Audio-Visual Emotion Analysis Using Semi-Supervised Temporal Clustering with Constraint Propagation

Rodrigo Araujo(✉) and Mohamed S. Kamel

Department of Electrical and Computer Engineering, Centre for Pattern Analysis and Machine Intelligence, University of Waterloo, 200 University Ave W, Waterloo, ONN2L 3G1, Canada
{raraujo,mkamel}@uwaterloo.ca

Abstract. In this paper, we investigate applying semi-supervised clustering to audio-visual emotion analysis, a complex problem that is traditionally solved using supervised methods. We propose an extension to the semi-supervised aligned cluster analysis algorithm (SSACA), a temporal clustering algorithm that incorporates pairwise constraints in the form of *must-link* and *cannot-link*. We incorporate an exhaustive constraint propagation mechanism to further improve the clustering process. To validate the proposed method, we apply it to emotion analysis on a multimodal naturalistic emotion database. Results show substantial improvements compared to the original aligned clustering analysis algorithm (ACA) and to our previously proposed semi-supervised approach.

Keywords: Semi-supervised · Facial expression · Speech emotion recognition · Clustering · Temporal segmentation · Kernel k-means

1 Introduction

Analysis of naturally occurring human emotions has become the main focus of recent research in the field of affective computing. Emotional analysis is considered a vital step towards building efficient and more realistic intelligent human-computer interfaces. The focus is now directed towards recognition in terms of dimensional and continuous description, rather than a small number of discrete emotion categories. Numerical representation of emotions in a multi-dimensional space is considered a more appropriate representation that can reflect the gradated nature of emotions. Moreover, human natural affective behavior is multimodal, subtle, and complex, which makes it challenging to map the affective human state into a single label or discrete number of classes [4].

Facial expressions and speech are the two modalities most commonly used to analyze emotions in human interaction. While facial expressions are considered the major modality in human communication, according to [8], speech is the fastest and most natural method of communication between humans.

A. Campilho and M. Kamel (Eds.): ICIAR 2014, Part II, LNCS 8815, pp. 3–11, 2014.
DOI: 10.1007/978-3-319-11755-3_1

The bulk of the approaches found in the literature are on supervised learning, despite the fact that the labeling process is demanding. With the abundance of data available in this domain and the burdensome nature of the labeling process, it is understandable that unsupervised methods should be pursued more intensively. However, purely unsupervised methods may not produce desirable results due to the complexity of the problem at hand. This need to balance the demands of process and accuracy of results motivated us to pursue a semi-supervised approach.

In this paper, we extend our previously proposed Semi-Supervised Aligned Cluster Analysis (SSACA) method [1] by using an exhaustive constraint propagation approach and apply it to the AVEC audio-visual database.

2 Related Work

Very few works have applied unsupervised methods to emotion analysis in general. This observation is true for both of the two most-used modalities: facial expressions and speech. A possible reason for the shortage of unsupervised work is due to the lack of the temporal aspect of the traditional clustering algorithms. The bulk of the methods are supervised, with HMM being the most used method in terms of audio. In terms of visual features, Relevance Vector Machine (RVM) has shown very good results [7].

Some supervised works that tackle the problem as a dimensional and continuous emotion recognition are [11] [7] [3]. Wollmer et al. [11] have studied the estimation of emotions from speech in the valence and activation dimensions using Long- Short-Term Recurrent Neural Networks. Nicolaou et al. [7] have proposed the use of Output-Associative Relevance Vector Machine (OA-RVM) for dimensional and continuous estimation of emotions from facial expressions. Grimm et al. [3] have compared the performance of Support Vector Regression, Fuzzy k-Nearest Neighbor, and Rule-based Fuzzy Logic classifiers as estimators of spontaneously expressed emotions in speech from three continuous-valued emotion primitives.

In terms of unsupervised methods for emotion analysis, one recent publication is the work of De la Torre *et al.* [2], who have proposed a temporal segmentation method of facial gestures to cluster similar facial actions. Zhou *et al.* [13] have examined facial events directly from naturally occurring videos, using temporal clustering. They use two algorithms for this task: Aligned Cluster Analysis (ACA) and a multi-subject correspondence for matching expressions.

Both of these works, however, analyze only one modality, and they use either categorical labels or action units. In the case of speech and unsupervised methods, most of the research is on speech segmentation for speech recognition or on speech separation.

