Università degli Studi di Milano
Master Degree in Computer Science

Information Management course

Teacher: Alberto Ceselli

Lecture 07: 06/11/2012

L. C. Molina, L. Belanche, A. Nebot
"Feature Selection Algorithms: A Survey and Experimental Evaluation", IEEE ICDM (2002)

and

L. Belanche, F. Gonzales "Review and Evaluation of Feature Selection Algorithms in Synthetic Problems", arXiv – available online (2011)

Feature Selection Algorithms

- Introduction
- Relevance of a feature
- Algorithms
- Description of fundamental FSAs
- Empirical evaluation
- Experimental evaluation

Introduction

- The Feature selection problem:
  - Given a set of candidate features, select a subset defined by one of the following approaches:
    - Having a fixed size and maximizing an evaluation measure;
    - Of smaller size that satisfies a constraint on an evaluation measure
    - Best tradeoff between size and evaluation measure

- FSA are motivated by a definition of *relevance* (not obvious)

- FSAs can be classified according to their output
  1) Giving a weighted linear order of features
  2) Giving a subset of original features (the one we focus on)
  - N.B. (2) is (1) with binary weighting

Feature Selection Algorithms

- Introduction
- Relevance of a feature
- Algorithms
- Description of fundamental FSAs
- Empirical evaluation
- Experimental evaluation

Relevance with respect to an objective

- Relevance must be defined with respect to an objective: assuming the objective is classification and the set of features is X:
  - A feature x ∈ X is relevant to an objective c() if there exist two examples A and B that
    - differ only in the value of x
    - c(A) ≠ c(B)
  - i.e. there are two elements that can be classified correctly only thanks to x
- However, our datasets are samples in the feature space:
  - A feature x ∈ X is strongly relevant to the sample S to an objective c() if there exist two elements A and B of S that
    - differ only in the value of x
    - c(A) ≠ c(B)
  - A feature x is weakly relevant if there exists a X' ⊂ X with x ∈ X', where x is strongly relevant with respect to S

Relevance as a complexity measure

- Idea: given a data sample S and an objective c(), define r(S,c) as the smallest number of relevant features to c() such that the error in S is the least possible for the inducer
- i.e. the smallest number of features required by a specific inducer to reach optimum performance in modeling c() using S
- Examples of such complexity measures:
  - Incremental usefulness:
    after choosing X', x is useful if the accuracy of c() computation is higher on x U X' than on X'
  - Entropic relevance:
    compute the amount of (Shannon) entropy in the dataset before and after the removal of a feature