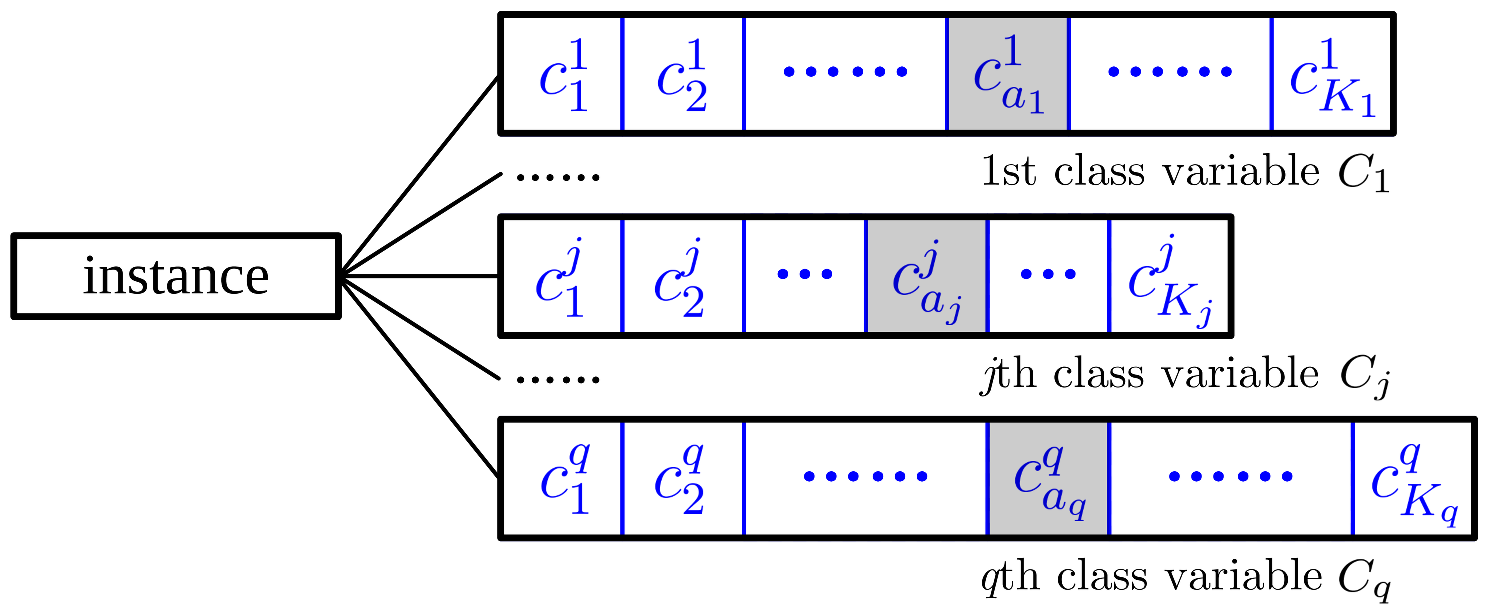
MD-KNN: An Instance-based Approach for Multi-Dimensional Classification

Bin-Bin Jia (jiabb@seu.edu.cn)†, *Min-Ling Zhang*(zhangml@seu.edu.cn)§

Introduction



Multi-Dimensional Classification (MDC)

Settings

$\mathcal{X} = \mathbb{R}^d$: d-dimensional input (feature) space

$\mathcal{Y} = C_1 \times C_2 \times \cdots \times C_q$: output space, where $C_j = \{c_1^j, c_2^j, \ldots, c_{K_j}^j\}$

Inputs

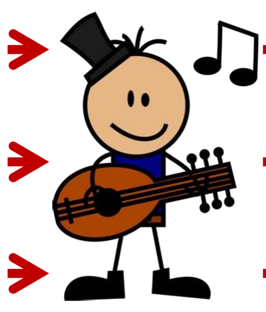
$\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \mid 1 \le i \le m\}$: training data set, where

$\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^\top \in \mathcal{X}$ and $\boldsymbol{y}_i = [y_{i1}, y_{i2}, \ldots, y_{iq}]^\top \in \mathcal{Y}$

Outputs

f : multi-dimensional classifier $\mathcal{X} \to \mathcal{Y}$

MDC example (A piece of music)

Dim. 1: Genre -->  --> rock, popular, classical, etc.

