A Propagation-based Method of Estimating Students' Concept Understanding

Rafael López-García, Makoto P. Kato, and Katsumi Tanaka

Department of Social Informatics, Graduate School of Informatics, Kyoto University, Kyoto, Japan
{rafael.lopez, kato}@dl.kuis.kyoto-u.ac.jp
tanaka.katsumi.85e@st.kyoto-u.ac.jp

Abstract. In this paper, we introduce a method to estimate the degree of students' understanding of concepts and relationships while they learn from digital text materials online. To achieve our goal, we first define a semantic network that represents the knowledge in a material. Second, we define students' behavior as the sequence of relationships they read in the material, and we create a probabilistic model for relationship understanding. We also create inference rules to include new relationships in the network. Third, we simulate the propagation of the new concept understanding through the network by using a method based on Biased PageRank, extending it with a method to represent prior knowledge and weighting the contribution of every concept according to the uniqueness of its relationships. Finally, we describe an experiment to compare our method against a method without propagation and a method in which propagation is inversely proportional to the distance between concepts. Our method shows significant improvement compared to the others, providing evidence that propagation of concept understanding through the entire network exists.

Keywords: Learning Data Analytics, Concept Understanding, Biased PageRank

1 Introduction

The use of online digital text materials and Virtual Learning Environments (VLEs) like Moodle [1] in traditional classrooms has been increasing in the recent years. One advantage of such technologies is that we can capture students' behavior via network while they are learning. The analysis of this behavior allows instructors to evaluate students' learning as they read, allowing students with different prior knowledge and understanding to receive personalized materials [2] on the fly.

The main goal of this paper is to provide a method to estimate students' understanding from their behavior while they read digital learning materials, either uploaded by the teacher to a VLE or found in an arbitrary website. One advantage of using this method for students' evaluation is that it happens immediately, while traditional methods such as comprehensive quizzes or interviews consume much of instructors' and students' time and effort. Our idea is also applicable to Massive Open Online Courses (MOOCs) [3] as many of them include reading assignments as well.

For that purpose we must: (1) show in a knowledge model which information the student can acquire from those materials, (2) when a student reads an atomic unit of text (i.e.: sentence), reflect the immediate effect of this behavior in the corresponding region of the model, and (3) simulate how this change in a small region affects the rest of the model, this is, how the new understanding propagates through all the concepts.

The first problem is finding a method to represent the knowledge in the materials and the students' understanding. Terms such as "knowledge" and "understanding" are very broad, as they include facts, procedures, principles and some other categories [4]. In this research, we target a set of concepts and their relationships, as concepts are the base of many of the other categories and they have been widely studied in the research literature [5,6] too. Moreover, well-known tools such as semantic networks are suitable to represent them.

The second problem is finding an accurate way to represent how the understanding of a relationship changes when the student reads an atomic unit of text that corresponds to it. A simple approach is marking in the network which relationships have been read. However, we cannot ensure that a student retains the relationship by reading it just once, so we discuss a probabilistic approach as an alternative.

The third problem is deciding how this change affects the understanding of the rest of the concepts and relationships in the network, this is, how understanding propagates through them. Piaget's constructivism describes how internal representations of knowledge are rebuilt by two operations: assimilation and accommodation. [7]. In addition, Vygotsky's Zone of Proximal Development (ZPD) shows how new knowledge is better acquired when it is closely related to the one that the student already has [8]. Nevertheless, these theories do not include mathematical methods to represent those processes with numbers. In this paper, we try to solve this problem by using a variant of the Biased PageRank formula [9] that incorporates the idea of prior knowledge for some concepts in the graph.

Finally, we evaluate the effectiveness of our method by comparing it against a baseline in which propagation does not happen and another baseline in which propagation is naïvely calculated according to the shortest distance between concepts. For that purpose, we performed an experiment in which different people in a crowdsourcing system had to read some learning materials. We applied their data to our understanding models, and found that our method outperforms both baselines.

The main contribution of this paper is that, to the best of our knowledge, our method is the first in estimating the degree of concept and relationship understanding by using the sequence of the relationships read from a text document, as well as the first in proposing a method to measure the propagation of concept understanding and in setting the baselines for its evaluation.

The rest of this paper is as follows: Section 2 reviews the related work. In Section 3 we discuss our proposed method. In Section 4 we propose an experiment to evaluate the model, and we discuss the results. In Section 5 we conclude and in Section 6 we propose the future work.

2 Related Work

2.1 Predicting Students' Performance

For more than two decades, researchers have been publishing papers on using data mining techniques on learning data to predict students' performance [10,11]. Much of that research uses classification methods to predict success after an academic year [12,13] or finding what students will drop out [14]. They use demographic data such as sex and age, and academic data such as grades in exams or course subjects, rate of completion of activities, etc. However, our problem focuses on performance in a very short period and their data is not appropriate for our task.

The two major statistical models for estimating students' knowledge are Bayesian Knowledge Tracing [15] and Learning Factor Analysis [16], and they present several differences with our method. First, they focus on understanding procedural knowledge, while we focus on conceptual knowledge. Second, our method can be used in non-academic environments, as we only need to record the text that the student reads from a file or webpage (e.g. by mouse tracking), while theirs require interactions with an Intelligent Tutoring System. Third, their methods require long term interactions and repetitions of the same task, as they analyze data such as the number of mistakes before learning a skill, etc., while we can estimate the degree of understanding while the student is reading.

Actually, many approaches in the literature are non-viable in an actual higher education classroom as methods require (1) technology that is not common in most institutions, such as the Intelligent Tutoring System, or (2) input that is not always available. For instance, it is possible to use students' vocabulary to predict the quality of their answers [17], although this would require writing activities. Something similar happens in the estimation of reading comprehension. It can be achieved by using students' oral fluency [18], but that would require them to read aloud. Eye tracking can be used instead [19], although such technology is non-viable in a real classroom either.

