Automatic Pain Intensity Estimation with Heteroscedastic Conditional Ordinal Random Fields

Ognjen Rudovic[1], Vladimir Pavlovic[2], and Maja Pantic[1,3]

[1] Comp. Dept., Imperial College London, UK
[2] Dept. of Computer Science, Rutgers University, USA
[3] EEMCS, University of Twente, The Netherlands

Abstract. Automatic pain intensity estimation from facial images is challenging mainly because of high variability in subject-specific pain expressiveness. This heterogeneity in the subjects causes their facial appearance to vary significantly when experiencing the same pain level. The standard classification methods (e.g., SVMs) do not provide a principled way of accounting for this heterogeneity. To this end, we propose the *heteroscedastic* Conditional Ordinal Random Field (CORF) model for automatic estimation of pain intensity. This model generalizes the CORF framework for modeling sequences of ordinal variables, by adapting it for heteroscedasticity. This is attained by allowing the variance in the ordinal probit model in the CORF to change depending on the input features, resulting in the model able to adapt to the pain expressiveness level specific to each subject. Our experimental results on the UNBC Shoulder Pain Database show that modeling heterogeneity in the subjects with the framework of CORFs improves the pain intensity estimation attained by the standard CORF model, and the other commonly used classification models.

1 Introduction

Automatic analysis of pain has received increased attention over the last few years mostly because of its applications in health care. For example, in intensive care units in hospitals, it has recently been shown that enormous improvements in patient outcomes can be gained from the medical staff periodically monitoring patient pain levels. However, due to the burden of work/stress that the staff are already under, this type of monitoring has been difficult to sustain, so an automatic system would be an ideal solution [1]. Recent research has evidenced the usefulness of facial cues for automatic pain analysis (e.g., see [2]), however, it has mainly focused on detection of presence/absence of pain.

In this paper, we address the problem of estimating the level of patients' shoulder pain from video recordings of their facial expressions, provided by the recently released UNBC-MacMaster Shoulder Pain Expression Archive Database [2]. The recorded patients suffer from chronic shoulder pain, intensity of which is quantified into discrete ordinal levels, ranging from no pain to the maximal level of pain, measured using the Prkachin and Solomon Pain Intensity (PSPI) metric [3]. As the patients perform a range of arm motion tests in front of the camera, the aim being to estimate their pain level in every frame of the video. This poses a number of challenges for the modeling task. First,

different pain levels are characterized by subtle changes of facial appearance within subjects, and large changes of facial appearance between subjects. This is because the latter depends on what constitutes the maximal level of the change in facial appearance of each subject. Consequently, the between-subject variation can easily overshadow the pain-intensity-related variation. Second, subjects' facial expressions are typically toned down due to a long-term exposure to chronic pain. Therefore, it is important to account for temporal dynamics of pain intensity changes.

To the best of our knowledge, only a few works ([1,4,5]) have addressed the problem of automatic pain intensity estimation so far. Lucey et al. [1] proposed a system for a three-level pain intensity estimation at the sequence level. The authors used the shape- and appearance-based features obtained using an Active Appearance Model (AAM). These features were then used to train separate Support Vector Machine (SVM) classifiers for each pain intensity level. To deal with spurious noisy signals, a moving-average smoothing filter was applied to the SVM output probability scores. Kaltwang et al. [4] proposed a feature-fusion approach for continuous pain intensity estimation based on the Relevance Vector Regression (RVR) model. As the input, the authors used the shape features, and the appearance features, obtained by computing the Discrete Cosine Transform (DCT) and Local Binary Patterns (LBPs) from the normalized facial appearance. As the targets for the regression model, the authors used the discrete pain intensity levels defined on a 16 point scale. Finally, Hammal and Cohn [5] performed the estimation of 4 pain intensity levels. The authors applied Log-Normal filters to the normalized facial appearance, which resulted in high-dimensional facial features. These features were then used to separately train SVMs for each pain intensity on a frame-by-frame basis. Note that the works mentioned above focus mainly on the feature extraction step. The classification/regression of the target pain intensity is performed consequently by applying the standard learning techniques for nominal data, therefore ignoring the fact that pain intensity is defined on the ordinal scale. Finally, none of these methods explore temporal dynamics of the pain intensity.

