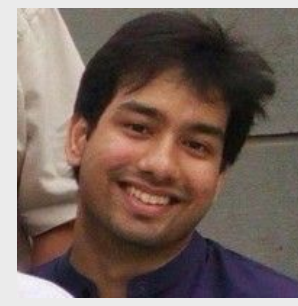
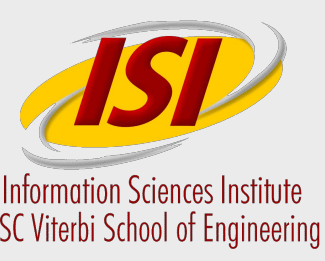


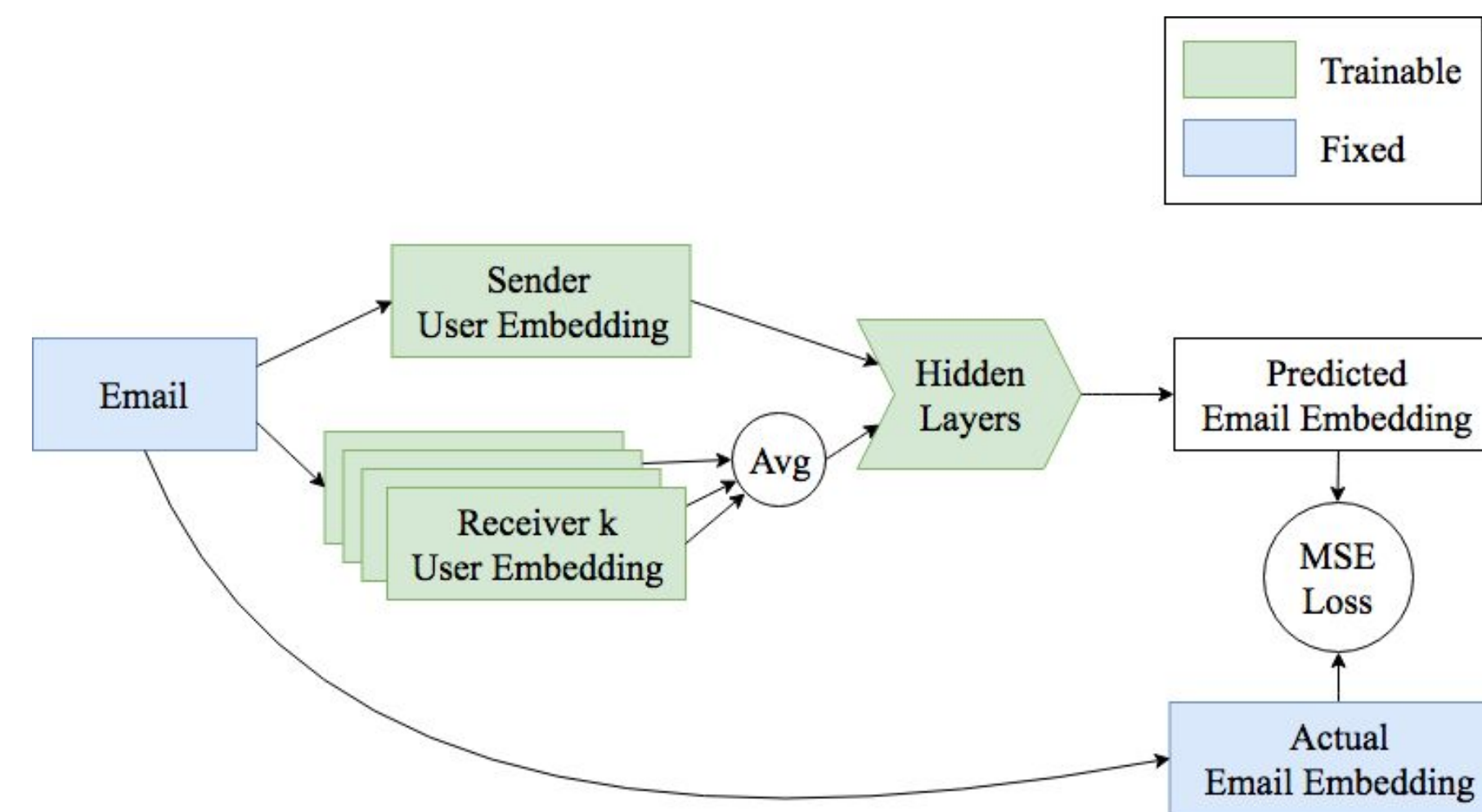
USC-ISI: Unsupervised user modeling to detect compromised email accounts