2.2 Semantic Networks

A s*emantic network* [20] is a way of knowledge representation formalized by a set of nodes representing concepts and a set of edges representing the relationships between those concepts. There are several types, such as *conceptual graphs* [21] and *simple concept graphs* [22]. In any case, semantic networks simply offer a picture of the knowledge in one instant, and therefore they are not sufficient to show the degree of students' understanding and its change. For that purpose, we use other mechanisms explained in Section 3.

Another problem of the semantic networks is the vast number of possible labels for the relationships. Some researchers addressed the need to restrict them [23]. In our work, we give a special treatment to IS_A relationships as their child nodes can always inherit the relationships of the parent, but we do not give any treatment to all the other labels.

3 Approach

3.1 Representation of Understanding

According to Gagné's instructional theory [5], a concept is a classification of things by using either physical features (concrete concept) or associations with other concepts (defined or relational concept). We understand concepts by building these associations (relationships), and this task involves linguistic operations and intensive use of prior knowledge.

In this research, we assume that the person or system assessing the student is interested in a set of concepts $C = \{c_1 \ldots c_n\}$ and a set of labeled, directed relationships R between them. Let L be the set of all the possible labels, we define the relationships as $R \subseteq C \times L \times C$. The triplet $(c_i, l, c_j) \in R$ with $c_i, c_j \in C$ is notated as r_{ij}.

There are several semantic networks capable of reflecting knowledge by using concepts and relationships. In this paper we define the Learning Material's Concept Graph (LMCG) as the representation of the knowledge expressed in the target learning materials. LMCGs are a variant of Simple Concept Graphs (SCG) [22] in which we do not include objects, hypergraphs or bipartite graphs, as these elements are unnecessary for our goal. Open relation extraction software [24] can be used to generate them automatically for arbitrary Web documents, although instructors can manually create more accurate ones for their own materials. We define a LMCG as $G = (C, R, L, W)$, where C is a set of nodes representing the concepts in the material, R is the set of labeled, directed edges representing the relationships between those concepts, and L is the set of labels. The set W, which does not exist in the original definition of SCG, assigns a weight to each relationship. Given a relationship $r_{ij} \in R$, w_{ij} represents how much c_i contributes to understand c_j. Fig. 1 shows a small example of LMCG with four concepts (c_1 to c_4) connected by three relationships labeled as $l_{2,1}$, $l_{3,2}$ and $l_{4,3}$ and weighted by $w_{2,1}$, $w_{3,2}$ and $w_{4,3}$ respectively.

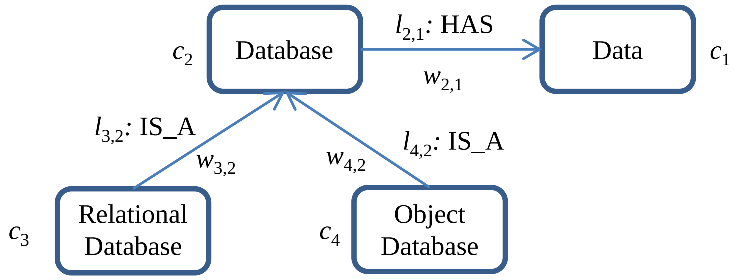

Fig. 1. Small example of LMCG.

Although relationships have direction, we assume that each relationship always has an "inverse" relationship that goes in the opposite direction, and if the student learns one, the inverse is automatically learned as well. For example, if we have the relationship "Database" HAS "Data", there is also the inverse relationship "Data" PART_OF "Database". We do not show these relationships in figures for simplicity.

Note that the content of a LMCG is completely independent from the actions of the students who read those materials. Therefore, it would only change if the instructor changes the materials themselves or the weights of the relationships.

Once concepts and relationships have been defined, we can mathematically define the degree of understanding of a concept as a function u: $C \times T \rightarrow [0, 1]$, in which T is a discrete representation of the time. This understanding degree changes as the student learns from a material. It is also possible that students had acquired some prior knowledge (pk) about the concepts in the past. We define the function pk: $C \rightarrow [0, 1]$ to represent it. One feature of pk is that it does not change during the reading process.

Regarding relationships, we cannot ensure that every student automatically understands every relationship they read in the text. Therefore, we need to define the function rund: $R \times T \rightarrow [0, 1]$ to estimate this relationship understanding. This relationship understanding changes as students read.

3.2 Assumptions on Understanding

Based on the characteristics of knowledge and understanding we stated in the previous sections, the first assumption of our method is that computation of the understanding of a concept must be based on (1) the prior knowledge on that concept, (2) the understanding of the relationships including that concept, and (3) the understanding of the other concepts included in those relationships. Since this method is recursive, we assume that understanding of a concept also depends on the understanding of concepts that are not directly connected to it. We call this effect propagation, and it leads to our first research question:

Research question 1: Does concept understanding propagate to concepts not directly connected in the concept graph, or does it only come from the prior knowledge of the given concept and the understanding of directly connected concepts?

Regarding relationship understanding, every relationship has a weight that changes its contribution to the understanding of a concept. This leads to our second question:

Research question 2: Do all the relationships in which a concept participates contribute in the same way to its understanding or is there a different contribution?

