Towards Creating Pedagogic Views from Encyclopedic Resources

Ditty Mathew   Dhivya Eswaran   Sutanu Chakraborti

Department of Computer Science and Engineering, Indian Institute of Technology - Madras, Chennai - 600036, India
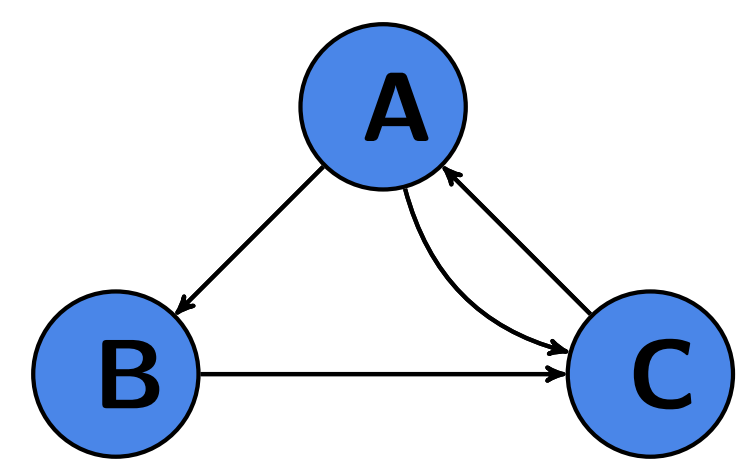
1. Introduction

Observation: Encyclopedic resources like Wikipedia have good reference value and broad coverage, but have limited pedagogic value. Textbooks on the other hand are often static and limited in coverage.

Motivation

How can we effectively create a pedagogic view of content from encyclopedic resources?

2. Concept network

- Edge from a concept A to a concept B signifies that A is used to define B
- Circularity in the concept network:
  - Description of Concept A assumes that Concept C is known
  - Description of Concept C assumes that Concept A is known
- Ideally, a pedagogic resource should ensure that its concept network is a directed acyclic graph

Goal

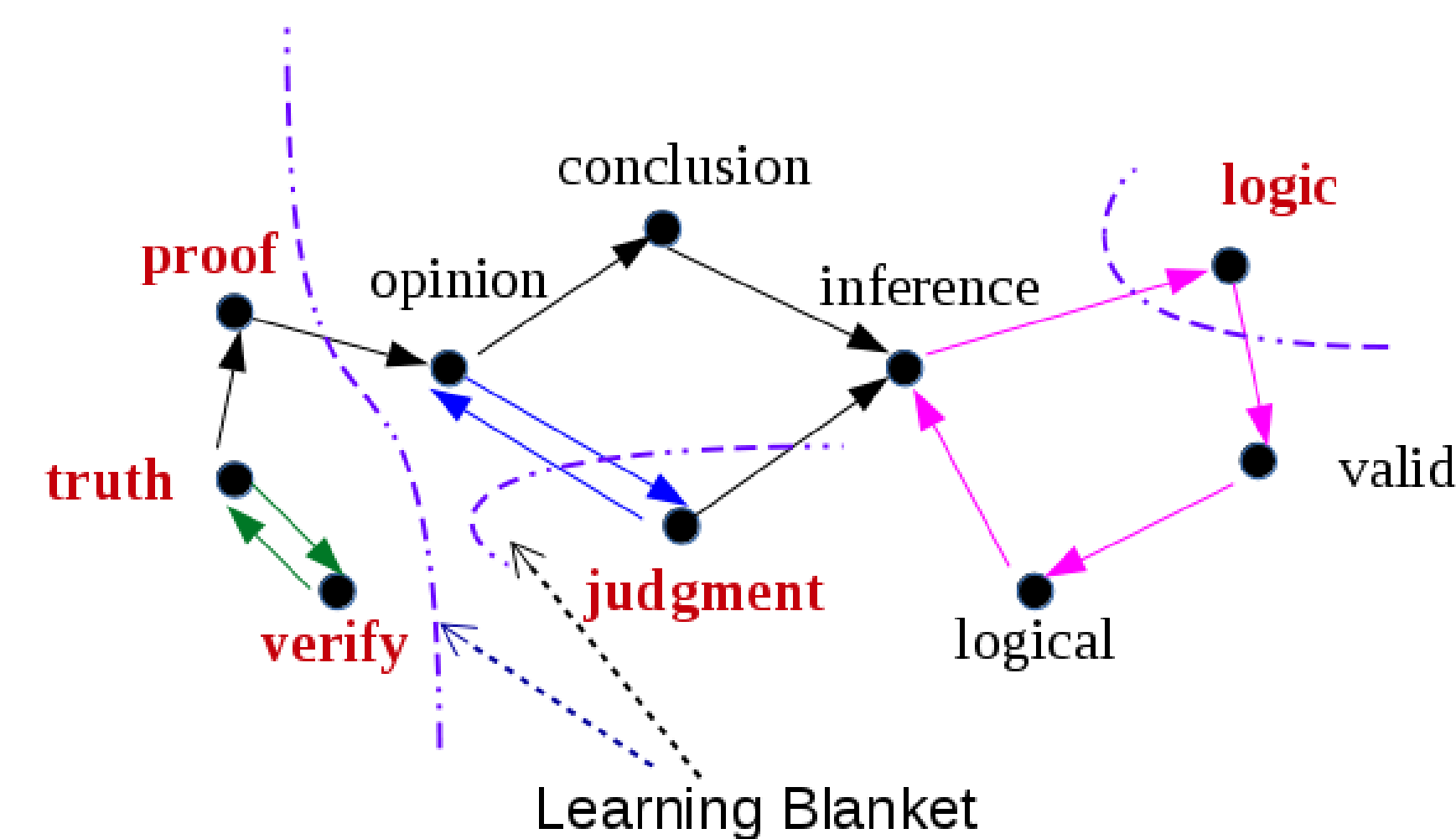Identify circular definitions and help content editors eliminate them