In this paper, we propose a model for pain intensity estimation that is based on the Conditional Ordinal Random Field (CORF) [6,7] model, specifically designed for estimation of sequences of ordinal variables. Although the CORF model can address the limitations of the existing methods mentioned above, its underlying assumption is that the noise on the ordinal targets (in our case, the pain intensity) is homogeneous, i.e., constant. To account for heterogeneity in the subjects, we need to relax this assumption. This is attained by allowing its variance to change depending on the input features, resulting in the *heteroscedastic* CORF model. In contrast to the existing methods for pain intensity estimation, the proposed model is able to adapt to varying pain expressiveness levels of different subjects. The benefit of this is reflected in the results of the experiments conducted on the ShoulderPain dataset [2].

The remainder of the paper is organized as follows. Sec.2 reviews standard ordinal regression models. In Sec.3 we introduce the proposed heteroscedastic CORF model. Sec.4 shows the results of the experimental evaluation, and Sec.5 concludes the paper.

2   Ordinal Regression Models

Different models for data with ordinal targets have been proposed (e.g., see [8] for an overview). In this paper we restrict the consideration to the popular probit threshold model proposed by McCullagh(1980) [9]. In this model, it is assumed that there is a latent continuous variable Y^* that underlies the observed ordinal response Y. For example, in the context of the target task, Y represents intensity of pain described as 'none', 'moderate' or 'severe'. These outcomes may literally be considered as resulting from pain severity, the unobserved continuous latent response Y^*. Since we are interested in the intensity of pain, we need to model the relationship between the unobserved variable Y^* (i.e., the latent process causing pain) and the observed response Y (i.e., the intensity of pain). This relationship can be expressed using the following probit threshold models.

2.1   Homoscedastic Threshold Model

Let $Y^* = f(x) + \sigma Z$ be a 1-D continuous latent variable, where x is a vector of covariates (i.e., image features), where $f : X \to \mathbb{R}$ and Z is a noise variable with the standard normal distribution $\mathcal{N}(0,1)$. The probability distribution function of Y^* is then given by $\Pr(Y^* \leq z) = \Phi\left(\frac{z-f(x)}{\sigma}\right)$, where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Under the threshold concept, the observed ordinal response Y is obtained as $Y := \{y \in \{1, ..., R\} | b_{y-1} < Y^* \leq b_y\}$, where $b_0 = -\infty \leq \cdots \leq b_R = \infty$ are increasing thresholds or cut-off points. The conditional probability of Y is then given by:

$$\Pr(Y = y|x) = \Phi\left(\frac{b_y - f(x)}{\sigma}\right) - \Phi\left(\frac{b_{y-1} - f(x)}{\sigma}\right). \tag{1}$$

2.2   Heteroscedastic Threshold Model

The homoscedastic threshold model has some limitations. In real-world data, the uncertainty of the labels may depend on the input x. That is, on some x the label y will almost certainly appear, and on other x the label Y may have nearly uniform distribution [8]. This can be leveraged by allowing the scale σ to depend on inputs x, i.e., $\sigma \equiv \sigma(x)$, where $\sigma : X \to \mathbb{R}_+$, with $\mathbb{R}_+$ denoting the set of positive real numbers. Using the notation from Sec.2.1, the continuous latent variable is defined as $Y^* = f(x) + \sigma(x)Z$. Then, the conditional distribution function of Y with heteroscedastic noise is

$$\Pr(Y = y|x) = \Phi\left(\frac{b_y - f(x)}{\sigma(x)}\right) - \Phi\left(\frac{b_{y-1} - f(x)}{\sigma(x)}\right), \tag{2}$$

where the uncertainty of labels is adjusted by the intensity of $\sigma(x)$.

3   Heteroscedastic Conditional Ordinal Random Fields

In this section, we present the proposed model for automatic pain intensity estimation. The model is based on the CORF model for temporal data with ordinal targets.

We extend this model by accounting for heterogeneity of subjects, which is incorporated in the model by using the modeling approach of heteroscedastic ordinal regression from the previous section.

3.1   The Model

Consider the standard Conditional Random Field (CRF) [10] model. It represents the conditional distribution $P(\mathbf{y}|\mathbf{x})$ as the Gibbs form clamped on the observation $\mathbf{x}$:

$$P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{Z(\mathbf{x}; \boldsymbol{\theta})} e^{s(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})}, \tag{3}$$

where $Z(\mathbf{x}; \boldsymbol{\theta}) = \sum_{\mathbf{y} \in \mathcal{Y}} e^{s(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})}$ is the normalizing partition function ($\mathcal{Y}$ is a set of all possible output configurations), and $\boldsymbol{\theta}$ are the model parameters[1] of the *score function*.