Furthermore, given a set of statements, humans often apply inference rules to them in order to obtain additional knowledge. Many of these rules depend on the domain of the text or on operators we do not provide in our LMCG (e.g.: complement of a set, choosing some elements of a set, etc. [25]). However, we can at least find two rules that are applicable to every domain. These are the two rules about relationship inheritance based on IS_A relationships:

$$\text{if } (c_i, l, c_j) \in R \wedge (c_h, IS_A, c_i) \in R \text{ then add } (c_h, l, c_j) \text{ to } R \tag{1}$$

$$\text{if } (c_i, l, c_j) \in R \wedge (c_h, IS_A, c_j) \in R \text{ then add } (c_i, l, c_h) \text{ to } R \tag{2}$$

For example, if we have $r_{1,2}$ “UNIX” IS_A “O.S.” and $r_{2,3}$ = “O.S.” MANAGES “Hardware”, then we can add $r_{1,3}$ “UNIX” MANAGES “Hardware”. By using these rules, we can create an extended version of the LMCG that includes every relationship of a parent concept in the successors of a hierarchy of IS_A relationships. Considering this, we formulate our third research question:

Research question 3: Do inference rules increase the accuracy of our prediction or should our model reflect the relationships explicitly written in the text?

3.3 Calculation of Relationship Understanding

As, in our model, relationship understanding is the previous step to concept understanding, we first present this part.

Most of the sentences in learning materials state relationships between concepts. Therefore, we can see the reading process as a sequence of steps in which the student understands these relationships. For example, let r be the relationship "Computer HAS C.P.U.". In a given instant t, the student did not know about it, so rund(r, t) = 0. Now, let us imagine that in the instant t + 1 the student reads r. Since students are not perfect at comprehension, we assume that this understanding is done with probability $p \in [0, 1]$. If this is the case, the equation of the relationship understanding for any relationship r_{ij} in the LMCG is as follows:

$$\mathrm{rund}\left(r_{ij}, t+1\right) = 1 - \left(1 - \mathrm{rund}\left(r_{ij}, t\right)\right)(1-p) \tag{3}$$

If we assume that students are perfect at understanding relationships, we have $p = 1$ and therefore rund(r_{ij}, t + 1) = 1. If this is not the case, the students need to read again the same relationship in later instants until they can fully understand it.

Let us now consider the inference rules presented in the equations (1) and (2). The added relationships must also be recursively updated accordingly by using equation (3), although the probability that they are understood is not exactly p, as it also depends on the understanding of the IS_A relationships. Let p^t_{ij} be the probability applied to relationship r_{ij} in an instant t, the probability to be applied to r_{hj} and r_{ih} is:

$$p_{hj}^{t} = p_{ij}^{t}\mathrm{rund}(r_{hi}, t-1), \forall r(c_h, IS\_A, c_i) \epsilon G \tag{4}$$

$$p_{ih}^{t} = p_{ij}^{t}\mathrm{rund}\left(r_{hj}, t-1\right), \forall r\left(c_h, IS\_A, c_j\right) \epsilon G \tag{5}$$

For example, if a student knows $r_{1,2}$ = "UNIX" IS_A "O.S." with rund($r_{1,2}$,t) = 0.5 and she now reads $r_{2,3}$ = "O.S." MANAGES "Hardware" with probability of understanding $p^t_{2,3}$ = 0.75, she also has a probability of understanding $r_{1,3}$ = "UNIX" MANAGES "Hardware" of $p^t_{1,3}$ = 0.75 · 0.5 = 0.375.

3.4 Calculation of Concept Understanding

Once a relationship has been read, we must simulate how this understanding propagates to the other concepts that the student has learned. For such simulation, we use a variant of the Biased PageRank formula [9], whose equation is as follows:

$$\mathbf{r} = \alpha\mathbf{Tr} + (1-\alpha)\mathbf{d} \tag{6}$$

In this equation, **r** is the vector that contains the PageRank values, **T** is called the transition matrix and **d** is the vector containing the bias. The solution to this equation is often calculated by an iterative process called the Jacobi method [26].

In our case, the transition matrix is based on the relationship understanding we calculated in section 3.3 and the weights of the relationships we explain in section 3.5. The static part of the equation is based on the idea of prior knowledge we stated in section 3.1. So our equation becomes as follows:

$$u\left(c_j, t' + 1\right) = \alpha \sum_i \frac{\operatorname{rund}(r_{ij}, t') w_{ij} u(c_i, t')}{\operatorname{in}(c_j)} + (1 - \alpha)\operatorname{pk}\left(c_j\right) \tag{7}$$

In this equation, in: $C \rightarrow \boldsymbol{N}$ is a function that returns the number of relationships pointing to a given concept. The parameter $\alpha \in [0, 1]$ balances the contribution of the prior knowledge against the contribution of the propagation in the graph. In order to give more contribution to propagation than to the prior knowledge, we use $\alpha = 0.9$. Note that we use t' instead of t because relationship understanding and propagation of concept understanding happen in two different timelines.

One problem in the previous equation is that, in our model, the concepts that have been marked as previously known with $\operatorname{pk}(c_i) = 1$ should not be affected by the understanding propagation, as we know that students know them perfectly well. In order to achieve this, when we have a concept c_i with pre-established prior knowledge, we set $\operatorname{rund}(r_{ii}, t') = 1$ and $w_{ii} = 1$, and then $\operatorname{rund}(r_{ij}, t') = 0$ and $w_{ij} = 0$ for all $j \neq i$ and for all t'. With this, such concepts do not receive understanding from others, but they offer it. This also implies the weights of their relationships in our calculations are not exactly the ones initially defined in the LMCG.