Recently, we proposed a semi-supervised method, SSACA [1], which uses pairwise constraints in the form of *must-links* and *cannot-links* as a way to add side information to help the clustering process, and to boost its performance with minimal supervised information. We applied this method to a naturalistic database, and the results showed improvements compared to the original approach. In this paper, we build on our previous approach, adding an exhaustive constraint propagation artifact to the framework, and we apply the proposed method to a larger multimodal naturalistic database, using audio-visual features.

3 Model Description

The emotional behavior of a person can be treated as a time series, wherein the specific emotion primitive being evaluated varies over time. The goal is to factorize (segment) multiple time series into disjointed segments that belong to k temporal clusters. Essentially, we have a temporal clustering problem. The idea is to have frames within a segment that are similar to each other and non-overlapping segments that belong to k temporal clusters.

3.1 Semi-Supervised Aligned Cluster Analysis (SSACA)

This section describes SSACA [1], a transformation of the temporal clustering algorithm ACA into a semi-supervised temporal clustering method. In contrast to ACA, SSACA adds some side information to its framework in the form of pairwise constraints, improving the accuracy and performance of the temporal clustering. ACA is a combination of kernel k-means and Dynamic Time Alignment Kernel (DTAK).

The goal of ACA is to decompose a segment $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ into m disjoint segments, where each segment belongs to a single cluster. Each segment is constrained by a maximum length n_{max}, which also serves as a way to control the temporal granularity of the segmentation. The segments begin at position s_i and end at $s_{i+1} - 1$, such that $n_i = s_{i+1} - s_i \leq n_{max}$. An indicator matrix $\mathbf{G} \in \{0, 1\}^{k \times m}$ assigns each segment to a cluster; $g_{ci} = 1$ if $\mathbf{Z}_i$ belongs to cluster c.

ACA combines kernel k-means with DTAK to achieve temporal clustering by minimizing:

$$
\begin{aligned}
J_{aca}(\mathbf{G}, \mathbf{s}) = \sum_{c=1}^{k} \sum_{i=1}^{m} g_{ci} \underbrace{\left\| \psi(\mathbf{X_{[s_i, s_{i+1})}}) - \mathbf{z_c} \right\|^2}_{dist^2_{\psi}(\mathbf{Y}_i,\, \mathbf{z}_c)} = \ \| [\psi(\mathbf{Y}_1, \ldots, \psi(\mathbf{Y}_m) - \mathbf{ZG} \|^2_F \,, \\
s.t. \mathbf{G}^T 1_k = 1_m \text{ and } s_{i+1} - s_i \in [1, n_{max}],
\end{aligned}
\tag{1}
$$

where $\mathbf{G} \in \{0, 1\}^{k \times m}$ is a cluster indicator matrix, and $s \in \mathbb{R}^{m+1}$ is the segment vector $\mathbf{Y} = \mathbf{X}_{[s_i, s_{i+1})}$, which is one of the differences between ACA and kernel k-means. In the case of ACA, the $dist^2_{\psi}(\mathbf{Y}_i,\, \mathbf{z}_c)$ is the squared distance between the ith segment and the center of cluster c in the nonlinear mapped feature space represented by $\psi(.)$.

In order to add the semi-supervised component to the proposed method, we rely on the discovery of Kulis *et al.* [5], which has shown that the objective function for semi-supervised clustering, based on Hidden Markov Random

Fields (HMRF), with squared Euclidean distance and a certain class of constraint penalty function, can be expressed as a special case of the weighted kernel k-means. SSACA, which is based on kernel k-means, may use the same framework of HMRF semi-supervised clustering. Thus, we can write the SSACA objective function as:

$$J_{ssaca}(\mathbf{G}, \mathbf{s}) = \sum_{c=1}^{k} \sum_{i=1}^{m} g_{ci} \underbrace{\left\| \psi(\mathbf{X}_{[\mathbf{s_i}, \mathbf{s_{i+1}})}) - \mathbf{z_c} \right\|^2}_{dist_\psi^2(\mathbf{Y}_i,\, \mathbf{z}_c)} - \sum_{\substack{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{M} \\ g_i = g_j}} w_{ij} + \sum_{\substack{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{C} \\ g_i = g_j}} w_{ij} \quad (2)$$

