Title: Actuarial Applications of Word Embedding Models

Presenter: Gee Lee, Michigan State University

Abstract:

In insurance analytics, useful information is often discarded, because traditional empirical analysis requires numeric descriptor variables. Although textual data contains a lot of information, how to utilize this gold mine of information has been in veil to actuaries because of technical difficulties. This paper will demonstrate how insurance analytics can be improved by allowing textual data to be easily incorporated into a standard regression analysis. Using the concept of word similarities, we illustrate how to extract variables from text and incorporate them into claims analysis using a standard regression model. This procedure is applied to the Wisconsin Local Government Property Insurance Fund (LGPIF) data, in order to demonstrate how insurance claims management and risk mitigation procedures can be improved. Word embedding matrices are used to transform insurance claim descriptions into a collection of vectors, and the resulting explanatory variables are used for the analyses. We illustrate two applications. First, we show how the claims classification problem can be solved using textual information. Second, we analyze the relationship between risk metrics and the probability of large losses. In the two applications, we use the generalized additive model framework as an underlying theme, treating the generalized linear model framework as a special case.