Now, for $|C| = n$, let $\mathbf{u}^{\mathbf{t}'}$ be $[u(c_1, t')\ldots u(c_n, t')]$, let $\mathbf{k}$ be $[\operatorname{pk}(c_1)\ldots\operatorname{pk}(c_n)]$, and $\mathbf{D}^{\mathbf{t}'}$ be the matrix whose $d_{ij}^{t'} = \sum_i \operatorname{rund}\left(r_{ij}, t'\right) w_{ij} u(c_i, t') / \operatorname{in}\left(c_j\right)$, we can express equation (7) as matrices:

$$\mathbf{u}^{\mathbf{t}'+\mathbf{1}} = \alpha \mathbf{D}^{\mathbf{t}'} \mathbf{u}^{\mathbf{t}'} + (1 - \alpha)\mathbf{k} \tag{8}$$

One difference between equations (8) and (6) is that we do not require $\Sigma_i^n u(c_i,t') = 1$, but $\Sigma_i^n u(c_i,t') \leq n$. For that purpose, we normalize the weights by forcing $\Sigma_i w_{ij} = 1$, although we do not require $\Sigma_i \operatorname{pk}(c_i) = 1$ or $\Sigma_i \operatorname{rund}(r_{ij}, t') = 1$. Therefore, we have that $\Sigma_i d_{ij}^{t'} \leq 1$. By allowing $\Sigma_i d_{ij} < 1$, we have a sub-stochastic transition matrix instead of a stochastic one. Hence, further analysis of convergence is necessary. This is done in the Appendix. A limitation of our method is that there has to be at least one concept c_i for which $\operatorname{pk}(c_i) = 1$, as smaller values do not grant convergence. This forces us to be conservative when setting the prior knowledge. In addition, if we start with $u(c_j,0) = \operatorname{pk}(c_j) = 0$ for all j, we can verify that $u(c_j,t') = 0$ for all t', respecting the theory that we need to build on prior knowledge to learn [8].

3.5 Estimation of Weights

Up to this point, for a concept c_j we just required that $\Sigma_i w_{ij} = 1$, but we have not stated anything about how to establish the values of each w_{ij}. The basic approach followed by PageRank [27] is that all weights are equal, this is, w_{ij} = 1 / in(c_j). However, our intuition is that some relationships contribute to the understanding of the students more than the others, so we provide an alternative model that includes this idea.

The model we propose is based on the uniqueness of relationships. We assume that for a certain concept c_j, neighbor concepts with unique relationships are more relevant for the understanding of c_j than neighbors with relationships that are shared by other neighbors. The reason is that the information provided by the former is entirely new for c_j, while the latter just convey information c_j already has. For example, let us assume we want to calculate weights $w_{1,0}$, $w_{2,0}$ and $w_{3,0}$ in the LMCG in Fig. 2:

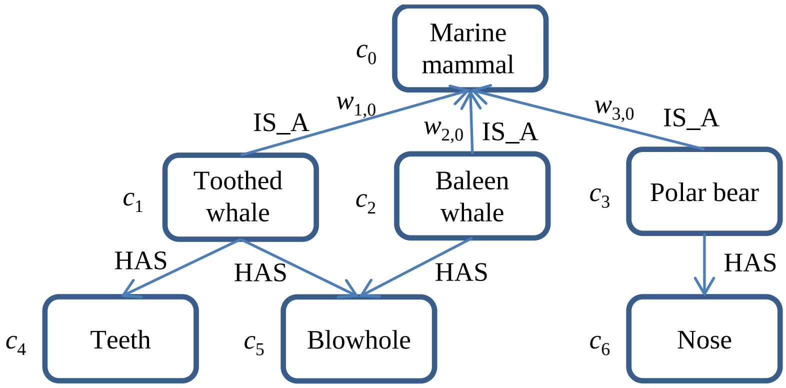

Fig. 2. A LMCG in which we calculate the weights of c_0. The neighbor of c_0 called c_1 has a unique relationship with c_4 and a non-unique one with c_5.

In Fig. 2, the relationship (c_1, HAS, c_5) is not carrying as much information for c_0 because (c_2, HAS, c_5) is also carrying similar information as (c_1, HAS, c_5). Therefore, c_1 and c_2 should distribute among themselves the contribution to the understanding of c_0 that comes from c_5.

Now, let neigh: $C \times G \rightarrow 2^C$ be the function that calculates the neighbors of a concept in a graph, and let edge: $C \times C \times G \rightarrow \{0, 1\}$ be the function that outputs 1 when there is an edge from the first to the second concept and 0 otherwise. Let G' be a version of G in which c_j does not exist. For a certain neighbor c_i, we define its degree of uniqueness δ_{ij} as:

$$\delta_{ij} = \prod_{c_h \in \mathrm{neigh}(c_i, G')} \frac{1}{\sum_{c_x \in \mathrm{neigh}(c_j, G)} \mathrm{edge}(c_x, c_h, G')} \tag{9}$$

The division calculates the degree of uniqueness of a certain relationship r_{ih} among the neighbors of c_j, while the product aggregates the uniqueness of all the relationships of c_i without penalizing concepts with few relationships. We use G' instead of G to avoid counting c_j itself as one of the c_h. After calculating the aggregated uniqueness of each neighbor of c_j, we still need to normalize them, so we can have $\Sigma_i\, w_{ij} = 1$. To achieve this, we just divide the aggregated uniqueness of every neighbor of c_j by the summation of all the aggregated uniqueness of all the neighbors of c_j:

$$w_{ij} = \frac{\delta_{ij}}{\sum_{c_x \in \text{neigh}(c_j, G)} \delta_{xj}} \quad (10)$$

Back to the example in Fig. 2, we have that $\delta_{1,0} = \delta_{2,0} = 0.5$ because c_1 and c_2 share a relationship with c_5, but $\delta_{3,0} = 1$ because the only relationship of c_3 in *G'* is unique. As $\Sigma_x\ \delta_{x0} = 2$, the normalized weights would be $w_{1,0} = w_{2,0} = 0.25$, and $w_{3,0} = 0.5$.

4 Experiment

4.1 Experimental Setup

In order to answer our research questions, we performed an experiment in which we measure the understanding of several students in the Japanese crowdsourcing service Lancers [28], and we compared their results against two baselines and several variants of our method. This section explains the experiment.