where $\mathcal{M}$ is the set of *must-link* constraints, $\mathcal{C}$ is the set of *cannot-link* constraints, w_{ij} is the penalty cost for violating a constraint $\mathbf{x}_i$ and $\mathbf{x}_j$, and g_i refers to the cluster label of $\mathbf{x}_i$. There are three terms in this objective function. The first term is the unsupervised k-means term of the objective function. Note that the distance $dist_\psi^2(\mathbf{Y}_i,\, \mathbf{z}_c)$ can be represented as a matrix of pairwise squared Euclidean distances among the data points (see proof in [5]). We refer later to this distance matrix as $\mathbf{S}$. The second term is based on the *must-link* constraints, and states that for every *must-link* $\mathbf{x}_i$ and $\mathbf{x}_j$ that are in the same cluster, the objective function is rewarded by subtracting some pre-specified weight. Similarly, the third term in the objective function states that for every *cannot-link* $\mathbf{x}_i$, $\mathbf{x}_j$ in the same cluster has violated that constraint, so the objective function is penalized by some pre-specified penalty weight. We will refer later to the second and third term of the function as $\mathbf{W}$.

Kulis *et al.* [5] have also shown that, for the equivalence of the HMRF k-means and the weighted kernel k-means to hold, it is necessary to construct a certain kernel matrix and set weights in a specific way. A kernel matrix $\mathbf{K}$ should have two components: $\mathbf{K} = \mathbf{S}+\mathbf{W}$. $\mathbf{S}$ is the similarity matrix, and comes from the unsupervised term, while $\mathbf{W}$ is the constraint matrix. This matrix $\mathbf{W}$ has a pre-specified w_{ij} weight for *must-link* and $-w_{ij}$ for *cannot-link*, and zero otherwise. Thus, this objective function is mathematically equivalent to the weighted kernel k-means objective function. In other words, we can run weighted kernel k-means to decrease the objective function.

Because the constraints are held in the segment level, we have two kernel matrices, $\mathbf{K}$ and $\mathbf{T}$. $\mathbf{K}$ is the frame kernel matrix, which defines the similarity between two frames, $\mathbf{x_i}$ and $\mathbf{x_j}$. $\mathbf{T} = [\tau_{ij}]_{m \times m} \in \mathbb{R}^{m \times m}$ is the segment kernel matrix that represents the similarity of the segments $\mathbf{X}_{[\mathbf{s_i}, \mathbf{s_{i+1}}]}$ and $\mathbf{X}_{[\mathbf{s_j}, \mathbf{s_{j+1}}]}$ using the distance DTAK. The segment kernel matrix $\mathbf{T}$ is be constructed as the sum of $\mathbf{T} + \mathbf{W}$. To avoid excessive notation we will also use $\mathbf{T}$ to designate the result of $\mathbf{T} + \mathbf{W}$.

In ACA, the method adopted to solve this optimization problem is a dynamic programming (DP)-based algorithm, which has a complexity of $O(n^2 n_{max})$ to exhaustively examine all possible segmentations. In SSACA, we adapt the DP-based search to a semi-supervised framework, incorporating the pairwise constraints into the algorithm. We call the new algorithm SS DPSearch (See Algorithm 1). SS DPSearch optimizes SSACA w.r.t $\mathbf{G}$ and s, as well as rewarding or penalizing the distance between segments $\tau(X_{[i,v]}, \dot{Y}_j)$ according to constraints.

Algorithm 1. SS DPSearch

parameter: n_{max}, k, n_{ml}, n_{cl}
input: $\mathrm{G} \in \{0,1\}^{k\times \dot{m}}$, $\dot{\mathrm{s}} \in \mathbb{R}^{(\dot{m}+1)}$, $\mathrm{K} \in \mathbb{R}^{n\times n}$, $\mathbf{T} \in \mathbb{R}^{m\times m}$, $\mathcal{M} \in \mathbb{Z}^{n_{ml}\times 2}$, $\mathcal{C} \in \mathbb{Z}^{n_{cl}\times 2}$
output: $\mathrm{G} \in \{0,1\}^{k\times m}$, $\dot{\mathrm{s}} \in \mathbb{R}^{(m+1)}$

headTail = getHeadTails($\mathcal{M}$, $\mathcal{C}$);
for $v = 1$ **to** n **do**
  $J(v) \leftarrow \infty$;
  if $v \geq$ headTail(:,1) **and** $v <$ headTail(:,2) **then**
    continue;
  end if
  if isTail(v) **then**
    for $j = 1$ **to** $\dot{m}$ **do**
      Retrive directly from $\mathbf{T}(X_{[i,v]}, \dot{Y}_j)$;
    end for
    $c^* \leftarrow \arg\min_c \, dist_\psi(X_{[i,v]}, \dot{z}_c)$;
    $J \leftarrow dist_\psi(X_{[i,v]}, \dot{z}_{c^*})$;
    $J([i,v]) \leftarrow J$, $g^*_{[i,v]} \leftarrow e^*_c$, $i^*_{[i,v]} \leftarrow i$;
  else
    for $n_v = 1$ **to** $min(n_{max}, v)$ **do**
      { *Same as DPSearch*}
    end for
  end if
end for
{*Perform backward segmentation*}