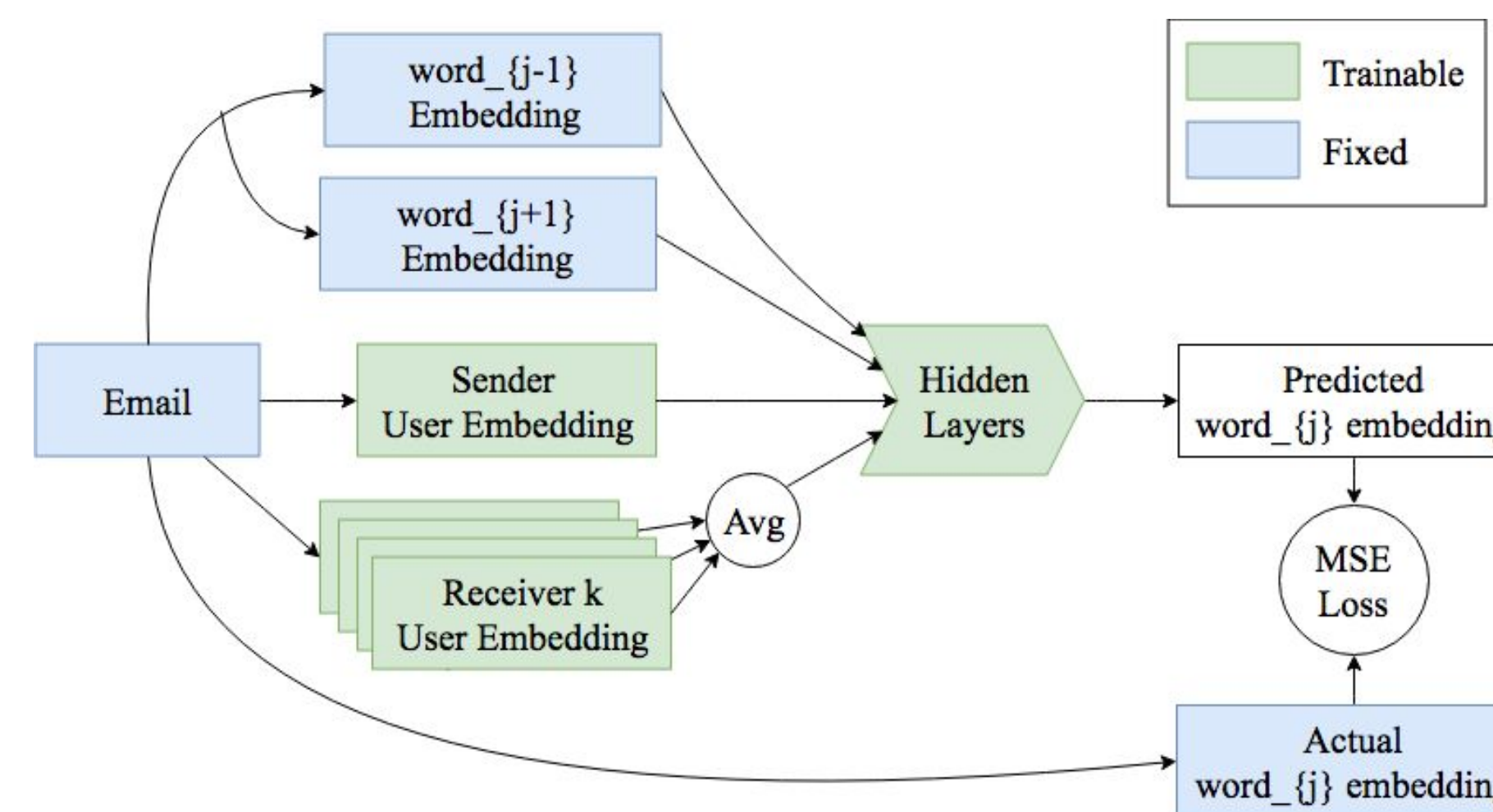
Mohit Surana, Mudit Bhargava, Rohini Kapoor, Saranya Krishnakumar
Mentors: Marjorie Freedman, Ryan Gabbard, John Lalor (UMass)



