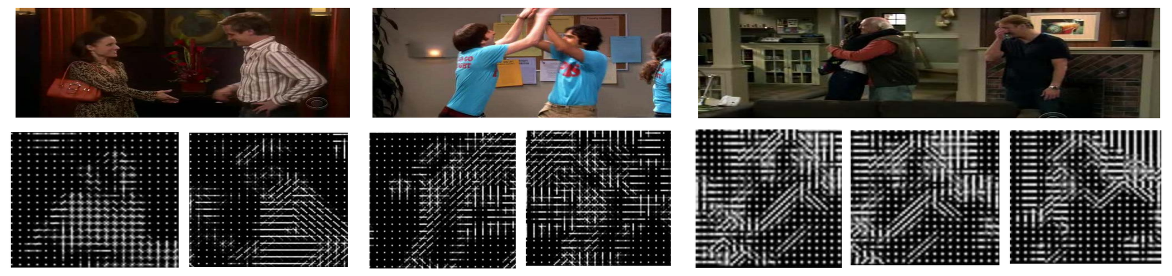
# Bilinear Programming for Human Activity Recognition with Unknown MRF Graphs

Zhenhua Wang, Qinfeng Shi, Chunhua Shen, Anton van den Hengel

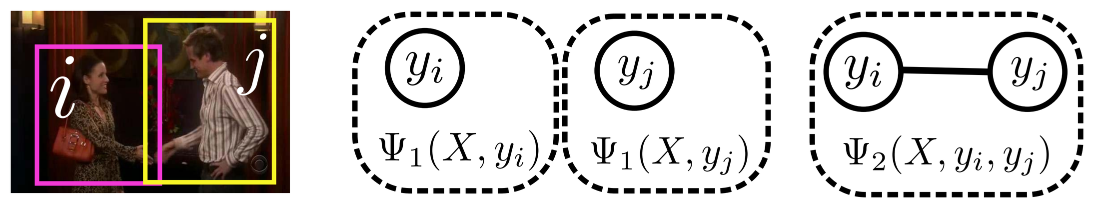
THE UNIVERSITY of ADELAIDE

## Overview

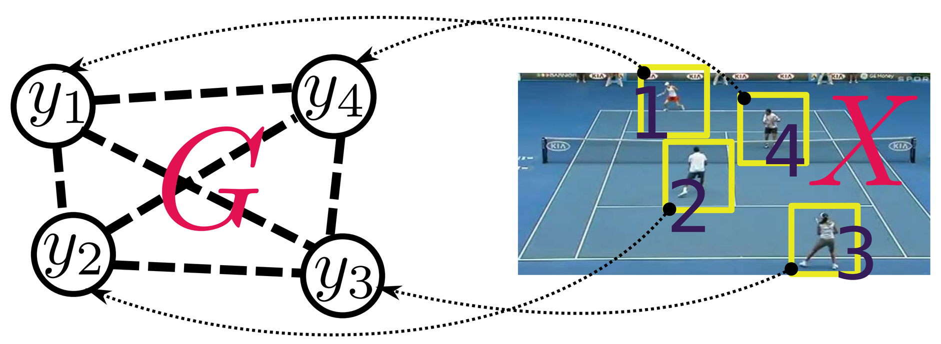
**Task:** Given an image $X$, we want to predict an action label $y$ for each person appeared in the image.



**Local feature** such as HoG (the second row), HoF are not robust and exploiting *interaction context* is necessary.




**Action context:** $\Psi_1(X, y_i)$ is the local feature for a single person $i$. $\Psi_2(X, y_i, y_j)$ is the action context feature, which encodes the joint action configuration $(y_i, y_j)$ given the observation $X$.




**The inference problem:** given $X$ with $m$ persons. Let $Y = \{y_1, y_2, \cdots, y_m\} \in \mathcal{Y}$ be action labels. An MRF $G = (\mathcal{V}, \mathcal{E})$ is used to model the action dependencies, where $\mathcal{V} = \{1, 2, \cdots, m\}$ and $\mathcal{E}$ is *yet to be determined*. The inference problem is formulated as

$$(Y^*, G^*) = \operatorname*{argmax}_{G,Y} \sum_{i\in\mathcal{V}} \underbrace{\mathbf{w}_1^\top \Psi_1(X, y_i)}_{-E_i(y_i)} + \sum_{(i,j)\in\mathcal{E}} \underbrace{\mathbf{w}_2^\top \Psi_2(X, y_i, y_j)}_{-E_{i,j}(y_i,y_j)}, \quad (1)$$

where $\mathbf{w} = [\mathbf{w}_1; \mathbf{w}_2]$ is the model parameter vector, which is to be learned during training.