3.2 Exhaustive and Efficient Constraint Propagation

The pairwise constraints are used to adjust the similarity matrix for the kernel k-means clustering algorithm. However, using this technique, only the constrained segment similarities are affected. In order to make the propagation of constraints more efficient, we borrow the idea of exhaustive and efficient constraint propagation from Lu and Ip [6] and adapt it to our framework. The rationale behind this method is to spread the effects of the constraints throughout the whole similarity matrix S.

Exhaustive and Efficient Constraint Propagation (E^2CP) tackles the problem of constraint propagation by decomposing it into sets of label propagation subproblems. Given the dataset $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$, a set of *must-link* $\mathcal{M}$ and a set of *cannot-link* $\mathcal{C}$, we can represent all the pairwise constraints in a single matrix $W = {Z_{ij}}_{N\times N}$:

$$W_{ij} = \begin{cases} +1, & (x_i, x_j) \in \mathcal{M} \\ -1, & (x_i, x_j) \in \mathcal{C} \\ \;\;0, & \text{otherwise} \end{cases} \tag{3}$$

Each j-th column of $W_{.j}$ can now be seen as a two-class semi-supervised learning problem, in which the *positive class* ($W_{ij} > 0$) represents the segments that should be on the same cluster, and the *negative class* ($W_{ij} < 0$) represents the segments that should not be in the same cluster. If $(W_{ij}) = 0$, x_i x_j are not constrained. Then, each column is solved by label propagation in parallel [12]. The same process is repeated for the rows, ensuring that all the segments will be affected by the propagation. The algorithm can be described as follows:

1. Create the similarity matrix T or a symmetric k-NN graph.
2. Create the matrix $\bar{\mathcal{L}} = D^{\frac{-1}{2}} T D^{\frac{-1}{2}}$, where D is a diagonal matrix with its (i,i)-element equal to the sum of the i-th row of T.
3. Iterate $F_v(t+1) = \alpha \bar{\mathcal{L}} F_v(t) + (1-\alpha) W$ for vertical constraint propagation until convergence, where $F_v(t) \in \mathcal{F}$ and α is a parameter in the range of $(0,1)$.
4. Iterate $F_h(t+1) = \alpha F_h(t) \bar{\mathcal{L}} + (1-\alpha) F_v^*$ for horizontal constraint propagation until convergence, where $F_h(t) \in \mathcal{F}$ and F_v^* is the limit of $\{F_{v(t)}\}$.
5. Output $F^* = F_h^*$ as the final representation of the pairwise constraints, where F_h^* is the limit of $\{F_{h(t)}\}$.

Intuitively, the algorithm receives information from its neighbor at each iteration, and the parameter α controls the relative amount of information passed from the neighbors. The final label of segments is set to be the cluster from which it has received the most information during the iteration process.

Without loss of generality, [12] shows that $\{F(t)\}$ can be calculated in a closed form. The output F^* represents an exhaustive set of pairwise constraints with the associated confidence scores $|F^*|$. Now, we can adjust the similarities in T with the output scores of F^*, as described in Equation 4.

$$\tilde{T_{ij}} = \begin{cases} 1 - (1 - F^*_{ij})(1 - W_{ij}), & F^*_{ij} \geq 0 \\ (1 + F^*_{ij}) W_{ij}, & F^*_{ij} < 0 \end{cases} \tag{4}$$

Algorithm 2 shows SSACA with the exhaustive propagation.

Algorithm 2. SSACA + Exhaustive propagation (EP)

input: $\mathbf{S} \in \mathbb{R}^{n\times n}$: input frame kernel matrix, $\mathbf{T} \in \mathbb{R}^{n\times n}$: input segment kernel matrix, $\mathbf{W} \in \mathbb{R}^{n\times n}$: constraint penalty, k: number of clusters, $\mathcal{M}$: set of must-link constraints, $\mathcal{C}$: set of cannot-link constraints, $\dot{s}$: initial segmentation.
output: G $\in \{0,1\}^{k\times n}$: Final partitioning of the points
1: Propagate the constraints $F^* \leftarrow W$
2: Form the matrix $\tilde{T}$ according to equation 4.
3: Diagonal-shift $\tilde{T}$ by adding σI to guarantee positive definiteness of $\tilde{T}$.
4: Get initial clusters $G^{(0)}$ using constraints.
5: Return s = SSDPSearch($G^{(0)}$, $\dot{s}$, S, $\tilde{T}$, $\mathcal{M}$, $\mathcal{C}$, k).

