Unambiguous Concept Mapping in Radiology Reports:Graphs of Consistent Concepts

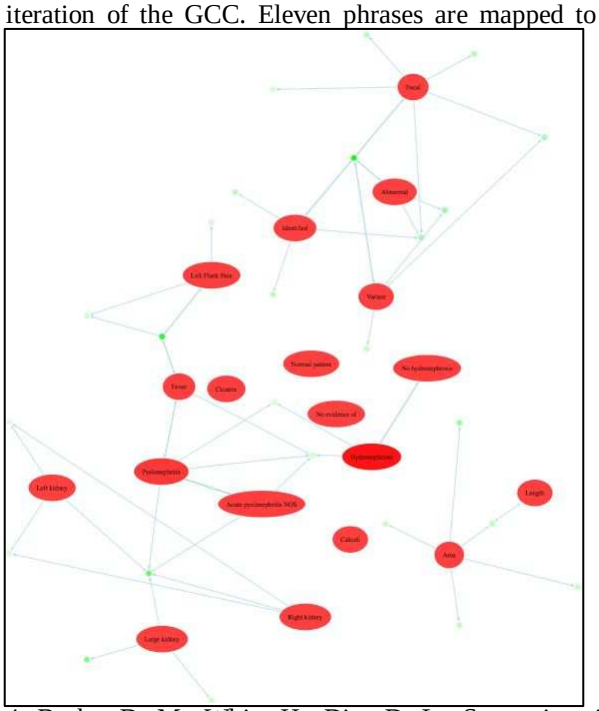
Pawel Matykiewicz[1, 2], John Pestian[1], Wlodzislaw Duch[2] Neil Johnson[1]
[1]Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
[2]Department of Informatics, Nicholaus Copernicus University,Torun, Poland

Background. A significant part of medical data remains stored as unstructured texts. The long-term goal of this project is to reach a human-level competence in medical text interpretation. One step in this process is to map words or phrases to a medical ontology. MetaMap[2] is a well-known tool that does this, but also produces ambiguous mappings. The classical word sense disambiguation methods[5, 6] use fixed windows context clustering to find the right meaning. This research develops a concurrent method of simulating *semantic priming*[3, 4] to discriminate between senses of concepts.

Methods. The Unified Medical Language System[1] (UMLS) is used to create a network of concepts and associations between them. Medical text is then mapped to the UMLS concepts. The unambiguous mappings create a backbone graph of consistent concepts (GCC). They spread activation across whole network and excite different concepts. The level of excitation of concepts from ambiguous mappings suggests the right meaning of the ambiguous concept.

Results. *Example*: "Fever, left flank pain, pyelonephritis. The right kidney is normal in sonographic appearance with no evidence of scarring, hydronephrosis or calculi. It measures XXXX cm which is normal for patient's age. The left kidney is enlarged. It measures XXXX cm in length. No focal areas of abnormal echogenicity or scarring are seen. No hydronephrosis or calculi are identified. Images of the bladder demonstrate no abnormality. Enlargement of the left kidney which may suggest acute pyelonephritis. This could also represent a normal variant. Normal appearing right kidney."

After automatically mapping these phrases from the example text to UMLS 19 concepts created the first iteration of the GCC. Eleven phrases are mapped to more than one concept. Excitation level of those concepts enables correct choice 64% time. Using supervised training it is possible to add new associations between concepts. This provides specific knowledge that increases the domain specific mapping accuracy. The Figure shows concepts, phrases and mappings. The red circles are the concepts; the green circles are connecting phrases. The intensity of the color depicts the intensity of the relationship.



Future work. Thorough radiology knowledge must be added to the network to increase mapping performance. For that purpose a training/learning set must be manually constructed. Different spreading activation and measures of consistency must be investigated.

REFERENCES

1. Unified Medical Language System (UMLS) web site: http://www.nlm.nih.gov/research/umls/

2. MetaMap web site: http://mmtx.nlm.nih.gov

3. Chertkow H., Bub D., Seidenberg M.: Priming and semantic memory loss in Alzheimer's disease. Brain and Language, 36 (1989), 420–446

4. Burke, D. M., White H., Diaz D. L.: Semantic priming in young and older adults: evidence for age constancy in automatic and attentional processes. Journal of Experimental Psychology: Human Perception Performance, 13 (1987), 79–88

5. Schutze, H.: Automatic Word Sense discrimination. Computational Linguistics, 24 (1998), 97–124

6. Pedersen, T., Bruce, R.: Distinguishing word senses in untagged text. Proceedings of the Second Conference on Empirical Methods in Natural Language Processing, 2 (1997), 197–207