Measuring Prerequisite Relations Among Concepts

Chen Liang [†]   **Zhaohui Wu**[‡]   **Wenyi Huang**[†]   **C. Lee Giles**[†]

[†]Information Sciences and Technology
[‡]Computer Science and Engineering
The Pennsylvania State University
University Park, PA
cul226@ist.psu.edu   zzw109@psu.edu   {wzh112,giles}@ist.psu.edu

Abstract

A *prerequisite* relation describes a basic relation among concepts in cognition, education and other areas. However, as a semantic relation, it has not been well studied in computational linguistics. We investigate the problem of measuring prerequisite relations among concepts and propose a simple link-based metric, namely *reference distance* (**RefD**), that effectively models the relation by measuring how differently two concepts refer to each other. Evaluations on two datasets that include seven domains show that our single metric based method outperforms existing supervised learning based methods.

1 Introduction

What should one know/learn before starting to learn a new area such as "deep learning"? A key for answering this question is to understand what a *prerequisite* is. A prerequisite is usually a concept or requirement before one can proceed to a following one. And the prerequisite relation exists as a natural dependency among concepts in cognitive processes when people learn, organize, apply, and generate knowledge (Laurence and Margolis, 1999). While there has been serious effort in understanding prerequisite relations in learning and education (Bergan and Jeska, 1980; Ohland et al., 2004; Vuong et al., 2011), it has not been well studied as a semantic relation in computational linguistics, where researchers focus more on lexical relations among lexical items (Miller, 1995) and fine-grained entity relations in knowledge bases (Mintz et al., 2009).

Instead of treating it as a relation extraction or link prediction problem using traditional machine learning approaches (Talukdar and Cohen, 2012; Yang et al., 2015), we seek to better understand prerequisite relations from a perspective of cognitive semantics (Croft and Cruse, 2004). Partially motivated by the theory of *frame semantics* (Fillmore, 2006), or, to understand a concept, one needs to understand all the related concepts in its "frame", we propose a metric that measures prerequisite relations based on a simple observation of human learning. When learning concept A, if one needs to refer to concept B for a lot of A's related concepts but not vice versa, B would more likely be a prerequisite of A than A of B. Specifically, we model a concept in a vector space using its related concepts and measure the prerequisite relation between two concepts by computing how differently the two's related concepts refer to each other, or *reference distance* (**RefD**).

Our simple metric **RefD** successfully reflects some properties of the prerequisite relation such as asymmetry and irreflexivity; and can be properly implemented for various applications using different concept models. We present an implementation of the metric using Wikipedia by leveraging the links as reference relations among concepts; and present a scalable prerequisite dataset construction method by crawling public available university course prerequisite websites and mapping them to Wikipedia concepts. Experimental results on two datasets that include seven domains demonstrate its effectiveness and robustness on measuring prerequisites. Surprisingly, our single metric based approach significantly outperforms baselines which use more sophisticated supervised learning. All the datasets are publicly available upon request.

Our main contributions include:

- A novel metric to measure the prerequisite relation among concepts that outperforms existing supervised learning baselines.

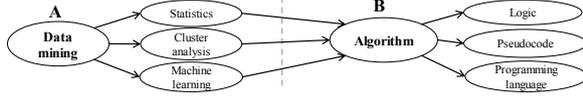- A new dataset containing 1336 concept pairs in Computer Science and Math.

Figure 1: An example of the reference structure for two concepts ("Data mining" and "Algorithm") with a prerequisite relation.

2 Measuring Prerequisite Relations

Our goal is to design a function $f : \mathcal{C}^2 \to \mathcal{R}$ that maps a concept pair (A, B) to a real value that measures the extent to which A requires B as a prerequisite, where $\mathcal{C}$ is the concept space. How should a concept be represented in $\mathcal{C}$? According to the theory of frame semantics, one cannot understand a concept without access to all essential knowledge related to it. Such knowledge can be actually viewed as a set of related concepts. Thus, a concept could be represented by its related concepts in $\mathcal{C}$. For example, the concept "deep learning" may be represented by concepts such as "machine learning", "artificial neural network", etc.