- WordNet is used as an example of encyclopedic resource, where concepts are words

3. Soft AND-OR composition
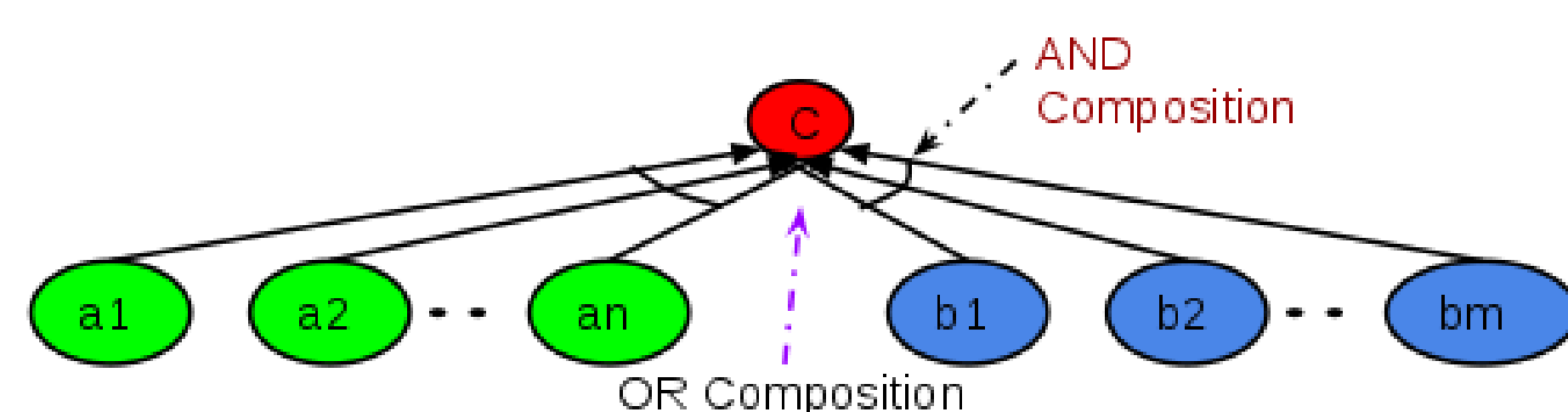
- User has to know either all a_i s or all b_j s to understand the Concept C. Soft AND-OR composition imposes relaxation on AND

