Data Science Seminar Series

Word and Graph Embeddings for Machine Learning

Steven Skiena, Ph.D.

Distinguished Professor
Department of Computer Science
Stony Brook University

Date: Wednesday, January 27th, 2021
Time: 4:00 PM – 5:00 PM EDT
Location: Zoom Virtual Room
Web Link: https://njit-institute-for-data-science.eventbrite.com

Distributed word embeddings (word2vec) provides a powerful way to reduce large text corpora to concise features readily applicable to a variety of problems in NLP and data science.   I will introduce word embeddings, and review several of our recent efforts to apply them for natural language processing (NLP) including the Polyglot system for entity recognition, POS tagging, and sentiment analysis) for over 100 different languages. DeepWalk is an approach we have developed to construct vertex embeddings: vector representations of vertices which be applied to a very general class of problems in data mining and information retrieval. DeepWalk exploits an appealing analogy between sentences as sequences of words and random walks as sequences of vertices to transfer deep learning (unsupervised feature learning) techniques from natural language processing to network analysis. DeepWalk has become extremely popular, having been cited by over 4000 research papers since its publication at KDD 2014.   In this talk, I will introduce the notion of graph embeddings, explain how DeepWalk constructs them, and demonstrate why they make such powerful features for machine learning applications.

Steven Skiena is Distinguished Teaching Professor of Computer Science and Director of the Institute for AI-Driven Discovery and Innovation at Stony Brook University. His research interests include data science, bioinformatics, and algorithms. He is the author of six books, including "The Algorithm Design Manual",   "The Data Science Design Manual", and "Who's Bigger: Where Historical Figures Really Rank". Skiena received his B.S. in Computer Science from the University of Virginia and his Ph.D. in Computer Science from the University of Illinois in 1988.   He is the author of over 150 technical papers. He is a Fellow of the American Association for the Advancement of Science (AAAS), a former Fulbright scholar, and recipient of the IEEE Computer Science and Engineer Teaching Award.