MPRSS 2020

6th IAPR Workshop on Multimodal pattern recognition for social signal processing in human computer interaction

Introducing Bidirectional Ordinal Classifier Cascades Based on a Pain Intensity Recognition Scenario

ulm university universität ulm

Peter Bellmann and Friedhelm Schwenker
Institute of Neural Information Processing

Ludwig Lausser and Hans A. Kestler
Institute of Medical Systems Biology

Agenda

- Formalisation & Error Correcting Output Codes
- Ordinal Classification & Ordinal Classifier Cascades
- BioVid Heat Pain Database
- Results
- Conclusion & Future Work

FORMALISATION & ERROR CORRECTING OUTPUT CODES

Formalisation

- $X \subset \mathbb{R}^d$, $d \in \mathbb{N}$: d-dimensional data set
- $\Omega = \{\omega_1, \dots, \omega_c\}$, $c \geq 3$: class label set
- $l(x)$: true label of data sample x

- $X_{i,j} \coloneqq \left\{x \in X \middle| l(x) = \omega_i \vee l(x) = \omega_j\right\} \subset X$
- $CM_{i,j}$: classification model that is trained in combination with $X_{i,j}$

- It holds: $CM_{i,j} = CM_{j,i} \ \forall\, i, j \in \{1, \dots, c\}$ (symmetric models)

Error Correcting Output Codes (ECOC)

- ECOC ensembles are popular divide-and-conquer approaches
- Each ensemble member is trained on a (binary) subtask
- ECOC Approaches
  - One versus All (1vsA)
  - One versus One (1vs1)
  - All Binary Combinations (ABC)
  - Ternary, etc.

ECOC – Prediction

- Each ensemble member provides a $\{0,1\}$-output
- Each class is represented by a $\{0,1\}$-code (vector)
- Compute the distance between ensemble output and class code
  - E.g., Hamming distance
  - Class corresponding to minimum distance is taken as prediction

ORDINAL CLASSIFICATION & ODINAL CLASSIFIER CASCADES

Ordinal Classification

- Ordinal classification $\nRightarrow$ ordinal-scaled features
- Ordinal classification $\Rightarrow$ ordinal class structure
- Notation: $\omega_1 \prec \cdots \prec \omega_c$
- We denote ω_1 and ω_c as edge classes or edges

Ordinal Classifier Cascades (OCCs)

- OCC architectures also represent divide-and-conquer techniques
- Each ensemble member is trained on a binary subtask
- The ensemble members constitute a classifier sequence

- Each ensemble member provides one of two options
  - Move the current input to the next model in the sequence
  - Take the current output as the ensemble‘s final decision
  - Exception: Last sequence member

OCCs – General Pairwise Approach

- CM 1 is trained to separate the first class from the second
- CM 2 is trained to separate the second class from the third
- Etc.

- Prediction
  - If CM 1 predicts the 1st class, take the prediction as final decision
  - Otherwise, move the current input to the second CM
  - Etc.

OCCs – Pairwise Approaches

- Motivation: The direction of OCCs is important!
- $\omega_1 \prec \cdots \prec \omega_c$ is equivalent to $\omega_c \prec \cdots \prec \omega_1$
- Separate the pairwise approach into
  - Current versus Next (CvsN)
  - Current versus Previous (CvsP)