4 Experiments

We performed experiments on a naturalistic emotion speech database, and compared the performance of the exhaustive propagation SSACA (SSACA+EP) with SSACA and ACA. The accuracy evaluation criterion is the same used in [1], and is based on the Hungarian algorithm.

4.1 AVEC Database

AVEC[10] is an audio-visual emotion recognition database created for the emotion recognition challenge (AVEC 2012). It consists of conversations between participants and four stereotyped characters. Each character has a specific emotion stereotype: sensible, happy, angry, and sad. The train partition of the database contains 31 sections, wherein each session contains one dialogue with a specific character. The database is labeled for arousal, valence, power, and expectancy.

We used the Word-Level Sub-Challenge (WLSC) portion of the database. Because we used a temporal-clustering-based approach, we categorized the continuous values of the affective dimensions, which range from [-1, +1], in 6 categories: [-1,-0.66], [-0.66, -0.33], [-0.33, 0], [0, 0.33], [0.33, 0.66], and [0.66, 1].

The audio features used consist of 1871 features, including 25 energy and spectral related low-level descriptors (LLD) x 42 functionals, 6 voicing related LLD x 32 functionals, and 10 voiced/unvoiced durational features. Details for LLD and functionals can be found in [10]. For visual features, we extracted Local Binary Patterns (LBP), based on the approach described in [9]. For arousal, we used audio features, since it has been shown consistently in other works [9] that audio features are more suitable for this type of affect dimension. For valence, power, and expectancy, we used visual features.

Figure 1 shows the average results of 20 random initializations on the train portion of the AVEC database for three different methods on arousal, valence, power, and expectancy, respectively. Note that both SSACA and SSACA+EP had superior performance compared to the baseline algorithm, ACA, in almost all of the sessions, with the addition of only 5% percent of the possible constraints as side information. SSACA+EP and SSACA showed very similar results; however, for the sessions with higher number of segments and high variability, SSACA+EP showed significantly better results, (*e.g.*, session 22, Figure 1(a)).

Emotion variability seems to play a big role in influencing the results of SSACA+EP. We define variability in this context as the variation of emotions. When there are a lot of transitions from one category to another, we say we have a high variability; when there are few transitions, we say we have low variability. In order to observe this aspect on the AVEC database, we set up another experiment. In this experiment, we used the dialogue of a participant with 4 different characters and combined them, so we had a longer conversation with the possibility of a high variability. Table 1 shows the average result of 20 random initializations.

Note that in this setup, SSACA+EP improved the results in all emotion dimensions. For arousal, it improved from 0.74 to 0.78 with a lower standard

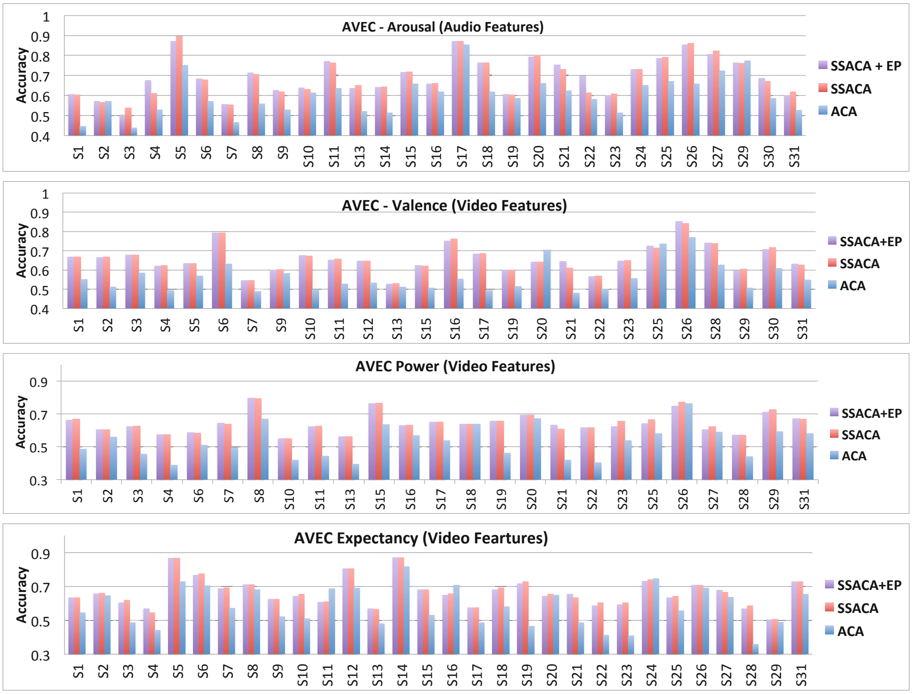

Fig. 1. Average accuracy of SSACA+EP, SSACA and ACA on the AVEC dataset for all sessions, at approximately 5% of the total number of possible constraints

