

# Fall 2005 and ahead..

Maresh Joshi  
joshi031@d.umn.edu  
November 18th 2005

## Status quo Fall 2005 and future work

- Work done so far
  - Continuation of work done during summer internship
  - Literature review on SVM Kernels
- To be accomplished
  - Development of NLP kernel for disambiguation in medical text
  - Integration of WSD Shell and WSD Gate

## Summer Work

- Supervised machine learning methods for disambiguating acronyms in medical text, e.g. APC is *Atrial Premature Contraction* or *Argon Plasma Coagulation*
- Developed WSD framework based on GATE, NSP and WEKA
- Presentation available at [http://www.d.umn.edu/~joshi031/files/mayo\\_seminar\\_summer2005\\_mahesh.pdf](http://www.d.umn.edu/~joshi031/files/mayo_seminar_summer2005_mahesh.pdf)

11/18/05

3

## Fall 2005

- Continued to develop the summer WSD framework
  - Extended it to enable dealing with any XML formatted data, specially Senseval-2 format
  - Modifications to GATE code submitted for approval, partly incorporated and others in process of review + incorporation
  - Will be released in the form of 2 packages on sourceforge
- Literature review on developing NLP kernels
  - Book: Kernel Methods for Pattern Recognition (1)
  - Papers: Domain Kernels (2), Latent Semantic Kernels (3)

11/18/05

4

## Fall 2005 contd..

- Tested the WSD framework on abbreviation data from Dr. Hongfang Liu at UMBC (4)
- Data used so far
  - National Library of Medicine dataset
  - Mayo Clinic clinical notes data
  - Abbreviation data from Dr. Hongfang Liu (4)

11/18/05

5

## Software and Scripts

- Will be releasing soon
  - NSPGATE: A wrapper component for NSP, which provides NSP functionality in GATE
    - <http://sourceforge.net/projects/nspgate>
  - WSDGATE: WSD framework that integrates GATE, NSPGATE and WEKA
    - <http://sourceforge.net/projects/wsdgate>
- Scripts
  - Will release conversion script to convert Liu abbreviation data into Senseval-2 format

11/18/05

6

## Future Work

- Analyze data to look for features useful for developing NLP kernel
- Developing an NLP kernel for the WSD task
- More experiments and more results!
- Enhancing WSDGATE with configurability features and experimental setup enhancements

11/18/05

7

## Future Work contd..

- Integrating the WSD GATE framework with WSD Shell
  - Feature annotation is done using WSDGATE, and extraction can be done in either WSDGATE or WSDShell

11/18/05

8

## References

1. J. Shawe-Taylor and N. Cristianini. Kernel Methods for Pattern Analysis. Cambridge University Press, 2004.
2. A. Gliozzo, C. Giuliano, and C. Strapparava. Domain kernels for word sense disambiguation. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), pages 403-410, University of Michigan, 2005.
3. N. Cristianini, J. Shawe-Taylor, and H. Lodhi: 2001, 'Latent Semantic Kernels'. In: C. Brodley and A. Danyluk (eds.): Proceedings of ICML-01, 18th International Conference on Machine Learning. Williams College, US, pp. 66--73.
4. H. Liu, V. Teller and C. Friedman Multi-Variable Comparison Study of Supervised Word Sense Disambiguation JAMIA 2004