Procedure. The main idea of our method is that concept understanding propagates through the graph according to the relationship understanding in every moment. Nevertheless, it is not feasible to ask every student about every relationship and concept as they read sentences of a text to observe the change. Furthermore, tracking the complete behavior of the students would require installing software in their computers, but this approach is non-viable in crowdsourcing. In order to solve these problems, we only ask about a few concepts and we simulate students do not understand some relationships by excluding them from their LMCG. We rewrote the text accordingly, so it does not lose coherence. We assume all the other relationships are read carefully and sequentially. We expect the answers of the students who read the original material to be better than the ones of the students who read the version we have modified.

As we want to generalize the result of our experiment, we repeat this procedure with 6 different learning materials of high school and first year university levels. We created materials about Computer Science, Biology and Chemistry because they focus on conceptual knowledge, while other topics such as Mathematics or Physics may focus too much on procedural knowledge. In addition, for reproducibility of the experiment, the topics we chose are cross-cultural, while other topics such as Literature or History may be localized to the place in which they are taught. For each topic, we chose two different "subtopics" in order to have LMCGs with different topology. We originally created the 6 texts in English with similar length, although the participants in the Japanese crowdsourcing site were shown a translation into their language, making the most technical documents a bit longer due to explanation of acronyms, etc.

For each subtopic, the procedure is as follows. We choose a target concept c_0 to evaluate its understanding. Then, we choose another 3 concepts c_1, c_2 and c_3, located in different points of the LMCG. With this, we can formulate 3 questions asking the relationships between c_1 and c_0, c_2 and c_0, and c_3 and c_0 respectively. This ensures that the student will have to traverse several relationships in the LMCG in order to answer correctly to all the questions, providing a more accurate estimation of the understanding of c_0. Now, given the original text t_0, we create a variant t_1 by removing relationships. These relationships are normally located between c_0 and c_1, c_2 or c_3, and we choose them trying to maximize the expectation of disconnecting a possible concept with prior knowledge from c_0, c_1, c_2 or c_3, or at least disabling the shortest path to these concepts. In order to traverse more relationships of a LMCG, we repeat this procedure for 3 different c_0. Therefore, apart from the original text t_0 we get 3 variants t_1, t_2 and t_3. Table 1 shows the statistics of the texts and LMCGs. Columns "$|C|$" and "$|R|$" contain the number of concepts and relationships in the LMCG, respectively, and the column "Rels. cut" shows the relationships removed in each variant of the original text. Both "|R|" and "Rels. cut" include inverse relationships.

Table 1. Statistics of the texts and the LMCGs

Subtopic	Words (Eng.)	Chars (Jap.)	|C|	|R|	Rels. cut (t_1/t_2/t_3)
Comp. Arch.	368	1162	34	110	4/18/10
Databases	363	1230	28	76	8/6/14
Genetics	396	976	25	66	6/8/6
Cetaceans	367	882	34	70	4/4/6
Compounds	389	883	33	92	10/16/6
Solutions	355	854	46	120	12/8/22

For our experiment, we created a web application in which the students are shown one of the above texts for 15 minutes, and they have to answer a questionnaire about it in another period of 15 minutes. In order to avoid cheating, the website does not let the students go back to the text when they reach the page of the questionnaire. Two "captcha" questions are also included in each questionnaire, in addition to the text areas in which the students have to write the relationship c_1-c_0, c_2-c_0 and c_3-c_0 respectively. Finally, if a student participated twice in the same subtopic the prior knowledge would be different, so we use a mechanism based on cookies in order to ensure that this does not happen, although one student can participate in the 6 subtopics.

Ground truth. We let 90 students participate in each subtopic. We divided them in 3 groups of 30 depending on the target concept assigned to them. Each group is again sub-divided in two groups of 15, one receiving the original text and the other receiving the text in which relationships are removed.

In order to grade students' answers, we hired two evaluators for each topic. These evaluators are master and Ph.D. students majoring a related subject at our institution. Each evaluator was asked to grade each answer in a 5-level likert scale, where 1 is assigned to blank or completely unrelated answers, while 5 is assigned to perfect answers. Intermediate values are assigned according to the number of missing relationships between the target concept and the other concept. Another task we requested to each evaluator is grading the contribution of concepts c_1, c_2 and c_3 to the understanding of c_0. These values were also given in a 5-level likert scale, where 1 means no contribution and 5 means very high contribution. Let w_{i0}, where i = 1..3, be these weights, and let score(c_i,c_0) be the score assigned to the answer about the relationship between c_i and c_0, we can calculate the ground truth of the understanding of a target concept as in equation (11). These values are normalized so $\hat{u}: C \rightarrow [0, 1]$.

$$\hat{u}(c_0) = \sum_{i=1}^{3} \text{score}(c_i, c_0) w_{i0} \tag{11}$$

Finally, since it is not viable to calculate the prior knowledge of each participant in every concept without unveiling the content of the text, we have decided to set it beforehand. For that purpose, we took a list of the 5000 most common words in English [29] and we decided that the terms appearing in that list would receive pk(c_i) = 1, while the others would receive pk(c_i) = 0. However, we found out that some of the terms appearing in the list are often used with a different meaning to the one in our text (e.g.: "class", "object", and so on in Object DBMS), so we had to apply human supervision to the concepts based on the following rules:

- If the concept has a meaning for daily life (e.g.: "class"), but the student probably has not studied the specialized meaning, set pk(c_i) = 0.
- If students may have studied the specialized meaning of the concept in school (e.g.: "character" in genetics), but it hardly ever appears in daily life, set pk(c_i) = 0.
- Otherwise, leave pk(c_i) = 1.

In addition we included some other terms because they are simple variants of the ones in the list (e.g.: "swimmer" does not appear but "swim" and "swimming" do).