deviation, using the same amount of constraints. Similar improvement was observed for valence, which improved from 0.71 to 0.75. For power, we observed improvements from 0.75 to 0.77. Finally, expectancy improved from 0.81 to 0.84. In terms of the baseline algorithm, the proposed method had a very significant increase in performance, in some cases doubling the accuracy with the addition of only 5% of the possible number of constraints.

Table 1. Average accuracy results on high variance segments

	Average Accuracy			
	Arousal	Valence	Power	Expectancy
SSACA+EP	0.78 ± 0.07	0.75 ± 0.08	0.77 ± 0.11	0.84 ± 0.08
SSACA	0.74 ± 0.09	0.71 ± 0.08	0.75 ± 0.12	0.81 ± 0.07
ACA	0.46 ± 0.02	0.51 ± 0.05	0.40 ± 0.12	0.40 ± 0.02

5 Conclusion

In this work, we propose SSACA+EP, a temporal clustering algorithm that extends SSACA. SSACA+EP incorporates a mechanism for constraint propagation into its framework, spreading the *must-link* and *cannot-link* constraints throughout the similarity matrix and making the process more efficient. Results on an audio-visual naturalistic emotion conversation database show improvement in all four dimensional emotions. One of the drawbacks of our approach is its complexity, which is quadratic in the number of frames. In future work, we plan on extending this approach to other temporal clustering methods and applying these methods to other temporal clustering problems.

References

1. Araujo, R., Kamel, M.: A semi-supervised temporal clustering method for facial emotion analysis. In: 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) (to appear, July 2014)
2. De La Torre, F., Campoy, J., Ambadar, Z., Conn, J.F.: Temporal segmentation of facial behavior. In: International Conference on Computer Vision, pp. 1–8 (2007)
3. Grimm, M., Kroschel, K.: Emotion estimation in speech using a 3d emotion space concept. In: Grimm, M., Kroschel, K. (eds.) Robust Speech Recognition and Understanding, pp. 281–300. I-Tech Education and Publishing, Vienna (2007)
4. Gunes, H., Pantic, M.: Automatic, dimensional and continuous emotion recognition. Int. J. Synth. Emot. **1**(1), 68–99 (2010)
5. Kulis, B., Basu, S., Dhillon, I.S., Mooney, R.J.: Semi-supervised graph clustering: a kernel approach. Machine Learning **74**(1), 1–22 (2009)
6. Lu, Z., Ip, H.H.S.: Constrained spectral clustering via exhaustive and efficient constraint propagation. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 1–14. Springer, Heidelberg (2010)
7. Nicolaou, M.A., Gunes, H., Pantic, M.: Output-associative rvm regression for dimensional and continuous emotion prediction. Image Vision Comput. **30**(3), 186–196 (2012)
8. Pantic, M., Member, S., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**, 1424–1445 (2000)
9. Sayedelahl, A., Araujo, R., Kamel, M.: Audio-visual feature-decision level fusion for spontaneous emotion estimation in speech conversations. In: 2013 IEEE International Conference on Multimedia and Expo Workshops, pp. 1–6 (July 2013)
10. Schuller, B., Valstar, M., Cowie, R., Pantic, M.: Avec 2012: The continuous audio/visual emotion challenge - an introduction. In: Proceedings of the 14th ACM International Conference on Multimodal Interaction, ICMI 2012, pp. 361–362. ACM, New York (2012)
11. Wöllmer, M., Eyben, F., Reiter, S., Schuller, B., Cox, C., Douglas-cowie, E., Cowie, R.: Abandoning emotion classes - towards continuous emotion recognition with modelling of long-range dependencies. In: Proceedings Interspeech (2008)
12. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: Advances in Neural Information Processing Systems 16, pp. 321–328. MIT Press (2004)
13. Zhou, F., De la Torre, F., Cohn, J.F.: Unsupervised discovery of facial events. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2010)