Compared to prerequisites, a more common and observable relation among concepts is a *reference*, which widely exists in various forms such as hyperlinks, citations, notes, etc. Although a single evidence of reference does not indicate a prerequisite relation, a large number of such evidences might make a difference. For example, if most related concepts of A refer to B but few related concepts of B refer to A, then B is more likely to be a prerequisite of A, as shown in Figure 1. In order to measure prerequisite relations, we propose a *reference distance* (**RefD**), which is defined as

$$RefD(A, B) = \frac{\sum_{i=1}^{k} r(c_i, B) \cdot w(c_i, A)}{\sum_{i=1}^{k} w(c_i, A)} - \frac{\sum_{i=1}^{k} r(c_i, A) \cdot w(c_i, B)}{\sum_{i=1}^{k} w(c_i, B)} \quad (1)$$

where $\mathcal{C} = \{c_1, ..., c_k\}$ is the concept space; $w(c_i, A)$ weights the importance of c_i to A; and $r(c_i, A)$ is an indicator showing whether c_i refers to A, which could be links in Wikipedia, mentions in books, citations in papers, etc.

$RefD$ enables several useful properties for the prerequisite relation: 1) normalized: $RefD(A, B) \in [-1, 1]$; 2) asymmetric: $RefD(A, B) + RefD(B, A) = 0$, which means if A is a prerequisite of B then B is not a prerequisite of A; and 3) irreflexive: $RefD(A, A) = 0$, which means A is not a prerequisite of itself. To capture all three possible prerequisite relations

between a concept pair, *RefD* is expected to satisfy the following constraints:

$$RefD(A,B) \in \begin{cases} (\theta, 1], & \text{if B is a prerequisite of A} \\ [-\theta, \theta], & \text{if no prerequisite relation} \\ [-1, -\theta), & \text{if A is a prerequisite of B} \end{cases}$$

where θ is a positive threshold.

Equation 1 provides a general framework to calculate *RefD*. In practice, we need to specify the concept space $\mathcal{C}$, the weight w, and the reference indicator function r.

3 Wikipedia-based RefD Implementation

We now implement *RefD* using Wikipedia. As a widely used open-access encyclopedia, Wikipedia provides relatively up-to-date and high quality knowledge and has been successfully utilized as explicit concepts (Gabrilovich and Markovitch, 2007). Moreover, the rich hyperlinks created by Wiki editors provide a natural way to calculate the reference indicator function r.

Specifically, the concept space $\mathcal{C}$ consists of all Wikipedia articles. $r(c, A)$ represents whether there is a link from Wiki article c to A. For $w(c, A)$, we experiment with two methods:

- *EQUAL*: A is represented by the concepts linked from it ($L(A)$) with equal weights.

$$w(c, A) = \begin{cases} 1 & \text{if } c \in L(A) \\ 0 & \text{if } c \notin L(A) \end{cases}$$

- *TFIDF*: A is represented by the concepts linked from it with TFIDF weights.

$$w(c, A) = \begin{cases} tf(c, A) * \log \frac{N}{df(c)} & \text{if } c \in L(A) \\ 0 & \text{if } c \notin L(A) \end{cases}$$

where $tf(c, A)$ is the number of times c being linked from A; N is the total number of Wikipedia articles; and $df(c)$ is the number of Wikipedia articles where c appears.

4 Experiments

In order to evaluate the proposed metric, we apply it to predicting prerequisite relations in Wikipedia, i.e., whether one article in Wikipedia is a prerequisite of another article. Given a pair of concepts (A,B), we predict whether B is a prerequisite of A or not. Both pairs where A is a prerequisite of

Dataset	Domain	# Pairs	# Prerequisites
CrowdComp	Meiosis	400	67
	Public-key Cryp.	200	27
	Parallel Postulate	200	25
	Newton's Laws	400	44
	Global Warming	400	43
Course	CS	678	108
	MATH	658	75