**Contributions**

- We propose a bilinear programming (BLP) formulation for solving the MRF with unknown graphs, *i.e.* the problem (1).
- A global solver (branch and bound based) for solving the BLP problem is invented.
- The proposed method improves the HAR rate greatly compared to the state-of-the-art.

## BLP relaxation

Introducing $q_i(y_i) \in [0,1], \forall i \in \mathcal{V}$, $q_{i,j}(y_i, y_j) \in [0,1], \forall i, j \in \mathcal{V}$, $z_{i,j} \in [0,1], \forall i, j \in \mathcal{V}$, problem (1) is relaxed into a **BLP problem**

$$\min \; f(q,z) = \sum_{i,j\in\mathcal{V}} \sum_{y_i,y_j} q_{i,j}(y_i,y_j) E_{i,j}(y_i,y_j) z_{i,j} + \sum_{i\in\mathcal{V}} \sum_{y_i} q_i(y_i) E_i(y_i). \quad (2a)$$

$$\text{s.t. } q_{i,j}(y_i, y_j) \in [0,1], \sum_{y_i,y_j} q_{i,j}(y_i,y_j) = 1, \quad (2b)$$

$$\sum_{y_i} q_{i,j}(y_i,y_j) = q_j(y_j), z_{i,j} = z_{j,i}, z_{i,j} \in [0,1],$$

$$\sum_{i\in\mathcal{V}} z_{i,j} \le d, \forall i, j \in \mathcal{V}, y_i, y_j.$$

Solving (2) returns $\{q_i(y_i)\}$, $\{q_{i,j}(y_i, y_j)\}$ and $\{z_{i,j}\}$, from which $G^*$ and $Y^*$ (*i.e.* best activities) can be computed.

**Obtaining the graph:** we start with $\mathcal{E}^* = \emptyset$. $\forall i, j \in \mathcal{V}, i \neq j$, if $z_{i,j} \ge 0.5$, $\mathcal{E}^* = \mathcal{E}^* \cup \{(i,j)\}$. Thus we have the estimated graph $G^* = (\mathcal{V}, \mathcal{E}^*)$.

**Obtaining the MAP:** assume $y_i \in \{1, 2, \cdots, K\}$, then $\forall i \in \mathcal{V}$, $y_i^* = \operatorname{argmax}_{k=1}^K q_i(k)$. We have the estimated label $Y^* = (y_1^*, y_2^*, \cdots, y_m^*)$.

## LP relaxation

A **LP relaxation** of (2) is

$$\min_{q,z,u,\gamma} \sum_{i,j\in\mathcal{V}} \sum_{y_i,y_j} u_{i,j}(y_i,y_j) + \sum_{k\in\mathcal{V}} \sum_{y_k} q_k(y_k) E_k(y_k), \quad (3a)$$

$$\text{s.t.} \begin{cases} q_{i,j}(y_i,y_j) \in [0,1] & \forall i,j \in \mathcal{V}, y_i, y_j, \\ z_{i,j} \in [0,1] & \forall i,j \in \mathcal{V}, \\ \sum_{y_i,y_j} q_{i,j}(y_i,y_j) = 1 & \forall i,j \in \mathcal{V}, \\ \sum_{y_i} q_{i,j}(y_i,y_j) = q_j(y_j) & \forall i,j \in \mathcal{V}, y_j, \\ z_{i,j} = z_{j,i} & \forall i,j \in \mathcal{V}, \\ \sum_{j\in\mathcal{V}} z_{i,j} \le d & \forall i \in \mathcal{V}, \\ u_{i,j}(y_i,y_j) \ge & \\ E_{i,j}(y_i,y_j)\gamma_{i,j}(y_i,y_j) & \forall i,j \in \mathcal{V}, y_i, y_j, \\ \gamma^l \le \gamma_{i,j}(y_i,y_j) \le \gamma^u & \forall i,j \in \mathcal{V}, y_i, y_j, \end{cases} \quad (3b)$$

where $\gamma^l$ is

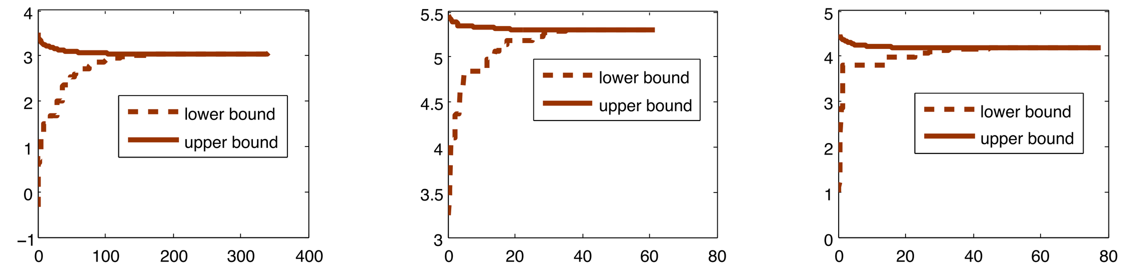
$$\max\{q_{i,j}^l(y_i,y_j) z_{i,j} + z_{i,j}^l q_{i,j}(y_i,y_j) - q_{i,j}^l(y_i,y_j) z_{i,j}^l,$$
$$q_{i,j}^u(y_i,y_j) z_{i,j} + z_{i,j}^u q_{i,j}(y_i,y_j) - q_{i,j}^u(y_i,y_j) z_{i,j}^u\},$$