Baseline. In order to know if propagation really affects understanding, we created the baseline in a way that the understanding of c_0 can only come from the prior knowledge of c_0 itself and the prior knowledge of the concepts that can reach c_0 in one hop, and we assume that the relationships between these concepts and c_0 have been read. Therefore, this baseline considers that propagation does not exist. Let points: $C \rightarrow 2^C$ be the function that calculates the concepts having an edge pointing to the given concept. We can define this baseline u': C → [0,1] as:

$$u'(c_0) = \max\left(\text{pk}(c_0), \frac{\sum_{c_i \in \text{points}(c_0)} \text{pk}(c_i)}{|\text{points}(c_0)|}\right) \tag{12}$$

Notice that this baseline outputs the same value in the original LMCG and in the version in which we remove relationships, as these relationships are not neighbors of c_0 with prior knowledge.

We also created a naïve method in which understanding propagation is calculated based on the distance between the target concept and the concepts with prior knowledge. Let dist: $C \times C \rightarrow \mathbf{N}$ the function returning the distance in hops between two concepts and prior: $G \rightarrow 2^C$ be the function that returns all the concepts of a graph with some prior knowledge assigned. Our function u'': C → [0,1] is as follows:

$$u''(c_0) = \max\left(\text{pk}(c_0), \frac{\sum_{c_i \in \text{prior}(G)} \frac{1}{\text{dist}(c_0, c_i)}}{|\text{prior}(G)|}\right) \tag{13}$$

4.2 Experimental Results

We collected 540 answers from students, but we invalidated 5 of them because the students answered in English, remaining 535. As we have two evaluators per task, we have 1070 evaluations in total. Agreement between reviewers was measured by using

Cohen’s kappa with equal weights [30] in the statistical software R. The agreement for the evaluators was $\kappa = 0.68$ in Biology, $\kappa = 0.45$ in Computer Science and $\kappa = 0.28$ in Chemistry, meaning fair to substantial agreement. We observed that, when κ is low, the reason is that one evaluator gives higher scores than the other (average difference in the 5-level likert scale of 0.122 in Biology, 0.489 in CS and 0.985 in Chemistry).

We grouped all the normalized grades to the same question together by taking the average of the scores of all the students. This gives 36 values in total, with $\mu = 0.466$ and $\sigma = 0.088$. We cannot reject that the distribution of these values is normal (Shapiro-Wilk W=0.947, p-value=0.108).

We checked for a significant difference between the grades for the original texts ($\mu = 0.527$, $\sigma = 0.081$) and the grades for the texts in which relationships have been cut ($\mu = 0.425$, $\sigma = 0.061$). Student’s T test found a significant difference between averages (t = 4.258, p-value = 0.000), while the Fisher’s F-test did not find a significant difference of variances (F = 1.726, p-value = 0.270). Significant average differences are also found if we analyze independently for each topic.

The way in which we compare our method to the baseline is through the analysis of correlations with the ground truth. For each result, we calculated Pearson’s coefficient r, Spearman’s rank coefficient ρ and Kendall’s rank coefficient τ. The baseline method without propagation performed quite poorly (r = -0.186, p-value = 0.278, ρ = -0.129, p-value = 0.445 and τ = -0.081, p-value = 0.549), not only showing a negative correlation with the ground truth, but also returning p-values bigger than 0.1, so we cannot deny that there may not be any correlation. For the naïve propagation method based on distance, we can find that some correlation coefficients are bigger than 0 while others are not (r = -0.039, p-value = 0.821, ρ = 0.040, p-value = 0.734 and τ = 0.046, p-value = 0.784), although with p-values bigger than 0.1 we cannot discard that there is no correlation either.

We tried 4 variants of our method by combining 2 features: equal weights vs. uniqueness-based weights (explained in section 3.5) and using inference rules vs. no inference rules (explained in sections 3.2 and 3.3). We found out that all the methods achieve their best performance when relationship understanding is perfect (p = 1), although which variant performs the best (equal weights vs. uniqueness-based weights + inference rules) depends on the correlation coefficient we use, as shown in Table 2.

Table 2. Correlations of the variants of our method (rel. understanding probability p = 1)

Correlation coefficient	Equal weights	Uniqueness-based weights	Inferences + equal weights	Inferences + uniqueness-based weights.
Pearson’s r	0.392	0.436	0.304	0.459
(p-value)	(0.018)	(0.008)	(0.071)	(0.049)
Spearman’s ρ	0.380	0.366	0.197	0.378
(p-value)	(0.025)	(0.030)	(0.245)	(0.026)
Kendall’s τ	0.302	0.279	0.137	0.294
(p-value)	(0.010)	(0.017)	(0.241)	(0.012)

In order to verify whether these methods perform better than the others, we compared pairs of correlations by using Fisher's r-to-z transformation for dependent samples [31,32] (only available for Pearson's r). Then, we corrected the p-values by using the Holm-Bonferroni method [33]. Table 3 shows the test for the statistic z and the corrected p-values (α = 0.1). All the variants of our method perform better than the baseline without propagation, and 3 of them (equal weights, uniqueness-based weights and uniqueness based weights + inference rules) perform better than the distance-based method. No significant difference occurs between two variants of our method.

Table 3. Differences between correlation pairs (statistic z)

Method	Equal weights	Uniqueness-based weights	Inferences + equal weights	Inferences + uniqueness-based weights.
No propagation (p-value)	-2.858 (0.026)	-3.158 (0.014)	-2.392 (0.072)	-3.316 (0.000)
Distance-based method (p-value)	-2.504 (0.060)	-2.752 (0.033)	-1.930 (0.216)	-2.954 (0,026)
Equal weights (p-value)		-0.955 (0.618)	1.092 (0.618)	-0.712 (0.618)
Uniqueness weights (p-value)			1.393 (0.574)	-0.294 (0.618)
Inference + equal weights (p-value)				-1.242 (0.618)

As we have seen in the table, the variant with inference rules and uniqueness-based weights performs quite well, while using only inference rules does not outperform the distance-based propagation method. The reason is as follows. Given a concept c_i with pk(c_i) = 1, if we remove a relationship nearby, the concept may not be able to propagate its understanding anymore in the original concept graph. However, if we add new relationships by using inference rules, the concept may propagate understanding through them. Fig. 3 illustrates this situation.

