CS565: INTELLIGENT SYSTEMS AND INTERFACES

Project Guideline

Semester: July – November 2020

Ashish Anand

Associate Professor, Dept of CSE

IIT Guwahati

Objective

- Hands on experience working with current challenging problems
- Chose flavor to explore
  - Research
  - Development and application
  - Contribute to open source projects

PROJECT ADMINISTRATION

Project Group

- 4 is maximum limit
- Assign Roles to each member
  - Marks will be based on Individual Contribution

Important Deadlines

- Project Group Information: 17th Sep, Thursday
- Topic and brief description of problem: 30th Sep, Wednesday
- Mid-Term Report: 25th Oct, Sunday
- Final Report and Code: 12th Nov, Thursday
- Presentation & Viva: 15th Nov onwards
- No extension feasible

Marks Distribution

- Project: 40
  - Group Work + Individual Contribution
  - Divided across several evaluation stages
  - Report across evaluation stages
  - Plagiarism Check

Topic and Brief Description of problem

- Problem Statement
- Major Challenges
- Brief Review [max 1 page] of existing models
- Proposed Direction [If you have thought already]
- Relevant References
- Total not to exceed 2-3 Page; 3 including references is max limit;

- Evaluation: Primarily based on how comprehensive and good a report is. Report must be in your own words with proper citation.

Mid-Term Report

- Objective: More formalized version of initial problem description submission
- Presented in form of extended abstract
  - We will provide latex template
- Organize into following sections
  - Abstract, Introduction, Method, Progress, Conclusion, References
- 3-4 Pages, Max 4 Pages

- Evaluation: Based on Progress made, clarity in the work direction, report and interaction

Final Report and Presentation

- Follow a full paper format
  - Latex template will be shared again
  - 6-7 Pages

- Evaluation: Based on clarity on the problem formulation, model selection, challenges handled, analysis of reported results, report and presentation

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 replicate results as reported in 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 THEMES

Themes: Pick One of Your Interests

- Word-Embedding [WE]
  - Multilingual
  - Subword structure
  - Choose one of the Indian Languages
- Fine-grained Named-entity recognition (NER) and Relation Extraction [FGNER-RE]
  - Automatic Data Generation: Heuristics/Model-based
  - Automatic Data Generation: Low-Resource Language/Multilingual
  - Noise Aware Models
  - Comparative studies on different types of loss functions: hierarchical, partial loss on one of the tasks
  - Analysis study on in-house generated datasets [Abhishek/Akshay]

More Themes to pick from

- Text Coherence Analysis
- Grammatical Correction
- Q & A Systems [QA]
  - Answer is available as a span of text in the given paragraph [SQUAD]
  - Answer requires reasoning over multiple paragraphs [HotpotQA]
  - Answers of complex questions requiring understanding over multiple paragraphs [natural question dataset]
  - BioASQ
- Conversational Agent [ConvAI]
  - RLLChatbot: Method as well as open-course resources
- Explainability/Interpretability of Neural/DL Models on NLP Tasks [ExpAI]

Relevant Resources

- Recent Conferences and workshops
- Challenges [Adv: Data availability]

Relevant Conferences

- NAACL-HLT, ACL, EACL, EMNLP, CoNLL
- IJCNLP, CoNLL, CoLING, AKBC
- SIGIR, ICLR, AAAI, IJCAI
- ICML, WWW, KDD, NeuRIPS/NIPS
- ArXiv: Computational Linguistics