The choice of the output graph $G = (V, E)$ and the cliques critically affects the representational capacity and the inference complexity of the model. For simplicity, a linear-chain model with *node* cliques ($r \in V$) and *edge* cliques ($e = (r, s) \in E$) is often assumed. By letting $\{\mathbf{v}, \mathbf{u}\}$ be the parameters of the node features, $\boldsymbol{\Psi}_r^{(V)}(\mathbf{x}, y_r)$, and the edge features, $\boldsymbol{\Psi}_e^{(E)}(\mathbf{x}, y_r, y_s)$, respectively, the score function $s(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ can be expressed as the sum:

$$s(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = \sum_{r \in V} \mathbf{v}^\top \boldsymbol{\Psi}_r^{(V)}(\mathbf{x}, y_r) + \sum_{e=(r,s) \in E} \mathbf{u}^\top \boldsymbol{\Psi}_e^{(E)}(\mathbf{x}, y_r, y_s). \tag{4}$$

The score function in (4) has a great modeling flexibility, allowing the node and edge features to be chosen depending on the target task.

Node Features. In the CORF framework, the node features are defined using the homoscedastic ordinal regression model in (1). In our model, we use the heteroscedastic ordinal regression model, defined in (2), to set the node features as:

$$\mathbf{v}^T \boldsymbol{\Psi}_r^{(V)}(\mathbf{x}, y_r) \rightarrow \sum_{c=1}^{R} I(y_r = c) \cdot \left[\Phi\left(\frac{b_{y_r} - f(\mathbf{x}_r)}{\sigma(\mathbf{x}_r)} \right) - \Phi\left(\frac{b_{y_r-1} - f(\mathbf{x}_r)}{\sigma(\mathbf{x}_r)} \right) \right]. \tag{5}$$

By applying the Representer Theorem to the regularized negative log-likelihood in (9), we obtain the optimal functional form for the location model $f(\cdot)$ as

$$f(\mathbf{x}_*) = \sum_{i=1}^{S} \alpha_i k_f(\mathbf{x}_i, \mathbf{x}_*), \tag{6}$$

where $k_f(\cdot, \cdot)$ is a Mercer kernel, and S is the number of kernel bases. Similarly, the scale model $\sigma(\cdot)$ is obtained as

$$\sigma(\mathbf{x}) = \exp(\beta_0 + \sum_{i=1}^{M} \beta_i k_\sigma(\mathbf{x}_i, \mathbf{x}_*)), \tag{7}$$

---

[1] For simplicity, we often drop the dependency on $\boldsymbol{\theta}$ in notations.

where we also include an intercept β_0, so when the data do not exhibit heterogeneity (or they do, but to a lesser extent), we recover the homoscedastic ordinal model. Also, to guarantee the non-negativity of σ, we use the exponential form of the kernel function.

The most important aspect of using the varying scale $\sigma(\mathbf{x})$ is that the inputs x can now directly influence the locations of the thresholds b in the ordinal model, which are constant in the homoscedastic CORF model. In this way, the proposed model with heteroscedastic (ordinal) node features can automatically adapt its thresholds to account for individual differences in pain tolerance and/or the level of individual pain expressiveness.

Edge Features. The edge features are defined as in the standard CRF model, i.e., using the absolute difference between the features of the temporally neighbouring frames, resulting in

$$\mathbf{\Psi}_e^{(E)}(\mathbf{x}, y_r, y_s) = \left[I(y_r = k \,\wedge\, y_s = l) \right]_{R \times R} \otimes \left| \mathbf{x}_r - \mathbf{x}_s \right|, \tag{8}$$

where $I(\cdot)$ is the indicator function that returns 1 (0) if the argument is true (false) and $\otimes$ denotes the Kronecker product. The role of the edge features is to enforce smooth predictions of the pain intensities across time.

With the node and edge features as defined above, we arrive at the following optimization problem:

$$\underset{\theta}{\arg\min} \sum_{i=1}^{N} - \ln P(\mathbf{y}_i | f(\mathbf{x}_i), \sigma(\mathbf{x}_i), \theta) + \Omega(\theta), \tag{9}$$

where N is the number of the training image sequences, $\Omega(\theta)$ is the (kernel-inducing) regularizer, and $\theta = \{\mathbf{b}, \alpha, \beta, \mathbf{u}\}$ are the model parameters.

