CS565: Intelligent Systems and Interfaces

Project Guideline

27th Jan, 2016

Semester: Jan – May 2016

Ashish Anand

IIT Guwahati

Objective

- Hands on experience working with current challenging problems in NLP domain

- Chose flavor to explore
  - Development and application
  - Research
  - Contribute to open source projects

Project Administration

Project Group

- Group of 4 Students [Current status: 90 Students]
  - Max of 25 Groups
- Group Information should be put on Canvas by **3rd Feb, Wednesday**
  - Saptarshi will update how and where to put that information.
- Assign Roles to each member

Important Deadlines

- Group Information: 3rd Feb, Wednesday
- Topic and brief description of problem: 15th Feb, Monday
- Mid-Term Report: 15th Mar, Tuesday
- Final Report and Code: 15th Apr, Friday [Tentative]
- Presentation & Viva: 16th Apr onwards [Tentative]
- No extension feasible

Marks Distribution

- Total Marks: 45; [Assignment: 25]
  - Group Work + Individual Contribution

Topic and Brief Description of problem

- Problem Statement
- Challenges
- Brief Review [max ½ page] of existing models
- Proposed Direction [If you have thought already]
- Relevant References
- Total not to exceed 1-2 Page

Mid-Term Report

- Objective: More formalized version of initial problem description submission
- Presented in form of extended abstract
  - Saptarshi will upload latex template
- Organize into following sections
  - Abstract, Introduction, Method, Progress, Conclusion, References
- 2-3 Pages

Final Report

- Follow a full paper format
  - Template – Saptarshi will share
  - 6 Pages

- Focus would be on clear description of problem, challenges, methods selected, Results and analysis

What I'll be expecting

- Report
  - Explain problem: definition or formulation, motivation, challenges, existing methods: adv. and disadv.
  - Explain your data: basic statistics, pre-processing
  - Explain your method: new proposal or comparative study, adv. of your method or project, novelty aspect
  - Explain Implementation: implemented yourself or used off-the-shelf libraries or tools.
  - Explain Result: Provide insights from obtained results.
  - Explain Future Scope: what next and what could be done differently
- Code
  - Well commented and readme file
  - Anybody should be able to execute and re-obtain your results as reported in your report.

What kind of work will be more appreciated

- Genuine new contribution
  - Could be in terms of tool development
  - Could be in terms of coming up with novel solution to existing problems
  - Coming with new problem formulation and solution
- Comprehensive studies on existing works
- Creating new benchmark corpus and its basic analysis

Project Topic

Themes

- Embedding: character, word, paragraph, document, knowledge-graph/base
- Unsupervised Relation Extraction
- Event Detection
- Co-reference Resolution
- Graphical Summarization
  - Document [Important relations between entities within document]
  - Search Queries
- Knowledge base generation

Challenges

- Opinion Mining – SemEval 2014, Task 4
- Relation Extraction – SemEval 2010, Task 8
- Sentiment Analysis – SemEval 2015-16 [Data available]
- Textual Similarity – SemEval
- NER in Tweets – ACL shared task 2015
- Semantic Role Labeling – CoNLL challenge
- i2b2 challenge
- Biomedical entity recognition
- Document classification: Multi-Label classification
- BioAsq Challenge
- BioNLP-ST 2016

Relevant Conferences

- NAACL-HLT
- ACL
- EMNLP
- IJCNLP
- CoNLL
- SIGIR
- ICLR