STATEMENT OF RESEARCH INTEREST

Candidate: Alessandro D'Alessandro

Topic *(selected from: http://www.m3es.unimore.it/site/home/research.html)**:***

Big Data Analytics for Industry 4.0

Motivation

I am a chemist. I've been working in a food company since 1994. I've always been working in Research & Development. Despite my age I am still curious, and I love to keep learning. With more fatigue than when I was 20th, to be honest.

Some years ago, I discovered the power of Chemometrics in interpreting complex results. During this time in my job, I have had the need to elaborate data coming from different analytical sources. Studying them with the classical approach, i.e. by looking at few specific items at time (variables, constituents, etc.) was often complex and poorly effective. The multivariate approach on other hand, offers tools to extract the proper information and synthesize results to present them effectively.

A further step, increasing my interest in this discipline, was the need to relate what I was analysing in laboratory with the characteristics of the raw materials, as well as with the production process parameters and/or quality indicators of the final products. And again, the multivariate approach gives more straightforward and robust results.

With the increase of the analytical performances, the spread of automation in devices and instrumentation, the data quantity has grown correspondently, and exponentially, so that new ways to "observe" data are required in order to find the useful information in this "ocean of figures".

Similarly, the tools to monitor the production processes have reached levels unimaginable just ten years ago.

Altogether, this new scenario indeed requires new approaches, and my "self-taught" knowledge of Chemometrics is now not enough.

Project to undertake during PhD

Objective

The proposed project fit the new paradigm of Industry 4.0, which is desirable perspective for both the Company and the R&D team in general. In particular, I would like to focus during my PhD studies on the objective of finding effective ways to handle the huge quantity of data, coming from different

sensors in a productive process, and to relate them with a correspondent large quantity of analytical data produced by off-line and/or on-line instruments.
For example, spectroscopic or image analysis techniques produce, especially when used on line, an enormous amount of data, each signal or image is constituted of hundreds to thousands of variables at a sampling rate in the order of minutes, or seconds. This data needs to be processed and correlated, in a time which the closer is to the production time, the better.
In this context, I envisage the main issues of Big Data: volume, velocity, diversity.

Possible methods

The multivariate approach could furnish proper tools to approach the theme. Starting from the planning of data collection, data pre-processing, through the elaboration, to conclude with an effective graphical representation and effective way to share the results outside the R&D environment. In this respect, I would like to learn and exploit in my context more recent areas such as Data Fusion, Multivariate Image Analysis, Multivariate latent variables model to Big Data in process monitoring.

There are several different steps to consider in an industrial process in which we need to understand how the characteristic of a raw material impacts on the final product quality. Especially because food products suffer season variability. The first step could be to select the proper raw material characteristic to be measured, i.e. which likely contains the relevant information we need. In other words, whether we are measuring the right characteristic or not. For example, could the colour in a fresh vegetable influence the colour of the final sauce? Probably yes, but we have also to consider the strong impact of the transformation process.

A second step could be related to the possibility of monitoring intermediate semi-finish products during the process (i.e. with a spectroscopic technique at-line or eventually on-line) and use the information obtained to optimally setup the process parameters. In this case, the development of a predictive mathematical model could allow to anticipate and manage unwanted situations.

Expected outcomes

At the end of the research project the idea is to have a proper “toolbox” to find the best way to approach big data analysis in a real–life context like a laboratory or a production plant.

Results could be more than one. Just to mention some possibilities:

- a predictive model that monitors the process parameters and predicts the final product characteristics;
- a correlation model between raw material parameters and final product quality;
- a multivariate control chart that monitors crucial process parameters;

In conclusion, more than solving a single case, my expected outcome would be to build my “tool box” I may use in future in many other different situations.