Table 1: Statistics of CrowdComp and Course Datasets

Domain	MaxEnt[†]	MaxEnt	EQUAL	TFIDF
Meiosis	51	60.2	53	55.7
Public-key Cryp.	67.1	60.3	55.1	57.7
Parallel Postulate	64.7	73.6	70.5	67.9
Newton's Laws	53.9	57.7	63.7	64.6
Global Warming	56.8	50.0	57.4	60.1
Average	58.7	60.4	60.0*	**61.2***

Table 2: Comparison of out-of-domain training accuracies of a MaxEnt classifier and *RefD* using EQUAL and TFIDF weighting. MaxEnt[†] is the number reported by Talukdar et al. (2012). MaxEnt shows the performance of our implementation. * indicates the difference between *RefD* and MaxEnt is statistically significant ($p < 0.01$).

B and pairs where no prerequisite relation exists will be viewed as negative examples.

RefD is tested on two datasets: CrowdComp dataset (Talukdar and Cohen, 2012) and a Course prerequisite dataset collected by us. We compare *RefD* with a Maximum Entropy (MaxEnt) classifier which exploits graph-based features such as PageRank scores and content-based features such as the category information, whether a title of concept is mentioned in the first sentence of the other concept, the number of times a concept is linked from the other, etc. (Talukdar and Cohen, 2012). All experiments use a Wikipedia dump of Dec 8, 2014.

4.1 Results on the CrowdComp Dataset

The CrowdComp dataset was collected using Amazon Mechanical Turk by Talukdar et al. (2012). It contains binary-labeled concept pairs from five different domains, including meiosis, public-key cryptography, the parallel postulate, Newton's laws of motion, and global warming. The label of the prerequisite relation for each pair is assigned using majority vote. Details of the dataset are shown in Table 1.

Following Talukdar et al. (2012), we evaluate

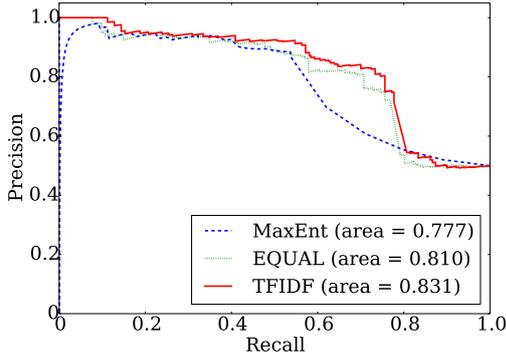|	CS				MATH			
	A	P	R	F	A	P	R	F
MaxEnt	72.8	87.6	53.2	66.1	69.0	78.1	53	63.1
EQUAL	76.4*	80.4	69.9	74.7*	**73.9***	78.4	67.3	**71.9***
TFIDF	**77.1***	82.3	69.1	**75.1***	70.3*	76.3	60.1	66.7*

Table 3: Comparison of in-domain training accuracies, precision, recall, and F1 measure of MaxEnt and *RefD* using EQUAL and TFIDF weighting. * indicates the improvement over MaxEnt is statistically significant ($p < 0.01$).

different methods in a "leave one domain out" manner, where data from one domain is used for testing and data from other four for training. Classes in the training and testing set are balanced by oversampling the minority class. Table 2 lists the accuracies of different methods. In terms of average performance, *RefD* achieves comparable average accuracy as MaxEnt. When TFIDF is used to calculate w, *RefD* performs better than MaxEnt. Also we notice that our implementation of MaxEnt classifier achieves higher accuracy than reported in the original paper, which may be due to the difference between Wiki dumps used. In addition, we can see that there are large differences in performance across different domains, which is mainly due to two reasons. First, the coverage of Wikipedia for different domains may vary a lot. Some domains are more popular and thus edited more frequently, leading to a better quality of articles and a more complete link structure. Second, since the ground-truth labels are collected by crowdsourcing and there is no guarantee for workers' knowledge about a certain domain, the quality of labels for different domains varies.
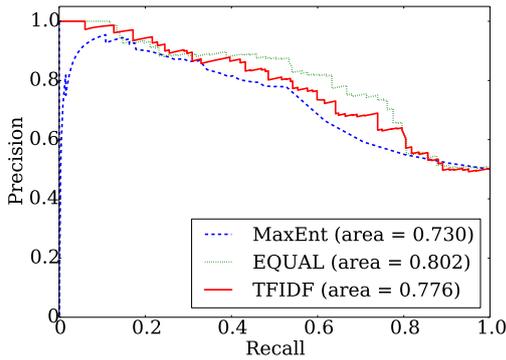
4.2 Results on the Course Dataset