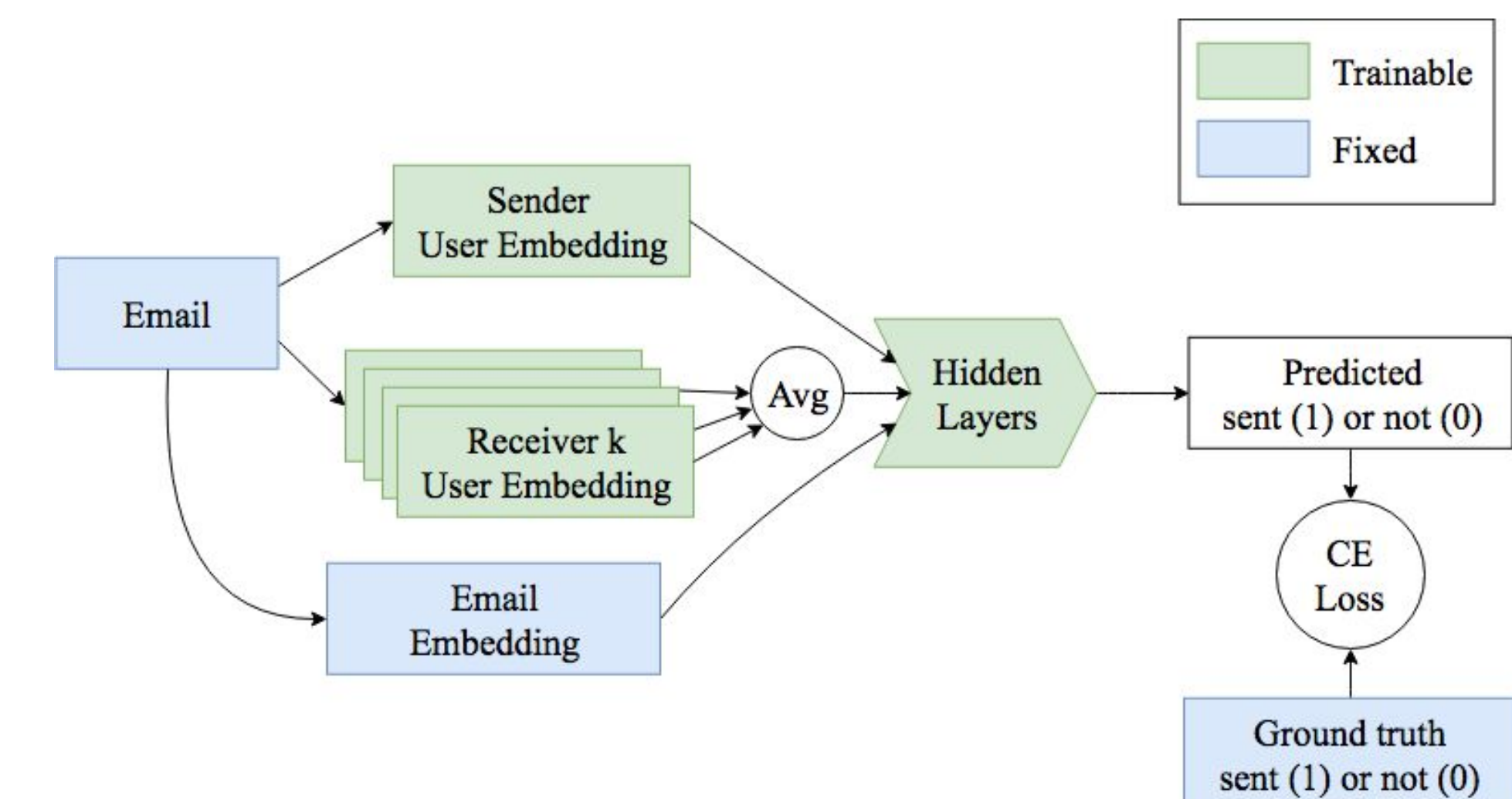
SR Model



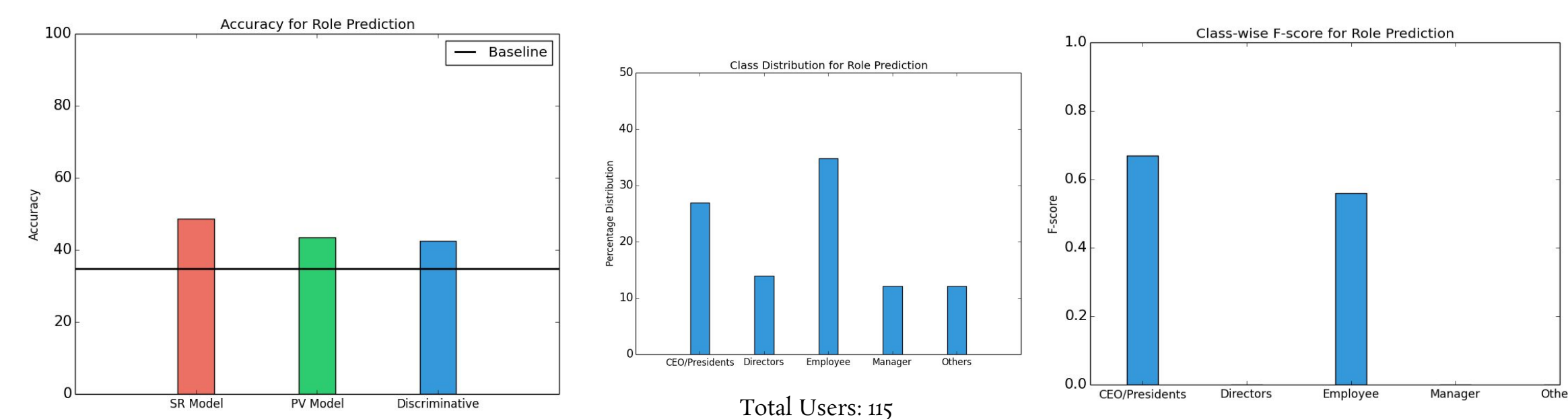
PV Model



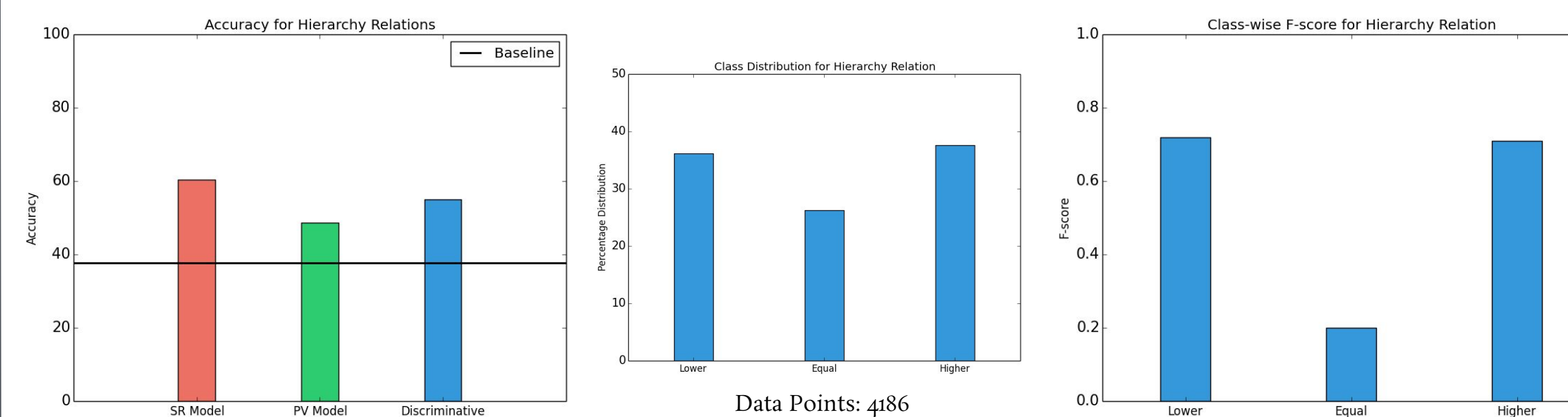
Discriminative



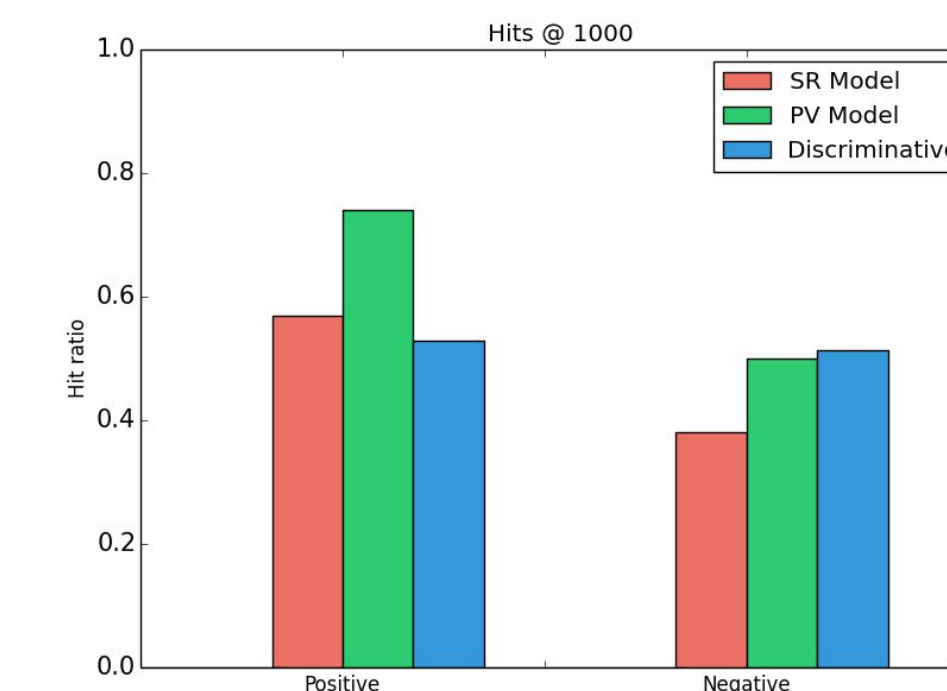
Role Prediction



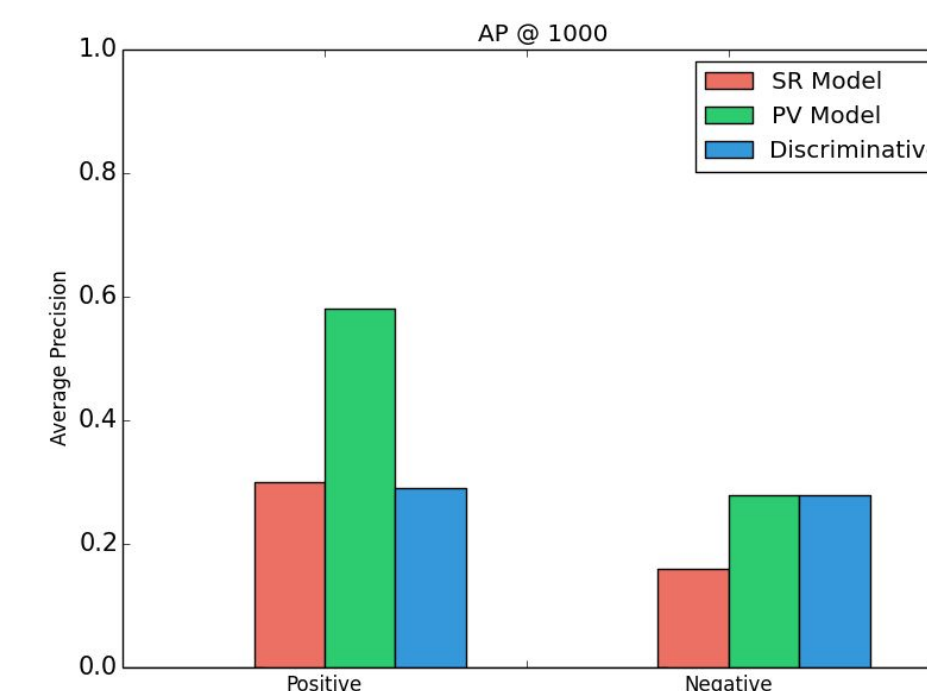
