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

Bin-Bin Jia **Min-Ling Zhang**

School of Computer Science and Engineering

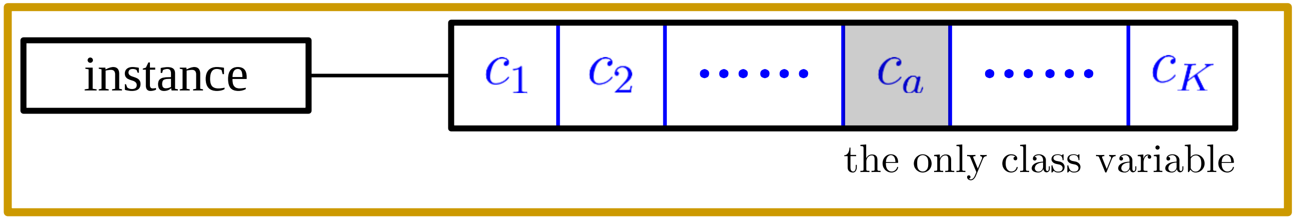
MOE Key Laboratory of Computer Network & Information Integration

Southeast University, China

Milan, Italy

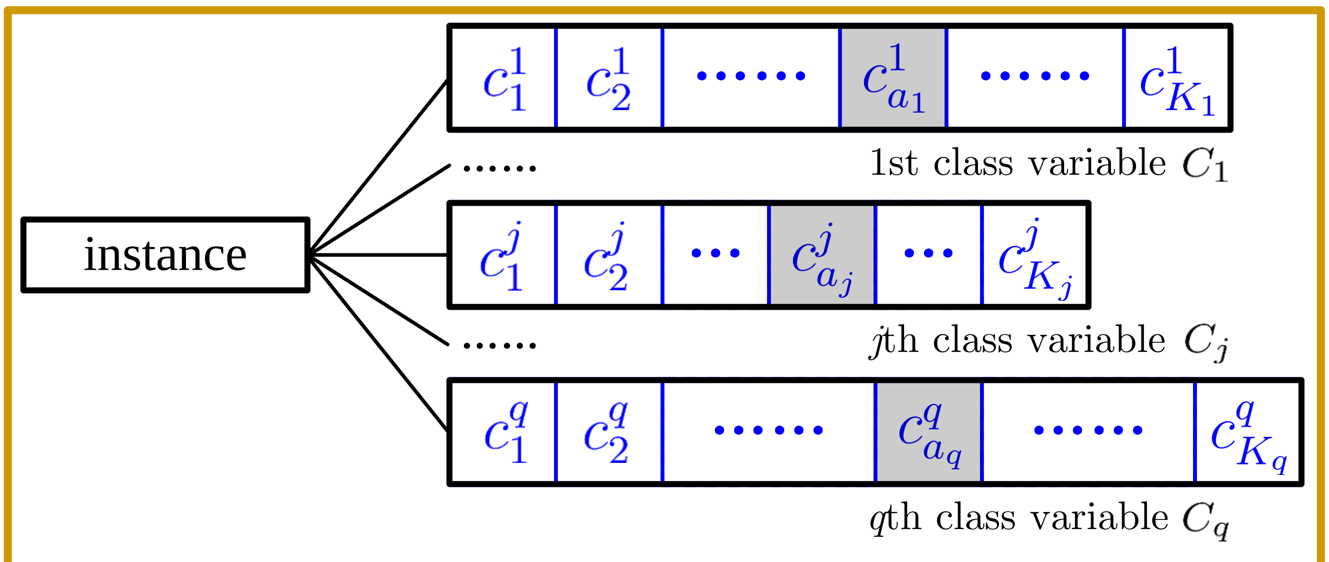
Multi-Dimensional Classification

Traditional Multi-class classification



- Only one class variable

Multi-Dimensional Classification (MDC)



- Multiple class variables

MDC examples

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

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

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

A piece of music

Dim. 1: Topic ---> sports, politics, social, Sci&Tech, etc.