3.2 Regularizers

As the objective function in (9) is nonlinear and nonconvex, it is critical to regularize it to improve the model's performance. We apply L_2 regularizer to the kernel weights and parameters $\mathbf{u}$ in order to avoid diverging solutions. To encourage the latent coordinates $f(\mathbf{x})$ to be close in the latent space, we employ the widely used Laplacian regularizer for kernels:

$$\Omega(\|f\|_K) = \sum_{i,j} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 W_{ij} = 2\alpha^T KLK\alpha, \tag{10}$$

where K is computed using the kernel function k_f, and $L = D - W$ is the graph Laplacian, with $D_{ii} = \sum_j W_{ij}$. The similarity W is derived using the target labels y as

$$W_{ij} = 1 - \frac{|y_i - y_j|}{R - 1}, \quad y_i, y_j = 1, ..., R. \tag{11}$$

Notice that when the absolute difference between two pain intensities increases, the extent of distance enlargement in (11) increases accordingly. This regularization approach has been shown effective in other facial-expression-related modeling tasks (eg. see [7]).

3.3   Learning and Inference

To minimize the objective in (9), we use the quasi-Newton limited-memory BFGS method. We briefly describe the learning strategy. Initially, we set the scale models σ to 1 to form a homoscedastic model. This is accomplished by optimizing the parameters of the location model f, the ordinal thresholds $\mathbf{b}$ and the transition parameters $\mathbf{u}$. In the next step, we fix the parameters of the homoscedastic model and optimize w.r.t. the parameters of the scale model. In the final run, we optimize all the parameters simultaneously. The regularization parameters are found using a cross-validation procedure on the training set. Once the parameters of the model are estimated, the inference of test sequences is carried out using Viterbi decoding.
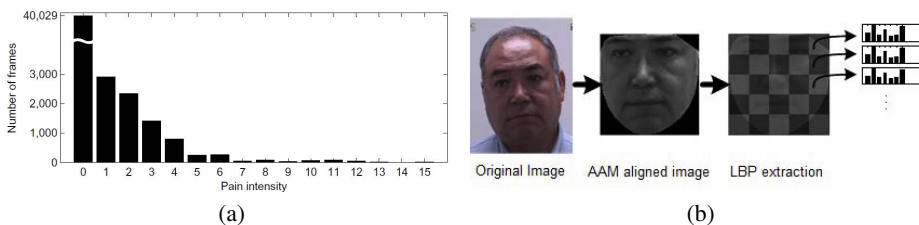

Fig. 1. a) Distribution of the pain intensity levels in The ShoulderPain dataset [2], b) Feature extraction process

4   Experiments

We conducted experiments on The ShoulderPain dataset [2] containing video recordings of patients suffering from shoulder pain while performing range-of-motion tests of their arms (see Sec. 1 for details). 200 sequences of 25 subjects were recorded (48,398 frames in total). For each frame, discrete pain intensities (0-15) according to Prkachin and Solomon [3] are provided by the database creators (see Fig.1(a)). All the image sequences with the pain intensity > 0 were pre-segmented, so that the number of frames with the intensity 0, the most frequent in the dataset, was balanced with the second most frequent intensity. The resulting intensity distribution was still highly imbalanced, so we discretized it into 6 pain levels as: 0 (none), 1 (mild), 2 (discomforting), 3 (distressing), 4-5 (intense), and 6-15 (excruciating). The ratio of the highest and the lowest pain level was 3:1. This data balancing was performed in order to avoid the tested methods overfitting the majority classes. To evaluate the methods, we selected 147 image sequences from 22 subjects, 10 of which were used as the training set, and the rest as the test set.

To obtain the input features, we first aligned the image frames using a piece-wise affine warp based on the 66 points of the AAM provided by the database creators (see [2,4] for details). The aligned images were then divided into 6x6 even patches to preserve local texture information. From each image patch we extracted Local Binary Patterns (LBP) [11] with radius 2, resulting in 59 histogram bins per patch. This process is outlined in Fig.1(b). We used LBPs as the input features since they have been shown to perform well for the facial affect data (e.g., see [7,4]).