Hierarchical Relations



Hits@k



AP@k

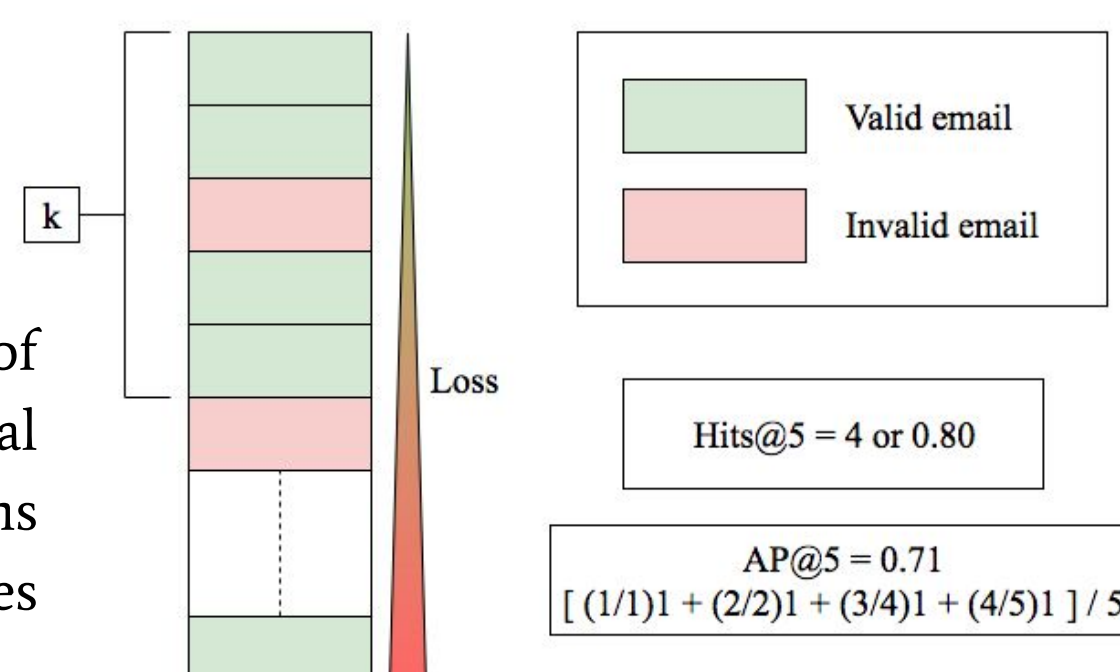


Dataset



The Enron dataset, collected as a part of 2001 Enron scandal, is the largest real world email dataset which contains roughly half a million email exchanges between Enron employees.

Metrics Description



Future Work

- Other ways of generating email representations
 - Averaging word representations leads to loss of information
- Improving the artificial negative example generation
 - How can we pick the best negative examples
- Robust models to tackle the cold start problem
 - Incorporating priors through a generic user representation

Conclusion

- Generative models that look at the words used are able to identify senders and receivers
- Email writing and receiving styles may not be indicative of the hierarchical position one holds