Dim. 2: Mood ---> good news, neutral news, bad news

Dim. 3: Zone ---> domestic, intra-/inter-continental

A news document

MDC examples

Dim. 1: Genre

rock, popular,

Widely exist in real-world applications

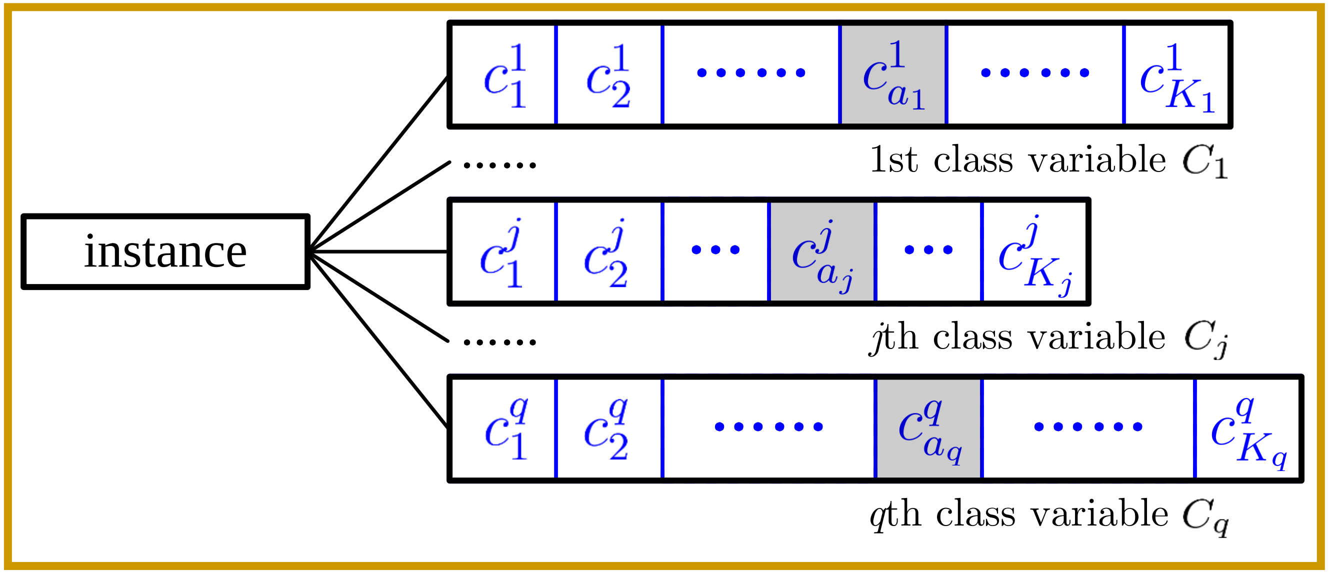
- Text classification [Ortigosa-Hernandez et al., Neurocomputing12] [Serafino et al., LNAI15] [Tu et al., ACM TIST17]
- Bioinformatics [Read et al., TKDE14] [Fernandez-Gonzalez et al., ICML Workshop'15] [Bolt et al., IJAR17][Benjumeda et al., IJAR18]
- Computer vision [Ma et al., Neurocomputing18]
- Resource allocation [Muktadir et al., IEICE TIS19]
- Other areas [Tekinerdogan, SoSE'19] [Verma et al. Sci Total Environ21]

Dim. 3: Zone

/inter-continental

A news document

The Problem



- Multiple class variables

Existing works: adapting parametric learning techniques to deal with MDC problem (e.g., probabilistic graphical model, distance metric learning, maximum margin, etc.).

Our work: make a first attempt to adapt **instance-based techniques** for MDC, and propose a novel approach named **MD-KNN**.

The MD-KNN Approach (1/4)

The workflow of MD-KNN:

(I). MD-KNN identifies ***k* nearest neighbors** of unseen instance in training set, and obtains its ***k*NN counting statistics** w.r.t. each class space.