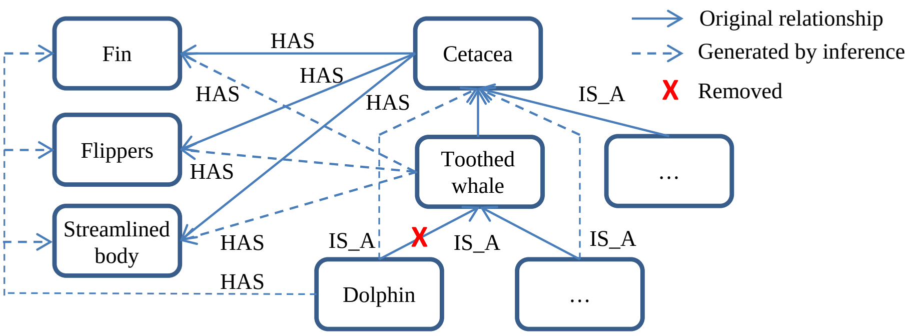

Fig. 3. A hierarchy in which inferred relationships allow propagation of understanding.

The problem now is that the amount of new relationships may be high, and if they all have equal weights, they may be propagating most of the understanding of c_i when they should not propagate that much. Nevertheless, by applying the uniqueness-based weights, we decrease this propagation without removing it completely. This is especially true when there is a large hierarchy of IS_A relationships, as all the ancestors of c_i share most of the relationships that c_i got by inference, so their uniqueness is low.

5 Conclusions

This paper presented a method that allows the estimation of the degree of concept understanding of a student while reading a learning material (text). Our model first reflects the knowledge contained in the material in a semantic network called Learning Material Concept Graph (LMCG). Then, it estimates the understanding by two operations: (1) relationship understanding and (2) concept understanding propagation. The first follows a probabilistic model, and we also considered the addition of inference rules for its calculation. The second is based on the Biased PageRank formula [9]. In order to enhance this formula, we added the notion of prior knowledge and a weighting system to balance the contribution of each concept to the understanding of the others. This system is based on the uniqueness of the relationships of the neighbors of the target concept.

In order to validate our model we performed an experiment with real people in a Japanese crowdsourcing platform [28]. First, we created the ground truth by evaluating their answers to a questionnaire. Then, we proposed a baseline that does not consider propagation of understanding and another in which propagation is calculated based on the distance between the concepts with prior knowledge and the concepts to be evaluated. We also considered four variants of our method: (1) with equal weights, (2) with uniqueness-based weights, (3) with equal weights and inference rules and (4) with uniqueness-based weights and inference rules. Finally, we analyzed the correlation between the ground truth and each model and compared the correlations. The experimental results show that all the variants of our method outperform the baseline with no propagation, suggesting the existence of concept understanding propagation. Moreover, we found that three variants also outperform the distance-based propagation method, and we analyzed why the variant with equal weights and inference rules did not.

6 Future Work

Our future work is the development of a system that registers the prior knowledge and the behavior of each student when they read learning materials on the computer (e.g.: by using mouse tracking), so we can compute individual estimations of their concept understanding. Then we can focus on automatically finding the most suitable materials on the Web, personalizing the search according to each student's concept understanding.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP15H01718 and JP26700009.

We would also like to thank Profs. Masatoshi Yoshikawa and Roi Blanco for their inestimable advice to develop this research.

Appendix

In order to prove the convergence of our method, we remind that in case of having an iterative method with transition matrix $\mathbf{M}$, convergence is proven if $\lim_{y\to\infty} \mathbf{M}^y\mathbf{u} = \mathbf{u}$. Equivalently, to grant we will achieve a steady state distribution, we need to ensure that the biggest dominant eigenvalue of $\mathbf{M}$ is 1. In our case, we need to verify $\alpha\mathbf{Du} + (1-\alpha)\mathbf{k} = \mathbf{u}$. If we rewrite this equation as $\mathbf{u} = \mathbf{Mu}$, and we let $\mathbf{S}$ be the matrix that verifies s_{ii} = pk(c_i) and s_{ij} = 0 for i ≠ j, we have that $\mathbf{M} = \alpha\mathbf{D} + (1-\alpha)\mathbf{S}$. Since it is difficult to prove that the dominant eigenvalue of $\mathbf{M}$ is 1, we will prove that the dominant eigenvalue for $\mathbf{M}^\mathbf{T}$ is 1 instead. Then, we can use the Perron-Frobenius theorem, which states that the dominant eigenvalue is the same for $\mathbf{M}$ and $\mathbf{M}^\mathbf{T}$. We remember that in the concepts where the prior knowledge had been set, we had established that rund(r_{ii},t) = *1* and rund(r_{ij},t) = *0* for all $j \neq i$. By using this, we can see that the dominant eigenvector $\mathbf{u}$ for the matrix $\mathbf{M}^\mathbf{T}$ is precisely $\mathbf{k}$, the vector representing prior knowledge and the eigenvalue for that vector is 1.

