Professor Enza Messina

Professor Enza Messina is a Professor in Operations Research at the University of Milano-Bicocca (Department of Informatics Systems and Communications), where she founded the research Laboratory MIND (www.mind.disco.unimib.it). She holds a PhD in Computational Mathematics and Operations Research from the University of Milano. She and her research team have over the years collaborated and worked with the research group CARISMA, Brunel University. With the Brunel team in general and Professor Mitra in particular she has participated in the design of SPInE : Stochastic programming integrated environment and also in supply chain logistics studies under uncertainty. Her research activity is mainly focused on the development of models and methods for decision making under uncertainty and more recently on statistical relational models for data analysis.. She developed decision models for addressing uncertainty in different application domains such as supply chain management, finance, system biology, ambient intelligence, web mining, e-forensics.

Talk 8: Constraint driven information extraction for natural language processing

Prof. Enza Messina
Professor of Operations Research
Dept. of Informatics, Systems and Communication (DISCo).
Building U14 – DISCo, Room 2048,
University of Milano-Bicocca,
Viale Sarca 336, 20126 Milan (ITALY)
E-mail: messina@disco.unimib.it

Abstract:

Advanced analytics requires technologies for predicting human or social behavior by analyzing massive amount of data and integrating information coming from social media with corporate business data. In this scenario, natural language processing technologies are needed for capturing various actions of people and society as reported in social media. This opens new challenges for OR practitioners.

In particular, the labeling of unstructured textual documents, to derive a structured representation of contents to be integrated with quantitative corporate data, represents a key challenge. The discovering of semantic information embedded within natural language documents - characterized by ambiguity and partial/imperfect information - can be viewed as a decision making process aimed at assigning a sequence of semantic labels to a set of interdependent variables. This problem can be modeled through a stochastic process involving both hidden variables (semantic labels) and observed variables (textual cues). We report the results of our investigation of one of the most recent and promising learning approach for semantic labeling, named Conditional Random Fields (CRFs).

We discuss how Integer Linear programming techniques can be applied to improve the performance of the inference procedure in CFR by enabling the inclusion of both specific domain knowledge and context knowledge learned from data.

The proposed approach has been validated by using a set of benchmark data and tested on real textual documents in the judicial domain.