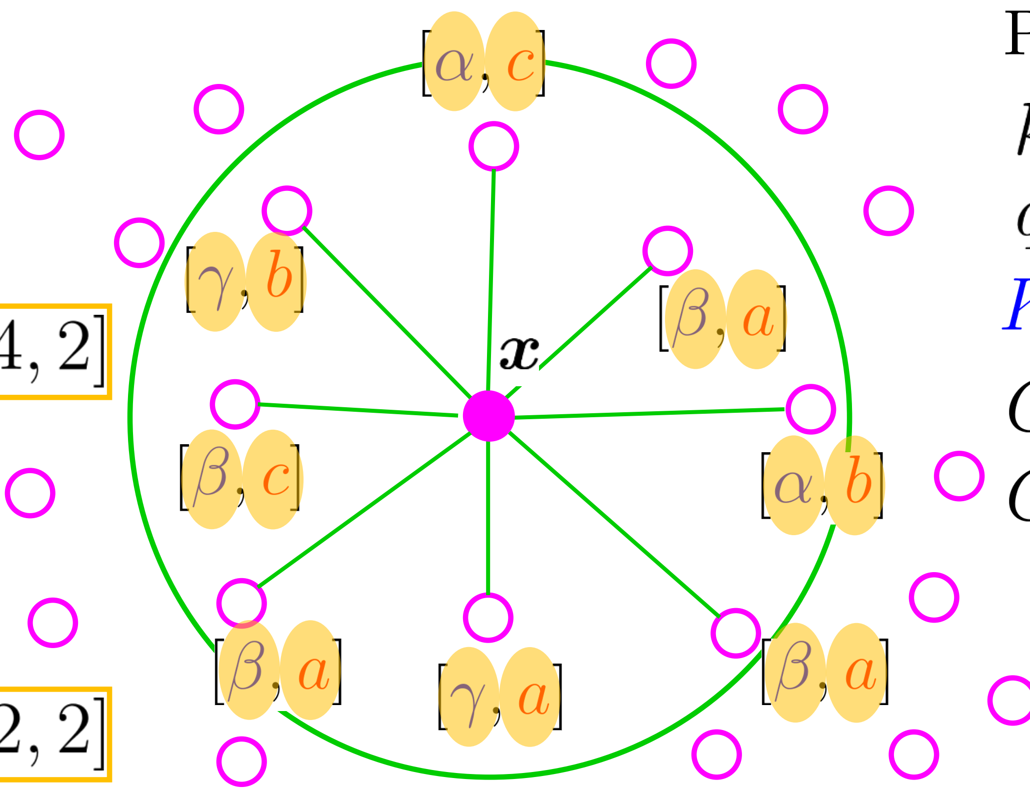
(II). For **each pair** of class spaces, **maximum a posteriori** (MAP) inference is made based on the obtained *k*NN counting statistics w.r.t. both class spaces.

(III). The class label w.r.t. each class space is determined by synergizing predictions from corresponding pairwise class spaces via consulting **empirical *k*NN accuracy**.

The MD-KNN Approach (2/4)

(I). MD-KNN identifies ***k* nearest neighbors** of unseen instance in training set, and obtains its ***k*NN counting statistics** w.r.t. each class space.

$$\delta_{11}^{\boldsymbol{x}} = 2$$
$$\delta_{12}^{\boldsymbol{x}} = 4$$
$$\delta_{13}^{\boldsymbol{x}} = 2$$
$$\boxed{\boldsymbol{\delta}_1^{\boldsymbol{x}} = [2, 4, 2]}$$
$$\delta_{21}^{\boldsymbol{x}} = 4$$
$$\delta_{22}^{\boldsymbol{x}} = 2$$
$$\delta_{23}^{\boldsymbol{x}} = 2$$
$$\boxed{\boldsymbol{\delta}_2^{\boldsymbol{x}} = [4, 2, 2]}$$