As in the case of the original PageRank [27], convergence of is only granted if the transition matrix is (1) irreducible and (2) aperiodic. We know that matrices verifying that one diagonal element is non-zero are aperiodic [34]. This is our case because we had set the pk(c_i) for at least one concept c_i and we also stated that in such case rund(r_{ii},t) = w_{ii} = 1 for all t. The problem is that the transition matrix is not irreducible because we allowed rund(r_{ij},t) = 0 for all j in a concept whose prior knowledge has not been set. However, by using the principle "each relationship has its inverse" in section 3.1, we have that if rund(r_{ij},t) = 0, then rund(r_{ji},t) = 0 too, so the concept c_i is completely isolated from the rest of the graph. In that case, we can just set $u(c_i,t)$ = 0 for all t, and we can remove the concept from the graph, having that the remaining graph is strongly connected and therefore its matrix is irreducible.

References

1. Moodle: https://www.moodle.org. Last accessed 2017/01/15.
2. Dagger, D., Conlan, O. and Wade, V. Fundamental Requirements of Personalised eLearning Development Environments. In World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (2005).
3. Kaplan, A.M., Haenlein, M.: Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. Business Horizons, 59(4), 441-450 (2016).
4. Marzano, R.J. and Kendall, J.S.: The New Taxonomy of Educational Objectives. Corwin Press. (2007)

5. Gagné, R.M.: The Conditions of Learning and Theory of Instruction. (1985)
6. Zentall, T.R., Galizio, M. and Critchfied, T.S.: Categorization, concept learning, and behavior analysis: an introduction. Journal of the Experimental Analysis of Behavior, 78(3), 237-248 (2002).
7. Piaget, J.: L´equilibration des structures cognitives: Problème central du développement. Presses Universitaires de France. (1975)
8. Vygotsky, L.S.: Interaction between Learning and Development. *Mind in Society*, 79-91. Cambridge, MA: Harvard University Press. (1978)
9. Gyöngyi, Z, Garcia-Molina, H and Pedersen, J. Combating Web Spam with TrustRank. Proceedings of the 30th VLDB Conference, 576-587. (2004)
10. Shahiri, A.M., Husain, W. and Rashid, N.A.: A Review on Predicting Student's Performance using Data Mining Techniques. The Third Information Systems International Conference. Procedia Computer Science, 72, 414-422 (2015).
11. Thakar, P. Mehta, A. and Manisha. Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. International Journal of Computer Applications, 110(15), 60-68 (2015).
12. Kabakchieva, D: Predicting Student Performance by Using Data Mining Methods for Classification. Cybernetics and Information Technologies, 13(1), 61-71 (2013).
13. Kotsiantis, S., Pierrakeas, C. and Pintelas, P. Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques. *Applied Artificial Intelligence*, 18(5) 411-426 (2004).
14. Dekker, G.W., Pechenizkiy, M and Vleeshouwers, J.M. Predicting Students Drop Out: A Case Study. Proceedings of 2nd International Conference on Educational Data Mining, 41-50 (2009).
15. Corbett, A.T. and Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. User modeling and User-Adapted Interaction, 4(4), 253-278 (1995.).
16. Cen, H. Koedinger, K. and Juncker, B.: Learning Factor Analysis - A General Method for Cognitive Model Evaluation and Improvement. Proc. of the 8th International Conference on Intelligent Tutoring Systems, 164-175 (2006)
17. Williams, C. and D'Mello, S.: Predicting Student Knowledge Level from Domain Independent Function and Content Words. In V. Aleven, J. Kay and J. Mostow (Eds.) ITS 2010, Part II, LNCS 6095, 62-71 (2010).
18. Fuchs, L.S., Fuchs, D. and Hosp, M.K.: Oral Reading Fluency as an indicator of Reading Competency: A Theoretical, Empirical and Historical Analysis. Scientific Studies of Reading, 5(3), 239-256 (2001).
19. Copeland, L., Gedeon, T., and Mendis, S. Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error. Artificial Intelligence Research, 3(3), 35-48 (2014).
20. Sowa, J.F.: Semantic Networks. In Stuart C. Shapiro Encyclopedia of Artificial Intelligence. (1987)
21. Sowa, J.F.: Conceptual Graphs for a Data Base Interface. IBM Journal of Research and Development, 20(4), 336-357 (1976).
22. Prediger, S.: Simple Concept Graphs: A logic approach. In Muniger and Chein, Conceptual Structures: Theory, Tools and Applications. 6th International Conference on Conceptual Structures ICCS'98, 225-239 (1998).
23. Van der Riet, R.P., Meersman, R.A.: Linguistic Instruments in Knowledge Engineering. Imprint Elsevier Science Ltd (1992).

24. Mesquita, F. Schmidek, J. and Barbosa, D.: Effectiveness and Efficiency of Open Relation Extraction. Proc. of the 2013 Conference on on Empirical Methods in Natural Language Processing, 447-457 (2013)
25. Atzeni, P. Parker Jr., D.S.: Set Containment Inference and Syllogisms. Theoretical Computer Science, 62(1-2), 39-65 (1988).
26. Golub, G.H. and Van Loan, C.F.: Matrix Computations. The Johns Hopkins University Press. (1996)
27. Page, L., Brin, S., Motwani, R. and Winograd, T.: The PageRank citation ranking: Bringing order to the Web. Technical Report. Stanford InfoLab. (1999)
28. Lancers: http://www.lancers.jp. Last accessed 2017/01/09
29. Word frequency data: http://www.wordfrequency.info. Last accessed 2017/01/31
30. Cohen, J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin. 70 (4): 213–220 (1968)
31. Lenhard, W. and Lenhard, A. Hypothesis Tests for Comparing Correlations. https://www.psychometrica.de/correlation.html. 2014. Last accessed 2017/03/15.
32. Eid, M., Gollwitzer, M and Schmitt, M.: Statistik und Forschungsmethoden Lehrbuch. Weinheim: Beltz. (in German) (2011)
33. Holm, S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 6(2), 65–70. (1979)
34. Meyer, C.D.: Matrix Analysis and Applied Linear Algebra, SIAM, ISBN 0-89871-454-0 (2000).