

An Indian-Australian research partnership

Project Title: **Active learning on graphical models (applied to Information Extraction)**

Project Number IMURA0408

Monash Main Supervisor
(Name, Email Id, Phone) Prof. Reza Haffari
Gholamreza.Haffari@monash.edu *Full name, Email*

Monash Co-supervisor(s)
(Name, Email Id, Phone)

Monash Head of Dept.
(Name,Email) Prof. Graham Farr *Full name, email*

Monash Department: Clayton School of IT

Monash ADRT
(Name,Email) Prof. Kai Ming Ting *Full name, email*

IITB Main Supervisor
(Name, Email Id, Phone) Prof. Ganesh Ramakrishnan
ganesh@cse.iitb.ac.in *Full name, Email*

IITB Co-supervisor(s)
(Name, Email Id, Phone)

IITB Head of Dept
(Name, Email, Phone) Prof.S.Sudarshan *Full name, email*

IITB Department: Computer Science and Engineering

Research Academy Themes:

Highlight which of the Academy's Theme(s) this project will address?

(Feel free to nominate more than one. For more information, see www.iitbmonash.org)

1. **Advanced computational engineering, simulation and manufacture**
2. Infrastructure Engineering
3. Clean Energy
4. Water
5. Nanotechnology
6. Biotechnology and Stem Cell Research

The research problem

Define the problem

Classification systems enabled with active learning capability iteratively select a sample of data to label based on some selection strategies suggesting the instances that deserve the most to be labeled. The goal, often, is to achieve comparable performance with supervised learners while using much less labeled data. Many algorithms have been proposed [see references below] which adopt uncertainty based principles for active learning. However, only the last one deals with active learning in the setting of structured output prediction (learning in non-iid settings in general). This area needs to be explored further

and that is what we aim to do in this project.

References:

S. Godbole, A. Harpale, S. Sarawagi, and S. Chakrabarti. Hiclass: Hyper interactive text classification by interactive supervision of document and term labels.

S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2:45–66, 2001.

P. Melville and V. Sindhvani. Active dual supervision: reducing the cost of annotating examples and features. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing, HLT'09*, pages 49–57, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

Project aims

Define the aims of the project

Expected outcomes

Highlight the expected outcomes of the project

How will the project address the Goals of the above Themes?

Describe how the project will address the goals of one or more of the 6 Themes listed above.

Capabilities and Degrees Required

List the ideal set of capabilities that a student should have for this project. Feel free to be as specific or as general as you like. These capabilities will be input into the online application form and students who opt for this project will be required to show that they can demonstrate these capabilities.

Potential Collaborators

Please visit the IITB website www.iitb.ac.in OR Monash Website www.monash.edu to highlight some potential collaborators that would be best suited for the area of research you are intending to float.