(a) KCORF [7]   (b) KCORF$_h$
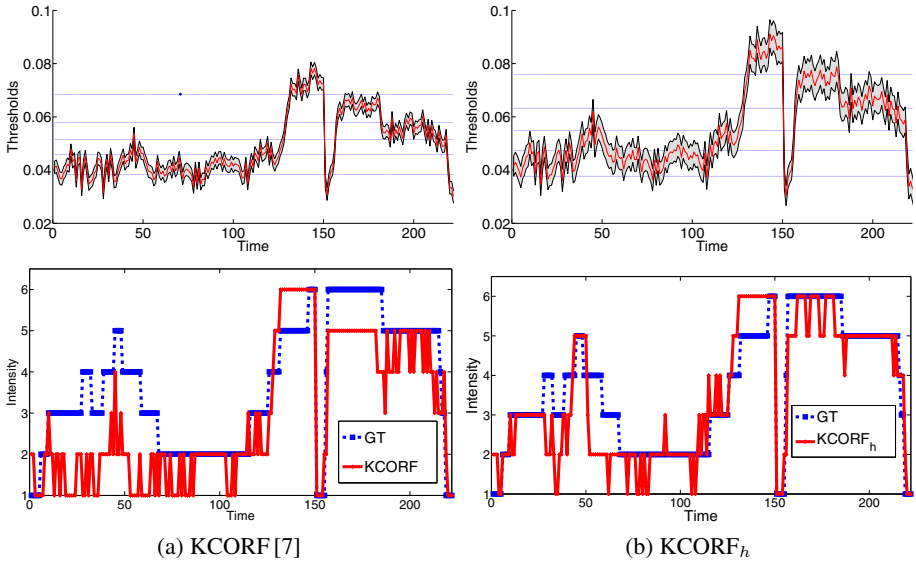
Fig. 2. Comparison of the: (a) homoscedastic and the proposed (b) heteroscedastic KCORF models with the same dynamic features. The upper row shows the values of the latent variable Y^* across time, where the horizontal lines are the learned thresholds. The estimated variance is also shown on Y^*. The $Time$ represents the frame number, where we concatenated two sequences of two test subjects (1-150 / subject 1, 151-222 / subject 2). Note the change in variance in the heteroscedastic model as the subjects change. The bottom row shows the intensity prediction by the two methods.

We compare the proposed heteroscedastic (kernel) CORF (KCORF$_h$) model with its homoscedastic counterpart, KCORF [12], recently proposed for AU temporal segmentation. We used 150 kernel bases for the location and scale models. The bases selection was performed by sampling 25 kernel bases from each pain intensity at random. It was found that this is a good trade-off between the performance and computational complexity of the models. Using the small number of kernel bases also helped to reduce the overfitting. For both the kernel methods, we used the Histogram Intersection (CHI) kernel [13], since it is a non-parametric kernel, and, therefore, it does not involve learning of additional parameters. The balancing trade-off between the regularization and the log-likelihood terms was estimated by grid search under cross validation on the training data.

As a baseline model, we used one-vs-all SVM [14], since most of the prior work on pain intensity estimation is based on this classifier. We also performed comparisons with the state-of-the-art static ordinal regression models, Support Vector Ordinal Regression with implicit constraints (SVOR) [15] and Gaussian Process Ordinal Regression [16]. For the kernel methods, we used the same kernel function as explained above. Finally, we performed the comparison with the base models for sequential data: Gaussian Hidden Markov Models (GHMM)[17] and linear-chain Conditional Random fields (CRFs) [18], since these models are commonly used for modeling sequential data. For the GHMM, each pain intensity level was treated as the model's state parametrized