We also built a Course dataset with the help of information available on a university's course website containing prerequisite relations between courses. For example, "CS 331 Data Structures and Algorithms" is a prerequisite for "CS 422 Data mining". We get the prerequisite pairs by crawling the website and linking the course to Wikipedia using simple rules such as title matching and content similarity. In order to get negative samples, we randomly sample 600 pairs using concepts appearing in the prerequisite pairs. All pairs are then checked by two domain experts by removing pairs with incorrect labels. Table 1 lists the information of the dataset.

Evaluation uses in-domain 5-fold cross-

(a) CS

(b) MATH

Figure 2: Comparison of Precision-Recall curves of MaxEnt and *RefD* (using EQUAL and TFIDF weighting) on the Course dataset.
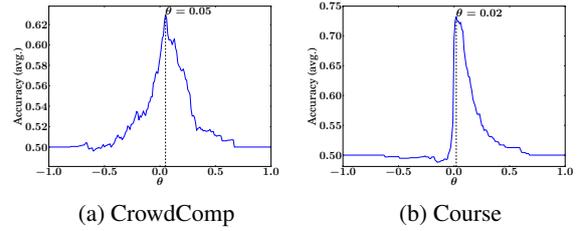

(a) CrowdComp   (b) Course

Figure 3: Average accuracy on two datasets with a given threshold of *RefD* using TFIDF weighting.

a concept using WordNet (Miller, 1995), Explicit Semantic Analysis (Gabrilovich and Markovitch, 2007), or Word2vec embeddings (Mikolov et al., 2013). Incorporating these representations may improve the performance of *RefD*.

4.3 Parameter Analysis and Case Study

Since using *RefD* to predict prerequisites requires setting a threshold θ, we also investigate the relation between the threshold and the performance of prediction, as shown in Figure 3. We can see that a threshold of 0.05 for *RefD* using TFIDF achieves the highest average accuracy on the CrowdComp dataset while a threshold of 0.02 works the best for Course dataset. Empirically we find that a threshold between 0.02 and 0.1 yields a good performance for prerequisite prediction task.

We further explore the performance of *RefD* through a case study for the concept "deep learning" (denoted as c'). Specifically, for any concept c linked from c' we calculate $RefD(c', c)$. Table 4 lists the *RefD* scores for different concepts using *EQUAL* weighting. The concepts on the left have negative *RefD* scores with high absolute values, which means that "deep learning" is a prerequisite of them. Meanwhile concepts on the right have high positive *RefD* scores, which means that "deep learning" requires knowing them first. For example, people may first need to have some knowledge of "machine learning", "artificial intelligence" and "algorithm" in order to learn "deep learning". Also we notice that concepts in the middle have *RefD* scores which are very close to 0, showing that there is no prerequisite relations between these concepts and "deep learning". However, since our *RefD* implementation is based on Wikipedia, it might not give an accurate measure for concepts if they have no Wikipedia articles or their articles are too short to provide an encyclopedic coverage, such as "discriminative model" and "feature engineering".

validation and classes are balanced by over-sampling the minority class. Table 3 lists the performance comparison of different methods on accuracy, precision, recall and F1 score. We can see that *RefD* outperforms MaxEnt in terms of accuracy, recall, and F1 score on both CS and MATH domain. Because MaxEnt relies on many features but there are only limited distinct positive samples in the dataset, it is more likely to overfit the training data, which leads to high precision but low recall on test set. In order to better compare precision and recall, we plot the Precision-Recall curves of different methods, as shown in Figure 2. *RefD* shows a clear improvement in the area under the Precision-Recall curve.

