Social Network Inspired Models of NLP and Language Evolution

Monojit Choudhury
Microsoft Research Lab India
196/36 2nd Main, Sadashivnagar, Bangalore, India 560080
monojitc@microsoft.com

Animesh Mukherjee and **Niloy Ganguly**
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur, India 721302
{animesh,niloy}@cse.iitkgp.ernet.in

Abstract

Human language with all its intricacies is perhaps one of the finest examples of a complex system. Therefore, it becomes absolutely necessary to study the faculty of language from the perspective of a complex system. Of late, there has been an upsurge in the use of networks in modeling the complex dynamics of various natural and artificial systems. While some of these works aim at using social network techniques to build certain end-user applications, others are more fundamental in the sense that they employ these techniques to explain the emergent properties of a complex system as a whole. A substantial amount of research have also been done in the field of linguistics to employ social networks in the design of efficient solutions for numerous problems in NLP and language evolution. The objective of this tutorial is to show how language and its dynamics can be successfully studied in the framework of social networks. The tutorial will particularly demonstrate the relevance of social network-based methods in the development of a large variety of NLP applications and in understanding the dynamics of language evolution and change.

The tutorial is divided into two parts. Part I begins with a brief introduction to this field showing how linguistic entities and the interactions between them can be respectively represented through the nodes and edges of a network. This will be followed by a comprehensive survey of the general theory of social networks with a special emphasis on the methods of analysis and models of synthesis for such networks.

Part II presents three case studies. The first case study is on unsupervised POS tagging, the second one involves modeling of the mental lexicon and applications of such models in spell checking and word sense disambiguation. The third case study demonstrates the usefulness of social networks in explaining some of the evolutionary dynamics of language pertaining to the sound inventories.

The tutorial is concluded by (a) comparing the above methods with more traditional methods of doing NLP, (b) providing pointers as to where to look for/publish materials in this area, and, (c) indicating some of the future research directions.

Biography

Monojit Choudhury is a post doctoral researcher at Microsoft Research, India. He has submitted his PhD from the Department of Computer Science and Engineering, IIT Kharagpur and earlier, received his B.Tech from the same department. Mr. Choudhury received the *Young Scientist Award* from the Indian Science Congress Association in 2003.

Animesh Mukherjee is a PhD student in the Department of Computer Science and Engineering, IIT Kharagpur and also a Microsoft Research fellow. He received his MTech from the same department, and BTech from Haldia Institute of Technology. Mr. Mukherjee received the *Young Scientist Award* from the Indian Science Congress Association in 2006.

Niloy Ganguly is an assistant professor in the Department of Computer Science and Engineering, IIT Kharagpur. He has received his PhD in Computer Science from Bengal Engineering and Science University, Calcutta and his Bachelors from IIT Kharagpur. He was a post doctoral fellow in Technical University of Dresden, Germany. He has numerous publications in international journals and conferences including ACL, PODC, SIGCOMM, ACM and IEEE Trans.