using a single Gaussian. We also included comparisons with the Laplacian-regularized Conditional Ordinal Random Field (CORF) [12] model, recently proposed for emotion intensity estimation. Because learning in the linear models (GHMM/CRF/CORF) is intractable due to the high dimensionality of the input features, we applied different dimensionality reduction techniques. The reported results are the best obtained, and they were achieved using the 6D features derived with the Kernel Locality preserving projections [19]. The performance of the tested models is reported using: (i) average F-1 measure computed from predictions for each pain intensity, (ii) the mean absolute loss computed between actual and predicted pain intensities, and (iii) Intra-Class Correlation (ICC(3,1) [20]). The ICC is commonly used in behavioral sciences to quantify agreement between different coders, and it is a measure of correlation or conformity of data with multiple targets. The higher the ICC the better.
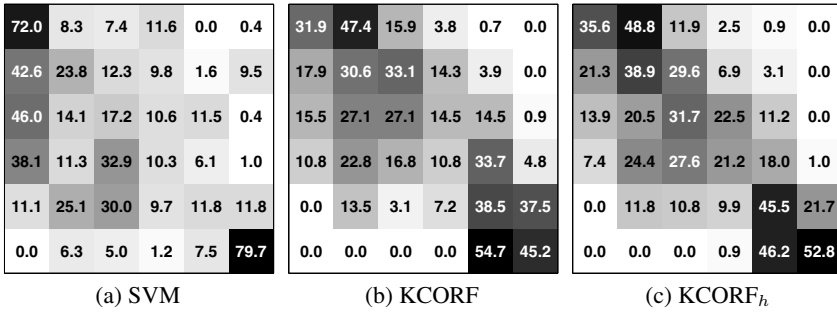
Fig.2 shows the latent variable learned in the homoscedastic KCORF and the proposed heteroscedastic KCORF$_h$ model. Note that the variance in the heteroscedastic model varies across time. This is especially true when switching between the subjects. The change in the variance helps to adjust the locations of the intensity thresholds in the heteroscedastic ordinal model depending on the test subject. Therefore, depending on the pain expressivness level of each subject, the model changes its parameters accordingly. Based on the prediction results shown in Fig.2, it is evident that this helps to improve estimation of the pain intensity levels, especially of the higher levels. For example, around the frame number 50, the heteroscedastic model correctly detects level 5, in contrast to the homoscedastic model. Also, the heteroscedatic model gives smoother predictions compared to the homoscedastic model. Since both models use the same dynamic features, we attribute this to the heteroscedastic component in the proposed model.

Table 1 shows the performance of different classification methods applied to the target task. First, note that all methods attain low the F-1 measure. This is expected because the large variation in facial appearance of different subjects poses a significant challenge for any classifier. We checked the training results of the tested methods and found that all methods attained significantly higher F1 values. This overfitting of the models is ascribed to the fact that subject-specific variation in the used features dominates over the pain-level-specific variation. We next examine how far off are the predictions from the labels. This is reflected in the absolute loss by the tested models. Note that the standard classification methods (SVM/GHMM/CRF) exhibit the highest loss, followed by the static ordinal regression models (SVOR/GPOR). The better results are attained by the dynamic ordinal models, i.e., KCORF and KCORF$_h$, with the latter performing the best. This evidences that both the ordinal and temporal modeling contribute to improving the pain intensity estimation. Furthermore, accounting for heterogeneity in subjects

Table 1. The performance of different methods applied to the task of automatic pain intensity estimation. The features for the linear models (GHMM/CRF/CORF) were pre-processed using KLPP[19].

Methods	SVM	SVOR	GPOR	GHMM	CRF	CORF	KCORF	KCORF$_h$
F-1 [%]	31.1	33.9	34.1	24.8	34.7	35.5	36.8	**40.2**
Abs. Loss	1.25	1.10	1.07	1.30	1.22	0.92	0.88	**0.80**
ICC [%]	46.5	57.1	57.8	39.4	49.0	63.2	66.5	**70.3**

72.0	8.3	7.4	11.6	0.0	0.4	31.9	47.4	15.9	3.8	0.7	0.0	35.6	48.8	11.9	2.5	0.9	0.0
42.6	23.8	12.3	9.8	1.6	9.5	17.9	30.6	33.1	14.3	3.9	0.0	21.3	38.9	29.6	6.9	3.1	0.0
46.0	14.1	17.2	10.6	11.5	0.4	15.5	27.1	27.1	14.5	14.5	0.9	13.9	20.5	31.7	22.5	11.2	0.0
38.1	11.3	32.9	10.3	6.1	1.0	10.8	22.8	16.8	10.8	33.7	4.8	7.4	24.4	27.6	21.2	18.0	1.0
11.1	25.1	30.0	9.7	11.8	11.8	0.0	13.5	3.1	7.2	38.5	37.5	0.0	11.8	10.8	9.9	45.5	21.7
0.0	6.3	5.0	1.2	7.5	79.7	0.0	0.0	0.0	0.0	54.7	45.2	0.0	0.0	0.0	0.9	46.2	52.8

| (a) SVM | (b) KCORF | (c) KCORF$_h$ |

Fig. 3. Confusion matrices obtained using different models. For a baseline, we include the results attained by the SVM-based method.

