of summation of node path weights of all nodes in the non-overlapping set to their respective roots, and the sum of the summation of node path weights of all nodes in the overlap set and summation of node path weights of all nodes in the non-overlap set.*Ť Consider a problem maps to set of concepts $C = [C_0, C_1, ..., C_n]$ with projections $P(C_0, \lambda), P(C_1, \lambda), ..., P(C_n, \lambda)$ and the non-overlapping and overlapping sets given by $N = [N_0, N_1, ..., N_p]^i$ and $O = [O_0, O_1, ..., O_q]^j$ where i and j are the local root parents of any element from N and O and $\forall i, j \in C$. Diversity is given by,

$$\Delta = \frac{\sum_{m=1}^{p} \eta(i, N_m^i)}{\sum_{m=1}^{p} \eta(i, N_m^i) + \sum_{m=1}^{q} \eta(j, O_m^j)} \tag{4}$$

4.3 Conceptual Distance

Conceptual distance is a measure of distance between two concepts with respect to the ontology root. Alternatively conceptual distance measures the similarity between two concepts by quantifying the distance of the concepts from the ontology root. Formally it is defined as *Şthe log of inverse of the minimum value of incident path weight (maximum value of threshold coefficient) which is required to encompass all the mapped concepts from the root conceptŤ.* The conceptual distance parameter is designed in such a way that it should be sensitive to the depth of the concepts. Hence it is a function of maximum threshold coefficient required to cover all the nodes from the ontology root. Incident path weight (γ) of a concept to the root is equivalent to the threshold coefficient (λ) required to encompass the node. If question asks concept set $C = [C_0, C_1, ..., C_n]$ then the conceptual distance from the root concept r is,

$$\delta(C_0, C_1, ..., C_n) = \log_2\left(\frac{1}{min[\gamma(r, C_0), \gamma(r, C_1), ..., \gamma(r, C_n)]}\right) \quad (5)$$

References

1. 2004 MIT OCW Program Evaluation Findings Report (June 2006). $http : //ocw.mit.edu/NR/rdonlyres/FA49E066 - B838 - 4985 - B548 - F85C40B538B8/0/05_Prog_Eval_Report_Final.pdf$
2. Edmondson, K., Concept mapping for Development of Medical Curricula. Presented at the annual meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993). 37p.
3. Li, T and S E Sambasivam.ăQuestion Difficulty Assessment in Intelligent Tutor Systems for Computer Architecture.ăIn The Proceedings of ISECON 2003, vă20 (San Diego): ğ4112. ISSN:ă1542-7382. (Also appears in Information Systems Education Journal 1:ă(51). ISSN:ă1545-679X.)
4. Rita Kuo, Wei-Peng Lien, Maiga Chang, Jia-Sheng Heh, Difficulty Analysis for Learners in Problem Solving Process Based on the Knowledge Map. International Conference on Advanced Learning Technologies, 2003, 386-387.
5. Silva, L., Oliveira, J.P., (2004). Adaptive Web Based Courseware Development using Metadata Standards and Ontologies. AH 2004, Eindhoven.