Dim. 2: Instrument --> --> piano, violin, guitar, etc.

Dim. 3: Language --> --> English, Chinese, Spanish, etc.

Existing Works: Adapting **parametric learning techniques** to solve the MDC problem (e.g., Bayesian learning, distance metric learning, maximum margin, etc.).

Our Work: Making a first attempt to adapt **instance-based techniques** for MDC.

The MD-KNN Approach

Step I: Obtain the unseen instance $\boldsymbol{x}_*$'s kNN counting statistics w.r.t. each class space (i.e., dimension):

- The kNN counting statistics $\boldsymbol{\delta}_j^{\boldsymbol{x}_*} = [\delta_{j1}^{\boldsymbol{x}_*}, \delta_{j2}^{\boldsymbol{x}_*}, \ldots, \delta_{jK_j}^{\boldsymbol{x}_*}]^\top$ for the j-th class space can be obtained as follows $(1 \le j \le q)$:

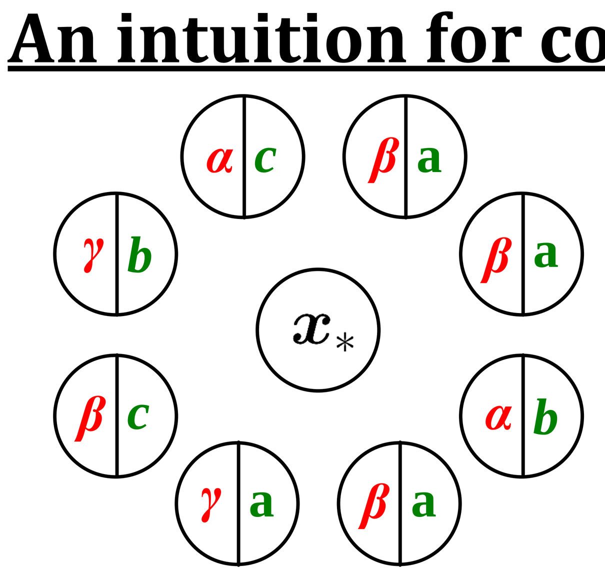
$$\delta_{ja}^{\boldsymbol{x}_*} = \sum_{i_t \in \mathcal{N}_k(\boldsymbol{x}_*)} \mathbb{I}(y_{i_t j}, c_a^j) \quad (1 \le a \le K_j)$$

$\mathcal{N}_k(\boldsymbol{x}_*) = \{i_t \mid 1 \le t \le k\}$ stores the set of indices for $\boldsymbol{x}_*$'s k nearest neighbors identified in training set $\mathcal{D}$;

$\boldsymbol{y}_{i_t} = [y_{i_t 1}, y_{i_t 2}, \ldots, y_{i_t q}]^\top$ corresponds to the class vector of the neighboring MDC example $\boldsymbol{x}_{i_t}$;

$\mathbb{I}(\pi_1, \pi_2)$ returns 1 if π_1 is identical with π_2 and 0 otherwise.

An intuition for computing kNN counting statistics:



Left: 1[st] dim.(α/β/γ)
Right: 2[nd] dim.(a/b/c)

$(k = 8, q = 2, K_1 = K_2 = 3)$

In the 8 nearest neighbors of $\boldsymbol{x}_*$:

1[st] dimension	2[nd] dimension
#Label **α**: 2	#Label **a**: 4
#Label **β**: 4	#Label **b**: 2
#Label **γ**: 2	#Label **c**: 2
$\boldsymbol{\delta}_1^{\boldsymbol{x}_*} = [2, 4, 2]$	$\boldsymbol{\delta}_2^{\boldsymbol{x}_*} = [4, 2, 2]$

The MD-KNN Approach (Con't)

Step II: For each pair of class spaces, MAP inference is made based on their kNN counting statistics.

