

# Kernel methods and Support Vector Machines - Application to NLP Problems

Jean-Michel Renders  
Xerox Research Center Europe

Kernel methods and Support Vector Machine (SVM) have recently been introduced to solve Natural Language problems. Kernel methods define a generalized "similarity" measure between objects of arbitrary structure, with three interesting properties:

- First, kernels can be designed to incorporate prior knowledge about the domain – mainly linguistic knowledge, derived from syntactical and semantic analysis.
- Secondly, kernels implicitly map the objects into a new feature space, where problem solving is easier: in general, the mapped structure is simpler and linear methods can be used.
- Thirdly, the new feature space can have very high (possibly infinite) dimension, which allows for very rich representation, while still being computationally tractable and efficient (because the mapping is left implicit).

On the other hand, Support Vector Machines (and a family of techniques based on similar concepts) have emerged as a powerful supervised learning method, combining a strong theoretical basis (structural risk minimization), natural coupling with kernels, efficient implementations and performances proved to be excellent, even for high dimensional problems and for small data sets.

After attending this tutorial, you should :

- Understand the fundamentals and the important ideas behind Kernel Methods and SVMs, with the help of illustrating examples in the domain of NLP; important NLP-kernels such as "bag-of-words kernels", implicit semantic kernels, string kernels, word-sequence kernels, tree and DAG kernels will be clear for you.
- Be aware of existing applications of both methods (individually or in combination) in Natural Language Problems. These applications cover tasks in Document Categorization and Clustering, Parsing and structural disambiguation, Part-of-speech tagging and Entity tagging/typing. In each case, you will know how to design specific kernels and how to couple them with SVMs or other classifiers, as well as their pros and cons in comparison with more traditional methods.
- Have a good idea of perspectives and potential impacts of these methods in solving NLP problems.