Networking Seminars Series
A.Y. 2023-2024 2nd Term

May, 13th 2024, 1:00-2:00 pm (CET)
Room C302, Via Pasubio 7/B, Dalmine (BG)
Online on Microsoft Teams ([Link](#))

Presenter:
Manoochehr Ghiassi

Professor of Information Systems
@Santa Clara University
Department of Information Systems & Analytics

TITLE – *"Sentiment Analysis with a Domain Transferable Feature Engineering Approach Using Supervised Machine Learning Tools"*

ABSTRACT – Social media services are a significant platform for customers and news consumers to express sentiments. Accurately capturing these sentiments has been challenging for researchers. The traditional approaches include dictionary-based and use supervised machine learning tools for sentiment classification. This research follows the supervised machine learning approach. A major challenge for the machine learning approach is feature selection, which is often domain dependent. We address this specific challenge and present a novel approach to identify a lexicon set unique to Twitter (X) Sentiment Analysis (TSA). We show that this Specific Lexicon Set (TSLS) is small, and most importantly, is domain transferable. This identification process generates a collection of vectorized tweets for input to machine learning tools. In traditional approaches, this vectorization often results in a highly sparse input matrix which produces low accuracy measures. In this research, we hierarchically reduce the feature set to a small set of seven "meta features" to reduce sparsity. We show that TSA based on these features can produce highly accurate results using a dynamic architecture for neural networks (DAN2) and SVM (machine learning tools) as measured by recall, precision, and F1 metrics (the harmonic average of precision and recall). Our results show that a Twitter Generic Feature Set (TGFS) derived from two datasets (@JustinBieber and @Starbucks) is domain transferable and when combined with only a few Twitter Domain Specific Features (TDSF) (less than 3%), can produce excellent sentiment classification values. We evaluate the effectiveness and transferability of the TGFS across three new and distinct domains (@GovChristie, @SouthwestAir, and @VerizonWireless). This presentation also offers an overall architecture for the DAN2 machine learning tools and show its effectiveness across a large set of applications.

METHOD – DAN2 machine learning tools

JOURNEY OF THE PAPER – First draft

For further information please refer to: seminars.digip@unibg.it