Measuring Prerequisite Relations Among Concepts

Chen Liang[1], Zhaohui Wu[2], Wenyi Huang[1], C. Lee Giles[1]

[1]Information Sciences and Technology

[2]Computer Science and Engineering

The Pennsylvania State University

Motivation

A *prerequisite* relation describes a basic relation among concepts in cognition, education and other areas.

What should one learn before starting to learn a new area such as "deep learning"?

However, as a semantic relation, it has not been well studied in computational linguistics.

Research questions:
How do we measure prerequisite relations among concepts?
Can we automatically detect the prerequisites of a concept?

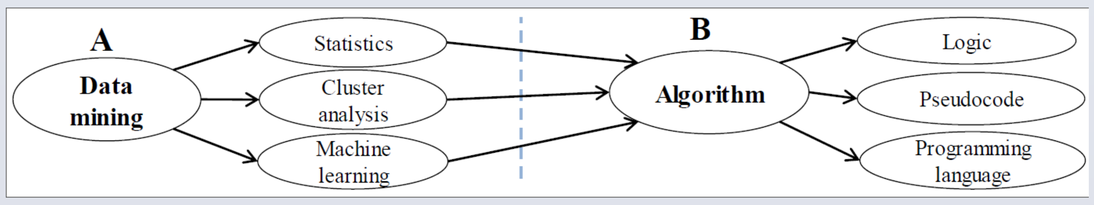
Our Contribution

- A novel link-based metric, *reference distance* (**RefD**), to measure the prerequisite relation among concepts that outperforms existing supervised learning baselines.
- A new dataset containing 1336 concept pairs in CS and Math.

Method

Goal: design a function $f: C^2 \rightarrow R$ that maps a concept pair (A, B) to a real value that measures the extent to which A requires B.

- One cannot understand a concept without access to all essential knowledge related to it. Thus a concept can be represented by its related concepts in C. (Frame semantics)
- If most related concepts of A refers to B but few related concepts of B refer to A, B is more likely to be a prerequisite of A.



We propose **RefD** to measure prerequisite relations:

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

C: concept space; $w(c_i, A)$: the importance of c_i to A; $r(c_i, A)$ is an indicator showing whether c_i refers to A.

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

Wikipedia-based RefD Implementation

- Concept space C: all Wikipedia articles
- $r(c, A)$: 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

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

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

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

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.

Experiments

Task: Given a concept pair (A, B), predict whether B is a prerequisite of A.

Datasets:

- CrowdComp dataset (Talukdar and Cohen, 2012)
- Course dataset (collected by us)

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

Baseline: Maximum Entropy (MaxEnt) classifier using graph-based features and content-based features (Talukdar and Cohen, 2012)

CrowdComp Results

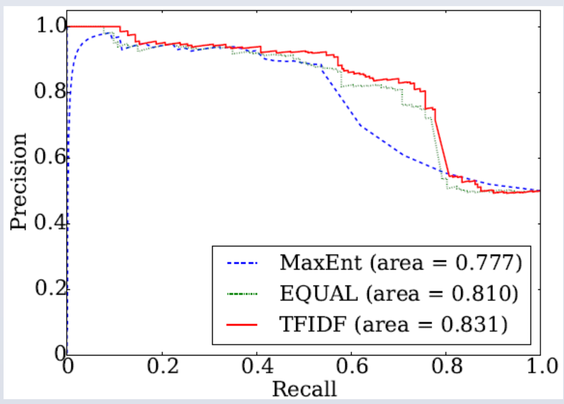
"leave one domain out"

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***

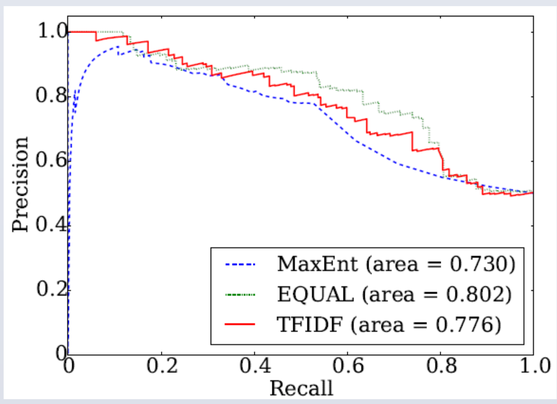
Course Results

In-domain 5-fold cross-validation

	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*

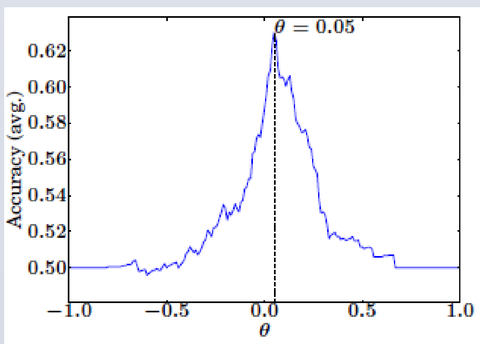


CS

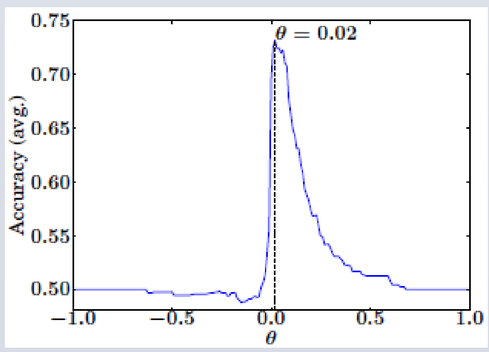


MATH

Parameter Analysis



CrowdComp



Course

Case Study

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

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

References

Talukdar, Partha Pratim, and William W. Cohen. "Crowdsourced comprehension: predicting prerequisite structure in Wikipedia." HLT-NAACL 2012 Workshop on Innovative Use of NLP for Building Educational Applications (BEA7)

Yang, Yiming, et al. "Concept Graph Learning from Educational Data." *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 2015.