additionally helps to improve the estimation. The same conclusions can be drawn from the ICC scores for the tested models. However, it is important to mention that the ICC used here is insensitive to bias in the predictions, in contrast to the absolute loss. Nevertheless, the obtained scores reveal that the ordinal models exhibit better conformity between the predictions and the labels, with the proposed model achieving the highest score. To further analyze the performance of the models, we plot in Fig.3 confusion matrices for the SVM, KCORF, and proposed KCORF$_h$. Note that both the ordinal models confuse mostly the neighboring intensity levels, which explains their high ICC scores and low absolute loss. On the other hand, the misclassification by the SVM does not conform to any pattern. We attribute this to the fact that SVM treats the output variables as nominal. From Fig.3(a), it is also evident that the SVM fails to differentiate well between intermediate intensity levels, as opposed to the ordinal models. Finally, compared to the homoscedastic KCORF model, the KCORF$_h$ reduces the misclassification with the classes being further from the diagonal, which, again, evidence the importance of modeling the heterogeneity in subjects.

5   Conclusion

In this paper, we proposed the heteroscedastic CORF model for automatic pain intensity estimation. The proposed model relaxes the homoscedasticity assumption in the CORF model, designed for modeling sequential ordinal data. Our experimental results indicate that, when LBPs are used as the image descriptors, the subjects in the dataset used do exhibit a certain level of heterogeneity. Based on the three performance measures used in our experiments, it is evident that accounting for this heterogeneity results in better pain intensity estimation attained by the proposed model compared to that attained by the homoscedastic ordinal model, and the other classification models.

References

[1] Lucey, P., Cohn, J., Prkachin, K., Solomon, P., Chew, S., Matthews, I.: Image and Vision Computing (42), 197–205

[2] Lucey, P., Cohn, J., Prkachin, K., Solomon, P., Matthews, I.: Painful data: The UNBC-McMaster shoulder pain expression archive database. In: FG, pp. 57–64. IEEE (2011)

[3] Prkachin, K., Solomon, P.: The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. Pain 139, 267–274 (2008)

[4] Kaltwang, S., Rudovic, O., Pantic, M.: Continuous pain intensity estimation from facial expressions. In: Bebis, G., et al. (eds.) ISVC 2012, Part II. LNCS, vol. 7432, pp. 368–377. Springer, Heidelberg (2012)

[5] Hammal, Z., Cohn, J.F.: Automatic detection of pain intensity. In: Proceedings of the 14th ACM International Conference on Multimodal Interaction, ICMI 2012, pp. 47–52. ACM (2012)

[6] Kim, M., Pavlovic, V.: Structured output ordinal regression for dynamic facial emotion intensity prediction. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 649–662. Springer, Heidelberg (2010)

[7] Rudovic, O., Pavlovic, V., Pantic, M.: Kernel conditional ordinal random fields for temporal segmentation of facial action units. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part II. LNCS, vol. 7584, pp. 260–269. Springer, Heidelberg (2012)

[8] Kanamori, T.: Statistical models and learning algorithms for ordinal regression problems. Information Fusion 14, 199–207 (2013)

[9] McCullagh, P.: Regression models for ordinal data. Journal of the Royal Statistical Society. Series B (42), 109–142

[10] Lafferty, J., McCallum, A., Pereira, F.: Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. In: ICML, pp. 282–289 (2001)

[11] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24, 971–987 (2002)

[12] Rudovic, O., Pavlovic, V., Pantic, M.: Multi-output laplacian dynamic ordinal regression for facial expression recognition and intensity estimation. In: CVPR (2012) (in press)

[13] Barla, A., Odone, F., Verri, A.: Histogram intersection kernel for image classification. In: ICIP 2003, vol. 3,2, pp. III-513–III-516 (2003)

[14] Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1– 27:27 (2011)

[15] Chu, W., Keerthi, S.S.: New approaches to support vector ordinal regression. In: ICML, pp. 145–152 (2005)

[16] Chu, W., Ghahramani, Z.: Gaussian processes for ordinal regression. JMLR 6, 1019–1041 (2005)

[17] Murphy, K.P.: The bayes net toolbox for matlab. Computing Science and Statistics 33, 2001 (2001)

[18] Lafferty, J.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data, pp. 282–289. Morgan Kaufmann (2001)

[19] He, X., Niyogi, P.: Locality Preserving Projections. In: NIPS (2004)

[20] Shrout, P., Fleiss, J.: Intraclass correlations: uses in assessing rater reliability. Psychology Bulletin (1979)