- Because the posterior probability cannot be explicitly estimated from training set (see our paper for details), the MAP inference is approximately done as follows:

for the (r, s)-th class space pair $(1 \le r < s \le q)$:
Train g_{rs} over $\mathcal{D}_{rs}^{\mathrm{MAP}}$, i.e., $g_{rs} = \mathcal{M}(\mathcal{D}_{rs}^{\mathrm{MAP}})$
$[y_{*r}^{rs}, y_{*s}^{rs}] = \phi_{rs}^{-1}(g_{rs}(\boldsymbol{x}_*^{rs}))$ where $\boldsymbol{x}_*^{rs} = [\boldsymbol{\delta}_r^{\boldsymbol{x}_*}, \boldsymbol{\delta}_s^{\boldsymbol{x}_*}]$

$\mathcal{D}_{rs}^{\mathrm{MAP}} = \{(\boldsymbol{x}_i^{rs}, \phi(y_{ir}, y_{is})) \mid 1 \le i \le m\}$ and $\boldsymbol{x}_i^{rs} = [\boldsymbol{\delta}_r^{\boldsymbol{x}_i}, \boldsymbol{\delta}_s^{\boldsymbol{x}_i}]$

$\phi_{rs}(\cdot, \cdot) : C_r \times C_s \to \mathbb{N}$ and inverse function: $\phi_{rs}^{-1}(\cdot, \cdot)$

Our Paper

Step III: Determine the unseen instance $\boldsymbol{x}_*$'s class label w.r.t. each class space by synergizing the q-1 predictions via consulting empirical kNN accuracy.

- The q-1 predictions for $\boldsymbol{x}_*$'s j-th class space $(1 \le j \le q)$:

$$\boldsymbol{y}_j^{\boldsymbol{x}_*} = \left[y_{*j}^{1j}, \ldots, y_{*j}^{(j-1)j}, y_{*j}^{j(j+1)}, \ldots, y_{*j}^{jq}\right]^\top$$

$h_{1j}(\cdot) = \phi_{1j}^{-1}(g_{1j}(\cdot))$ $\quad h_{j(j+1)}(\cdot) = \phi_{j(j+1)}^{-1}(g_{j(j+1)}(\cdot))$ $\quad h_{(j-1)j}(\cdot) = \phi_{(j-1)j}^{-1}(g_{(j-1)j}(\cdot))$ $\quad h_{jq}(\cdot) = \phi_{jq}^{-1}(g_{jq}(\cdot))$

Here, each element is obtained by different classifier and the one obtained by the h with best kNN accuracy is returned.

Experiments

Experimental Setup

- **10 Benchmark Data sets** and **3 Evaluation metrics**
- **Compared Approaches:** BR, CP, ECC, ESC, gMML
- **Experimental Protocol:** Ten-fold cross-validation + Pairwise t-test

Experimental Results

Win/tie/loss counts of pairwise t-test (at 0.05 significance level) between MD-KNN and each MDC approach.

Evaluation metric	MD-KNN against				
	BR	CP	ECC	ESC	gMML
HS	8/1/1	4/1/1	8/1/1	5/1/1	8/1/1
EM	8/2/0	4/1/1	7/3/0	4/2/1	8/2/0
SEM	6/3/1	3/3/0	6/3/1	4/3/0	6/3/1
In Total	**22/6/2**	**11/5/2**	**21/7/2**	**13/6/2**	**22/6/2**

- MD-KNN achieves superior or at least comparable performance against the five compared approaches in 119 cases.
- All the 10 under-performing cases and 23 out of 30 comparable cases for MD-KNN against other compared approaches occur for the three data sets with nominal features.
- More details about experimental results and some further analysis (effectiveness of algorithmic design, sensitivity analysis) , can be found in our paper.

Conclusion

A first attempt towards adapting instance-based techniques for MDC is investigated, and a novel approach named MD-KNN is proposed which makes use of instance-based techniques in two levels (MAP inference and empirical kNN accuracy).

School of Computer Science and Engineering, Southeast University, China†, §
College of Electrical and Information Engineering, Lanzhou University of Technology, China†
Key Laboratory of Computer Network and Information Integration, Ministry of Education, China †, §