Comparing two weighting methods, we find that TFIDF performs slightly better than EQUAL on CS while EQUAL has higher scores than TFIDF on MATH. Since how to compute w in *RefD* is a crucial problem, our ongoing work is to explore more sophisticated semantic representations to measure prerequisite relations. A natural extension to the two simple methods here is to represent

1671

Concept	RefD	Concept	RefD	Concept	RefD
Deep belief network	-0.38	List of Nobel laureates	0.009	Machine learning	0.32
Neocognitron	-0.28	Neural development	0.009	Artificial neural network	0.31
Word embedding	-0.24	Watson (computer)	0.003	Artificial intelligence	0.15
Vanishing gradient problem	-0.22	Self-organization	8e-5	Algorithm	0.14
Feature learning	-0.17	Language model	-0.004	Statistical classification	0.13

Table 4: *RefD* scores between "deep learning" and the concepts linked from it. All scores are calculated by *RefD*('deep learning', concept).

Please note that our Wikipedia-based implementation is computationally efficient especially after precomputing weights and references and can be easily incorporated as a feature into existing supervised learning based methods.

5 Related Work

In the area of education, researchers have tried to find prerequisites based on the assessment data of students' performance (Scheines et al., 2014; Vuong et al., 2011). However, prerequisite relations have not been well studied in computer science, with only a few exceptions. Liu et al. (2011) studied learning-dependency between knowledge units using classification where a knowledge unit is a special text fragment containing concepts. We focus on more general prerequisite relations among concepts. Talukdar and Cohen (2012) applied a Maximum Entropy classifier to predict prerequisite structures in Wikipedia using various features such as a random walk with restart score and PageRank score. Instead of doing feature engineering, we propose to measure prerequisite relations using a single metric. Yang et al. (2015) proposed Concept Graph Learning to induce relations among concepts from prerequisite relations among courses, where the learned concept prerequisite relations are implicit and thus can not be evaluated. Our method is more interpretable for measuring prerequisite relations.

Our work is closely related to the study of semantic relations. One direction is automatic lexical relation extraction. Different methods have been proposed to discover hypernym-hyponym relations based on lexical patterns (Hearst, 1992; McNamee et al., 2008; Ritter et al., 2009), distributional similarity (Kotlerman et al., 2010), semantic word embeddings (Fu et al., 2014), etc. Another line is entity relation extraction, which can be performed by distant supervision (Mintz et al., 2009; Riedel et al., 2010), Open IE (Fader et al., 2011), and neural networks (Bordes et al., 2011; Lin et al., 2015).

In addition, semantic relatedness measures have been widely studied, where the key is to model the semantic representation based on a latent space, such as LSA (Deerwester et al., 1990), PLSA (Hofmann, 1999), LDA (Blei et al., 2003) and distributed word embeddings (Huang et al., 2012; Mikolov et al., 2013), or an explicit concept space, such as ESA (Gabrilovich and Markovitch, 2007), SSA (Hassan and Mihalcea, 2011), and SaSA (Wu and Giles, 2015). Our work can also be served as a basis for building concept hierarchy (Wang et al., 2015) and teaching/learning assistant tools (Liang et al., 2015).

6 Conclusions and Future Work

We studied the problem of measuring prerequisite relations among concepts and proposed *RefD*, a general, light-weight, and effective metric, to capture the relation. We presented Wikipedia-based implementations of *RefD* with two different weighting strategies. Experiments on two datasets including seven domains showed that our proposed metric outperformed existing baselines using supervised learning.

Promising future directions would be applying the framework of *RefD* to other contexts such as measuring the prerequisite relations or reading orders between papers and textbooks. In addition, *RefD* can be incorporated into existing supervised models for a more accurate measure. Also it would be meaningful to explore ranking different prerequisites of a concept. Besides the rich link structure we could take advantage of more content information from Wikipedia and other resources such as textbooks and scientific papers.

Acknowledgments

References

John R Bergan and Patrick Jeska. 1980. An examination of prerequisite relations, positive transfer among learning tasks, and variations in instruction for a seriation hierarchy. *Contemporary Educational Psychology*, 5(3):203–215.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. In *Conference on Artificial Intelligence*, number EPFL-CONF-192344.