OCCs – Pairwise Approaches

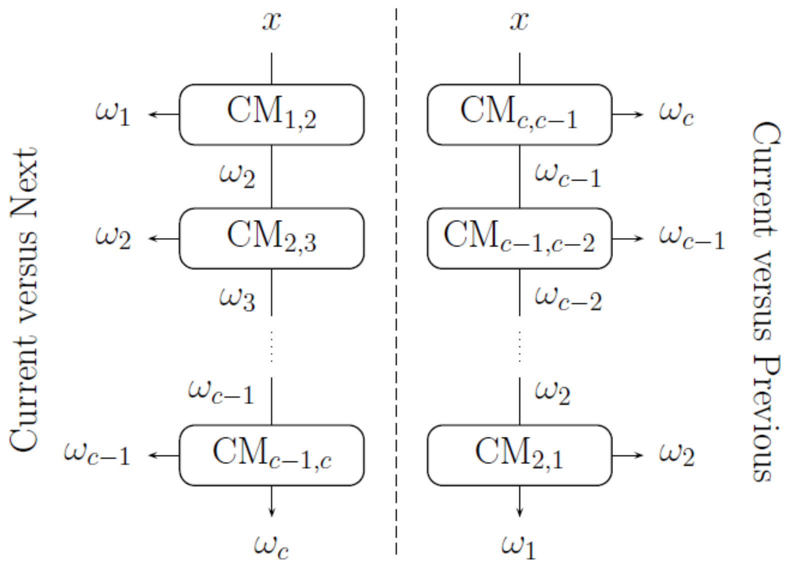

Fig. 1. Cascaded Classification Architectures. Left: Current vs. Next (CvsN). Right: Current vs. Previous (CvsP). x: Input. c: Number of classes. $\omega_1, \ldots, \omega_c$: Class labels. $\mathrm{CM}_{i,j}$: Classification model that separates the classes ω_i and ω_j. Arrowheads indicate that the corresponding output is taken as the architecture's final prediction.

Bidirectional Ordinal Classifier Cascades (bOCCs)

- Simple idea: Combine the pairwise approaches, CvsN & CvsP
- Selector component: $CM_{1,c}$

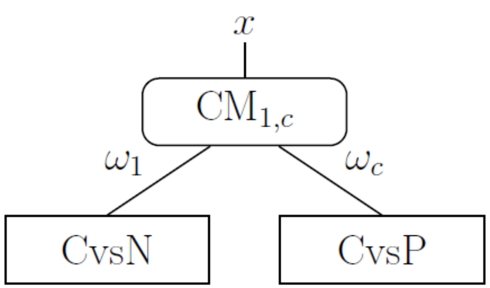

Fig. 2. Bidirectional Ordinal Classifier Cascade Architecture. x: Input. c: Number of classes. $\mathrm{CM}_{1,c}$: Classification model that separates the edge classes ω_1 and ω_c. CvsN/CvsP: Current vs. Next/Current vs. Previous architectures (see Fig. 1).

- Study Question:
  How does this straightforward modification affect the classification performance?

BIOVID HEAT PAIN DATABASE

BioVid Heat Pain Database (Part A) – General Information

- Consists of 5 publicly available parts (Part A – Part E)
- Healthy test subjects participated in pain elicitation experiments
- Pain was induced in form of heat by a professional thermode
- The thermode was attached to the participant‘s forearm
- Maximum allowed temperature was set to $50.5°C$

BioVid Heat Pain Database (Part A) – Individual Calibration

- T_0: Baseline Temperature – $32°C$ for each participant
- T_1: Pain Threshold – participant feels a change from heat to pain
- T_4: Tolerance Threshold – participant classifies pain as unbearable
- T_2/T_3: Intermediate Pain Levels – all pain levels are equidistant (not T_0)

- Naturally occurring order of class labels: $T_0 \prec T_1 \prec T_2 \prec T_3 \prec T_4$

BioVid Heat Pain Database (Part A) – Pain Experiments

- Each pain level was applied 20 times per participant in randomised order
- Each pain level was held for 4 seconds
- Each pain level was followed by a baseline stimulus of 8-12 seconds