Parameters:

$$k = 8$$
$$q = 2$$
$$K_1 = K_2 = 3$$
$$C_1 = \{\alpha, \beta, \gamma\}$$
$$C_2 = \{a, b, c\}$$

The MD-KNN Approach (3/4)

(II). For **each pair** of class spaces, **maximum a posteriori** (MAP) inference is made based on the obtained *k*NN counting statistics w.r.t. both class spaces.

for $r = 1$ to $q-1$ **do** % *q*: number of dimensions

for $s = r+1$ to q **do**

$$y_{*r}^{rs}, y_{*s}^{rs} \leftarrow \underset{c_{a_r}^r \in C_r, c_{a_s}^s \in C_s}{\arg\max} \mathbb{P}\left(c_{a_r}^r, c_{a_s}^s \mid \boldsymbol{\delta}_r^{\boldsymbol{x}_*}, \boldsymbol{\delta}_s^{\boldsymbol{x}_*}\right)$$

end for

end for

Note1: Consider pairwise class dependencies

Note2: Obtain *q*-1 predicted class labels for each dimension

The MD-KNN Approach (3/4)

(II). For **each pair** of class spaces, **maximum a posteriori** (MAP) inference is made based on the obtained *k*NN counting statistics w.r.t. both class spaces.

for $r = 1$ to $q-1$ **do** % *q*: number of dimensions

  for $s = r+1$ to q **do**

$$y_{*r}^{rs}, y_{*s}^{rs} \leftarrow \arg\max \ \mathbb{P}\left(c_{a}^{r}, c_{a}^{s} \mid \boldsymbol{\delta}^{\boldsymbol{x}_*}, \boldsymbol{\delta}^{\boldsymbol{x}_*}\right)$$

  end for

end for

Cannot be effectively estimated from training set due to combinatorial complexity

Note1: Consider pairwise class dependencies

Note2: Obtain *q*-1 predicted class labels for each dimension

The MD-KNN Approach (3/4)

(II). For **each pair** of class spaces, **maximum a posteriori** (MAP) inference is made based on the obtained *k*NN counting statistics w.r.t. both class spaces.

$$
\begin{aligned}
&\textbf{for } r = 1 \text{ to } q-1 \textbf{ do} \quad \text{\% } q\text{: number of dimensions}\\
&\quad \textbf{for } s = r+1 \text{ to } q \textbf{ do}\\
&\qquad \text{Train } g_{rs} \text{ over } \mathcal{D}_{rs}^{\text{MAP}}, \text{ i.e., } g_{rs} = \mathcal{M}(\mathcal{D}_{rs}^{\text{MAP}})\\
&\qquad [y_{*r}^{rs}, y_{*s}^{rs}] = \phi_{rs}^{-1}(g_{rs}(\boldsymbol{x}_*^{rs})) \quad \text{where } \boldsymbol{x}_*^{rs} = [\boldsymbol{\delta}_r^{\boldsymbol{x}_*}, \boldsymbol{\delta}_s^{\boldsymbol{x}_*}]\\
&\quad \textbf{end for}\\
&\textbf{end for}
\end{aligned}
$$

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

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

The MD-KNN Approach (4/4)

(III). The class label w.r.t. each class space is determined by synergizing predictions from corresponding pairwise class spaces via consulting **empirical *k*NN accuracy**.