4. Example

- Learning blanket encompasses the set of concepts in the concept graph that learner is familiar with

Circularities situated below the learning blanket do not challenge the learner

5. Methods to resolve circular dependencies

- Perceptual grounding
  - Eg: Concepts like *red* can be defined by contrasting against other colors
- Collapsing
  - Eg: Concepts such as *polite* and *courteous* can be defined using a single definition *showing good manners*
- Linguistic grounding
  - Eg: the circular definition of *opinion* in the example depicted in section 5 can be broken by redefining it as *a personal view* instead of the current definition *a judgment of a person.*

6. Greedy discovery of concepts for grounding

- Identify the set of concepts that do not take part in any cycle
- Rank the remaining concepts based on the extent to which they affect learning due to cycles
- Add concepts one by one in the ranked order, until there are no more cycles in the graph

7. Ranking concepts for grounding

Relative Coverage

$helpsUnderstand(a,b) \iff a$ occurs in the definition of b

$Coverage(a) = \{b \,|\, helpsUnderstand(a,b)\}$

$Reachability(b) = \{a \,|\, helpsUnderstand(a,b)\}$

$RelativeCoverage(a) = \sum_{b \in Coverage(a)} \frac{1}{|Reachability(b)|}$

Weakness: Ignores transitive closures and implicitly assumes OR composition

PageRank

- Recursively estimate the importance of concepts in the concept network

Limitation: Score of a concept increases (decreases) with increase (decrease) in the score of any of its in-neighbours

Proposed Solution : Weighted PageRank with weight as Relative Coverage

8. Identifying regions for collapsing

- **Hypothesis**: Nodes in Strongly Connected Components(SCC) are related
- Collapse SCCs in which the number of nodes is less than some threshold
- Rank SCCs using topological sort

9. Resources

- **Learning resources**: content words in Brown and Gutenberg corpora
  - Brown Corpus - 23,238 content words
  - Gutenberg Corpus - 18,361 content words
- **Concept network**: using the word definitions in WordNet
- **Average learning blanket**: words present in Indian English Textbook published by National Council of Educational Research And Training(NCERT)

10. Comparison of methods

Values denote % of concepts flagged to experts

Avg. level of learning blanket	Relative Coverage		Pagerank		Pagerank (Rel Cov)		Random	
	Brown	Gut	Brown	Gut	Brown	Gut	Brown	Gut
1	13.9	14.7	14.7	14.8	**13.6**	**13.9**	28.5	29.5
2	13.0	12.9	12.7	12.5	**11.4**	**11.3**	24.1	25.9
3	12.5	12.3	12.5	10.9	**10.6**	**10.7**	25.7	23.8
4	11.2	9.9	10.4	9.2	**9.0**	**8.8**	19.3	20.3
5	13.4	**10.8**	9.3	12.2	**8.5**	12.9	18.1	20.2

- Smaller fraction of concepts should be flagged for editing to reduce the human effort. Values in bold correspond to the best reductions.

11. Example - concepts suggested for collapsing

displeasure	magnificent	pasture	sleeve
displease	grandeur	herbage	armhole
deceit, deceive	stubborn	existence	enfold
defraud	obstinate	extant	enclose
dishonest	tenaciously	exist	

12. Conclusion and future work

- **Contribution**:
  - Identified characteristics of the encyclopedic resources that hinder learning
  - Proposed approaches that help experts create a pedagogic view of an encyclopedic resource. Experiments show significant reduction in the number of concepts flagged for editing
- **Future work**:
  - Explore personalized extensions to discover the learning blanket of a learner to help her explore the pedagogic space of concepts
  - Extend to resources like Wikipedia

References

[1] D Levary, J P Eckmann, E Moses and T Tlusty. *Loops and Self-Reference in the Construction of Dictionaries*, Physical Review X 2, 031018, 2012

[2] B Smyth and E McKenna. *Footprint-Based Retrieval*, Proceedings of the Third International Conference on Case Based Reasoning and Development, Pages 343-357, 2012

[3] Deb Roy. *Grounding words in perception and action: computational insights*, TRENDS in Cognitive Sciences Vol.9 No.8, Pages 389-396, 2005