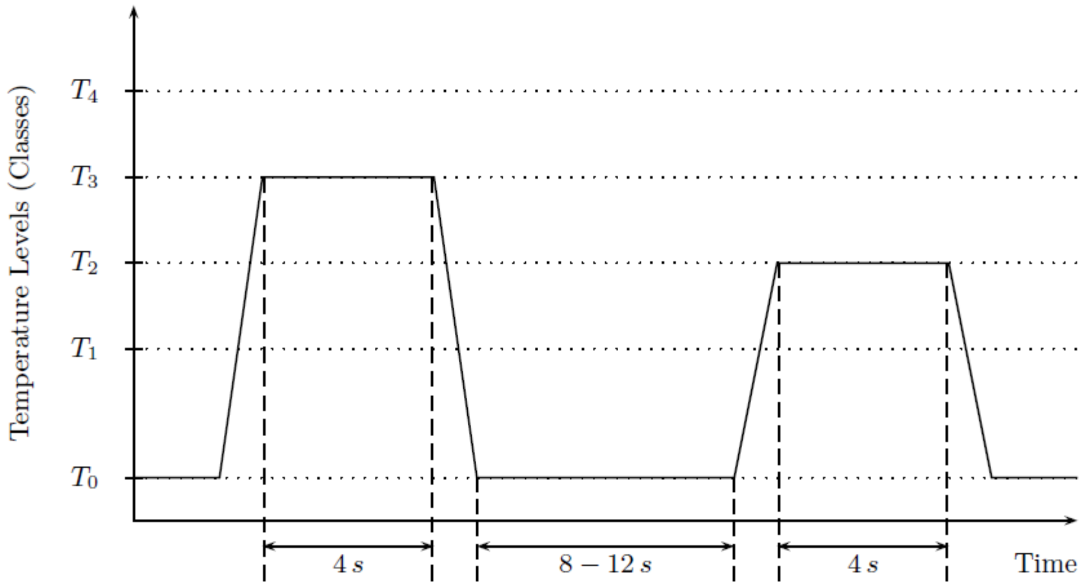

Fig. 3. An example for a participant's stimuli and recovering sequence.

BioVid Heat Pain Database (Part A) – Recorded Modalities

- Videos from 3 different angles (not used in the experiments)
- Electrocardiogram (ECG)
- Electromyogram (EMG) of the trapezius muscle (located at the shoulder)
- Electrodermal Activity (EDA) measured at the ring and index fingers

Table 2. Properties of the BioVid Heat Pain Database. Num. of samples: 8700.

Number of participants	87 (43 f, 44 m)
Number of classes	5 (T_0, T_1, T_2, T_3, T_4)
Number of samples per class and participant	20
Number of features (ECG + EDA + EMG)	194 (68 + 70 + 56)

RESULTS

Results – Experimental Settings

- **Classification Models**
  Support Vector Machines with linear kernel
- **Fusion Approach**
  Feature fusion, i.e. concatenate ECG, EDA, EMG data to one vector
- **Evaluation Approach**
  Leave-One-Person-Out (LOPO) Cross Validation

Results – Experimental Settings

- **Performance Measures**
  - Mean Absolute Error (MAE)
  - Accuracy
- **Statistical Testing**
  - Two-sided Wilcoxon signed-rank test
  - Significance level 5%

Results – CvsN & CvsP

Table 3. CvsN & CvsP: Confusion tables with absolute values. Rows denote true labels. Columns denote predicted labels. Σ: Sum of predictions. Left: CvsN. Right: CvsP. The averaged LOPO cross validation accuracies are equal to 34.54% and 34.69%, for the CvsN and CvsP approaches, respectively. Bold figures denote the number of correctly classified test samples. Each class consists of 1740 test samples.

	Current vs. Next					Current vs. Previous				
	T_0	T_1	T_2	T_3	T_4	T_0	T_1	T_2	T_3	T_4
T_0	**1035**	418	182	72	33	**573**	312	401	319	135
T_1	674	**551**	302	153	60	384	**397**	441	378	140
T_2	561	488	**352**	216	123	254	317	**470**	465	234
T_3	411	414	316	**291**	308	130	213	380	**532**	485
T_4	300	247	198	219	**776**	41	89	203	361	**1046**
$\sum$	2981	2118	1350	951	1300	1382	1328	1895	2055	2040

Results – Comparison of all Methods

Table 6. Averaged LOPO performance values (± standard deviations). An asterisk (*) denotes a statistically significant difference, in comparison to our proposed method (bOCC), according to the two-sided Wilcoxon signed-rank test, at a significance level of 5%. The best performing method is underlined. The accuracy is given in %. Chance level accuracy: 20%. MAE: Mean absolute error. #: Number of (classifiers).

Approach	Accuracy	MAE	#Classifiers
CvsN	$34.54 \pm 8.06^*$	$1.139 \pm 0.229^*$	4
CvsP	$34.69 \pm 9.31^*$	$1.063 \pm 0.258^*$	4
1vsA	$35.11 \pm 7.81^*$	$1.155 \pm 0.267^*$	5
bOCC	$\underline{37.01 \pm 8.97}$	1.040 ± 0.267	5
1vs1	36.95 ± 9.41	$\underline{1.002 \pm 0.259^*}$	10
ABC	36.29 ± 8.07	$1.098 \pm 0.279^*$	15

Conclusion & Future Work

- **Conclusion**
  - The direction of an OCC sequence has a significant impact
  - The addition of one single CM significantly improves the performance
  - The only superior performing method was the 1vs1 approach (MAE)
- **Future Work**
  - Optimisation of the selector component
  - Combination of 1vs1 and bOCC

THANK YOU VERY MUCH FOR LISTENING TO ME ☺

PLEASE FEEL FREE TO ASK YOUR QUESTIONS!