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

$$\boldsymbol{y}_j^{\boldsymbol{x}_*} = \left[y_j^{\boldsymbol{x}_*}(1), y_j^{\boldsymbol{x}_*}(2), \ldots, y_j^{\boldsymbol{x}_*}(q-1)\right]^\top$$

$$\text{where } y_j^{\boldsymbol{x}_*}(i) = \begin{cases} y_{*j}^{ij}, & \text{when } 1 \le i \le j-1 \\ y_{*j}^{j(i+1)}, & \text{when } j \le i \le q-1 \end{cases}$$

obtained by $h_{ij}(\cdot) = \phi_{ij}^{-1}(g_{ij}(\cdot))$

obtained by $h_{j(i+1)}(\cdot) = \phi_{j(i+1)}^{-1}(g_{j(i+1)}(\cdot))$

The MD-KNN Approach (4/4)

(III). The class label w.r.t. each class space is determined by synergizing predictions from corresponding pairwise class spaces via consulting **empirical *k*NN accuracy**.

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

$$\boldsymbol{y}_j^{\boldsymbol{x}_*} = \left[y_j^{\boldsymbol{x}_*}(1), y_j^{\boldsymbol{x}_*}(2), \ldots, y_j^{\boldsymbol{x}_*}(q-1)\right]^\top$$

Return the one corresponding to the *h* with best accuracy over $\boldsymbol{x}_*$'s *k* nearest neighbors

where

obtained by $h_{ij}(\cdot) = \phi_{ij}^{-1}(g_{ij}(\cdot))$

obtained by $h_{j(i+1)}(\cdot) = \phi_{j(i+1)}^{-1}(g_{j(i+1)}(\cdot))$

Experimental Setup

Experimental data sets

Characteristics of the experimental MDC data sets.

Data Set	#Exam.	#Dim.	#Labels/Dim.	#Features†
Flare1	323	3	3,4,2	$10x$
WaterQuality	1060	14	4	$16n$
Scm20d	8966	16	4	$61n$
Rf1	8987	8	4,4,3,4,4,3,4,3	$64n$
Thyroid	9172	7	5,5,3,2,4,4,3	$7n$, $22x$
Pain	9734	10	2,5,4,2,2,5,2,5,2,2	$136n$
Scm1d	9803	16	4	$280n$
Disfa	13095	12	5,5,6,3,4,4,5,4,4,4,6,4	$136n$
Fera	14052	5	6	$136n$
Adult	18419	4	7,7,5,2	$5n$,$5x$

† n, x denote numeric and nominal features respectively.

Experimental Setup

Evaluation Metrics

Testing set: $\mathcal{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) \mid 1 \leq i \leq p\}$, where $\boldsymbol{y}_i = [y_{i1}, y_{i2}, \ldots, y_{iq}]^\top$

Predicted class vector: $\hat{\boldsymbol{y}}_i = f(\boldsymbol{x}_i) = [\hat{y}_{i1}, \hat{y}_{i2}, \ldots, \hat{y}_{iq}]^\top$

For each MDC test example $(\boldsymbol{x}_i, \boldsymbol{y}_i): \; r^{(i)} = \sum_{j=1}^{q} [\![y_{ij} = \hat{y}_{ij}]\!]$

Hamming Score:

$$\mathrm{HS}_{\mathcal{S}}(f) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{q} \cdot r^{(i)}$$

Exact Match:

$$\mathrm{EM}_{\mathcal{S}}(f) = \frac{1}{p} \sum_{i=1}^{p} [\![r^{(i)} = q]\!]$$

Sub-Exact Match:

$$\mathrm{SEM}_{\mathcal{S}}(f) = \frac{1}{p} \sum_{i=1}^{p} [\![r^{(i)} \geq q - 1]\!]$$

Experimental Setup

Compared Algorithms

BR: Learn q independent multi-class classifier, one per dimension

CP: Learn a single multi-class classifier via powerset transformation

ECC: Learn a chain of q multi-class classifier, one per dimension

ESC: Group the class variables into groups

gMML: Learn a regressor for each class label as well as a Mahalanobis distance metric to train all regressor in a joint manner

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**

Detailed experimental results and some further analysis (effectiveness of algorithmic design, sensitivity analysis) can be found in our paper.

Thanks！

http://palm.seu.edu.cn/zhangml/files/MD-kNN.zip

Email: jiabb@seu.edu.cn