William Croft and D Alan Cruse. 2004. *Cognitive linguistics*. Cambridge University Press.

Scott C. Deerwester, Susan T Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. 1990. Indexing by latent semantic analysis. *JAsIs*, 41(6):391–407.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics.

Charles J Fillmore. 2006. Frame semantics. *Cognitive linguistics: Basic readings*, 34:373–400.

Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu. 2014. Learning semantic hierarchies via word embeddings. In *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics: Long Papers*, volume 1.

Evgeniy Gabrilovich and Shaul Markovitch. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *IJCAI*, volume 7, pages 1606–1611.

Samer Hassan and Rada Mihalcea. 2011. Semantic relatedness using salient semantic analysis. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011*.

Marti A Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pages 539–545. Association for Computational Linguistics.

Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57. ACM.

Eric H Huang, Richard Socher, Christopher D Manning, and Andrew Y Ng. 2012. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 873–882. Association for Computational Linguistics.

Lili Kotlerman, Ido Dagan, Idan Szpektor, and Maayan Zhitomirsky-Geffet. 2010. Directional distributional similarity for lexical inference. *Natural Language Engineering*, 16(04):359–389.

Stephen Laurence and Eric Margolis. 1999. Concepts and cognitive science. *Concepts: core readings*, pages 3–81.

Chen Liang, Shuting Wang, Zhaohui Wu, Kyle Williams, Bart Pursel, Benjamin Brautigam, Sherwyn Saul, Hannah Williams, Kyle Bowen, and C. Lee Giles. 2015. Bbookx: An automatic book creation framework. In *Proceedings of the 2015 ACM Symposium on Document Engineering*.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of AAAI*.

Jun Liu, Lu Jiang, Zhaohui Wu, Qinghua Zheng, and Yanan Qian. 2011. Mining learning-dependency between knowledge units from text. *The VLDB Journal*, 20(3):335–345.

Paul McNamee, Rion Snow, Patrick Schone, and James Mayfield. 2008. Learning named entity hyponyms for question answering. In *IJCNLP*, pages 799–804.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1003–1011. Association for Computational Linguistics.

Matthew W Ohland, Amy G Yuhasz, and Benjamin L Sill. 2004. Identifying and removing a calculus prerequisite as a bottleneck in clemson's general engineering curriculum. *Journal of Engineering Education*, 93(3):253–257.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Machine Learning and*

Knowledge Discovery in Databases, pages 148–163. Springer.

Alan Ritter, Stephen Soderland, and Oren Etzioni. 2009. What is this, anyway: Automatic hypernym discovery. In *AAAI Spring Symposium: Learning by Reading and Learning to Read*, pages 88–93.

Richard Scheines, Elizabeth Silver, and Ilya Goldin. 2014. Discovering prerequisite relationships among knowledge components. In *Proceedings of Educational Data Mining*, pages 355–356.

Partha Pratim Talukdar and William W Cohen. 2012. Crowdsourced comprehension: predicting prerequisite structure in wikipedia. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 307–315. Association for Computational Linguistics.

Annalies Vuong, Tristan Nixon, and Brendon Towle. 2011. A method for finding prerequisites within a curriculum. In *Proceedings of Educational Data Mining*, pages 211–216.

Shuting Wang, Chen Liang, Zhaohui Wu, Kyle Williams, Bart Pursel, Benjamin Brautigam, Sherwyn Saul, Hannah Williams, Kyle Bowen, and C. Lee Giles. 2015. Concept hierarchy extraction from textbooks. In *Proceedings of the 2015 ACM Symposium on Document Engineering*.

Zhaohui Wu and C. Lee Giles. 2015. Sense-aware semantic analysis: A multi-prototypeword representation model usingwikipedia. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 2188–2194.

Yiming Yang, Hanxiao Liu, Jaime G. Carbonell, and Wanli Ma. 2015. Concept graph learning from educational data. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM 2015, Shanghai, China, February 2-6, 2015*, pages 159–168.