Linguistic summaries of time series using fuzzy sets and their application for performance analysis of mutual funds

Anna Wilbik
Postdoctoral Fellow
Department of Electrical and Computer Engineering
University of Missouri

Date and Time: March the 8th 2011 from noon to 1pm
Place: ECE conference room (EBW room 350)

ABSTRACT

We consider some decision making issues related to financial investments, to be more specific investments in mutual funds. In our context, we assume that what is important to the decision maker is, first, the past performance of the mutual fund, and, second, how it has performed against its benchmark(s). We propose the use of linguistic data summaries in case of time series data using the seminal idea of the approach by Yager, Kacprzyk, Kacprzyk and Yager, and Kacprzyk, Yager and Zadrozny, Kacprzyk and Zadrozny, and extending it to a dynamic context of time series data. The summaries of time series we propose refer in fact to the summaries of trends (segments) identified with straight line segments of a piecewise linear approximation of time series. Basically, the linguistic summaries proposed are interpreted in terms of the number or proportion of elements possessing a certain property. Such summaries exemplified by "among all segments, most are short" or in a more sophisticated form by "among all long segments, most are slowly increasing" can be easily interpreted using Zadeh's calculus of linguistically quantified propositions. The most important element of this interpretation is a linguistic quantifier exemplified here by "most" which is interpreted in terms of a proportion of elements possessing a certain property (e.g., the length of a segments) among all the elements considered (e.g., all segments). An important aspect is how the behavior of quotations of the investment fund has followed the behavior of the benchmark assumed, over some specified past time period. That is, equivalently, how similar they have been. We present our novel approach for the comparison of times series via linguistic summaries. This approach is based on the assumption that if the simultaneously occurring segments can be described by the same characteristics, then the time series composed of those segments are similar. By extending this idea we also propose a method for the evaluation of similarity of two time series based on linguistic summaries which characterize those time series in the sense considered.

Biosketch

Anna Wilbik received M.S. degree in Computer Science from Warsaw University of Technology in 2004. Since then she worked at the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, from where she received the Ph.D. Degree in computer science in 2010. She taught students at Warsaw School of Information Technology – the courses on data mining and text mining. She translated 3 books from those topics for Polish Scientific Publishers PWN, for the last translation she received the Jerzy Kurylowicz Award for the best translation of technical literature in year 2009.

For more information contact Dr. Derek Anderson at AndersonDT@missouri.edu