Inductive Fuzzy Classification

Overview

By Dr. Michael A. Kaufmann
Introduction: Fuzzy Classification

- The sorites paradox: How many grains does it take to constitute a heap? (Hyde 2012)
- This puzzle can be resolved by fuzzy classification. In general, fuzzy classification can be defined as any procedure which determines a degree of membership of objects $x$ in classes $y$
- (Meier et al. 2008)
Inductive Fuzzy Classification (IFC)

- If a membership function $\mu_y: \text{dom}(x) \rightarrow [0,1]$ is given a priori, the degree of classification of $x$ in $y$ is equal to $\mu_y(x)$.
- In contrast, if the membership function is not given a priori, how can it be induced from data? That task is called membership function induction (MFI).
- Inductive logic can be applied to induce membership functions from sample data. The likelihood ratio can be normalized in $[0,1]$ by adding the numerator to the denominator (Normalized Likelihood Ratio, NLR).

\[ \mu_y(x) \overset{\text{def}}{=} \frac{p(x|y)}{p(x|y) + p(x|\neg y)} \in [0,1] \]

- The algorithm can be extended to cover numerical attributes.
Multivariate IFC

- **Membership function induction**: For every input variable $X_k$, a membership function $\mu_y: X_k: \rightarrow [0,1]$ is estimated using the Target variable $Y$ and the NLR.

- **Inductive fuzzification**: Using the estimated membership functions, for every record $i$ the values of all input variables are transformed into estimated degrees of membership, $\mu_y(X_k(i))$.

- **Multivariate aggregation**: The $n$ degrees of membership are merged together, e.g. with a supervised multivariate model with numeric output or with a data fusion operator.
Empirical Results

- 30 data sets from the UCI machine learning repository were tested. Each target was predicted with logistic regression, with and without prior IFC of the input variables (Kaufmann 2012)
- Logistic regression was significantly improved by IFC on average, and this improvement is correlated significantly with the non-linearity of the prediction task
IFC: Implementation in WEKA

- Induces membership functions from data
- Visualizes these membership functions
- Can transform the data (inductive fuzzification)
- (Kaufmann & Graf 2012)
References