and $\gamma^u$ is

$$\min\{q_{i,j}^u(y_i,y_j) z_{i,j} + z_{i,j}^l q_{i,j}(y_i,y_j) - q_{i,j}^u(y_i,y_j) z_{i,j}^l,$$
$$q_{i,j}^l(y_i,y_j) z_{i,j} + z_{i,j}^u q_{i,j}(y_i,y_j) - q_{i,j}^l(y_i,y_j) z_{i,j}^u\}.$$

## Branch and bound solver

To solve (2), we employ a B & B method.
**Branch**: split $\Omega \in \Omega_{\text{init}}$ along $z$ variables only.
**Bounds**: $\Phi_{\text{lb}}(\Omega)$ is the solution of (3). Let $(q^*, z^*)$ be the optimisers. $\Phi_{\text{ub}}(\Omega) = f(q^*, z^*)$.




## Feature representation

The two features in (1) are:

$$\Psi_1(X, y_i) = \mathbf{s}_i \otimes \mathbf{e}_1(y_i),$$
$$\Psi_2(X, y_i, y_j) = \mathbf{t}_i \otimes \mathbf{t}_j \otimes \mathbf{e}_1(y_i) \otimes \mathbf{e}_1(y_j) \otimes \mathbf{e}_2(r_{i,j}).$$

$\mathbf{s}_i$ is the local move confidence score, $t_i$ is the local pose confidence score, $r_{i,j}$ is the 2D body relative position, $e_1, e_2$ are indicator vectors, $\otimes$ denotes the Kronecker tensor product.

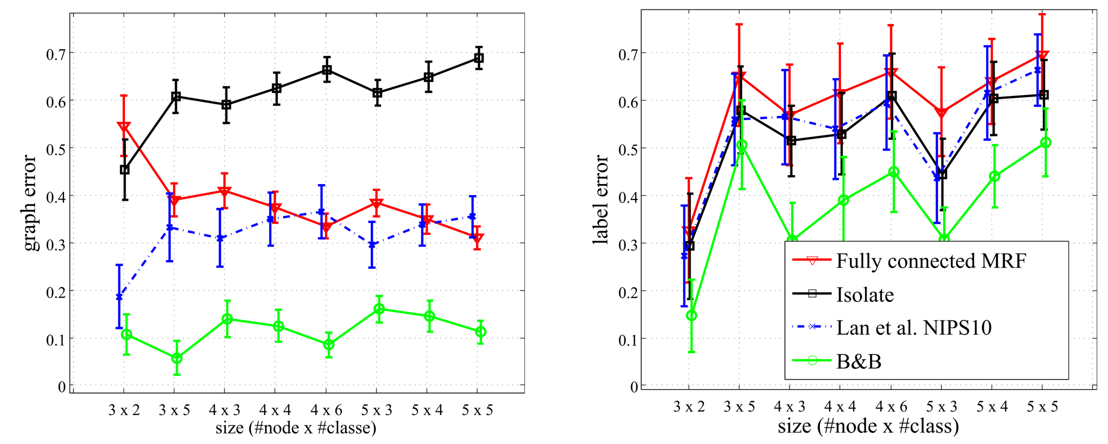
## Latent structured SVM training

We train our model via the latent structured SVM:

$$\min_{\mathbf{w}} \; \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{\ell} \left[ \max_{Y,G'} \left[ \mathbf{w}^\top \Psi(X^i, Y, G') + \Delta(Y^i, Y) \right] - \max_G \mathbf{w}^\top \Psi(X^i, Y^i, G) \right]_+, \quad (4)$$

which is solved via Convex-Concave Procedure (CCCP). Here $\Delta(Y^i, Y) = \frac{1}{m}\sum_{j=1}^m \delta(y_j^i \neq y_j)$. The two inference problems: 1) $\max_G \mathbf{w}^\top \Psi(X^i, Y^i, G)$, 2) $\max_{Y,G} \mathbf{w}^\top \Psi(X^k, Y, G) + \Delta(Y^k, Y)$ are solved using our B & B solver.

## Experiments



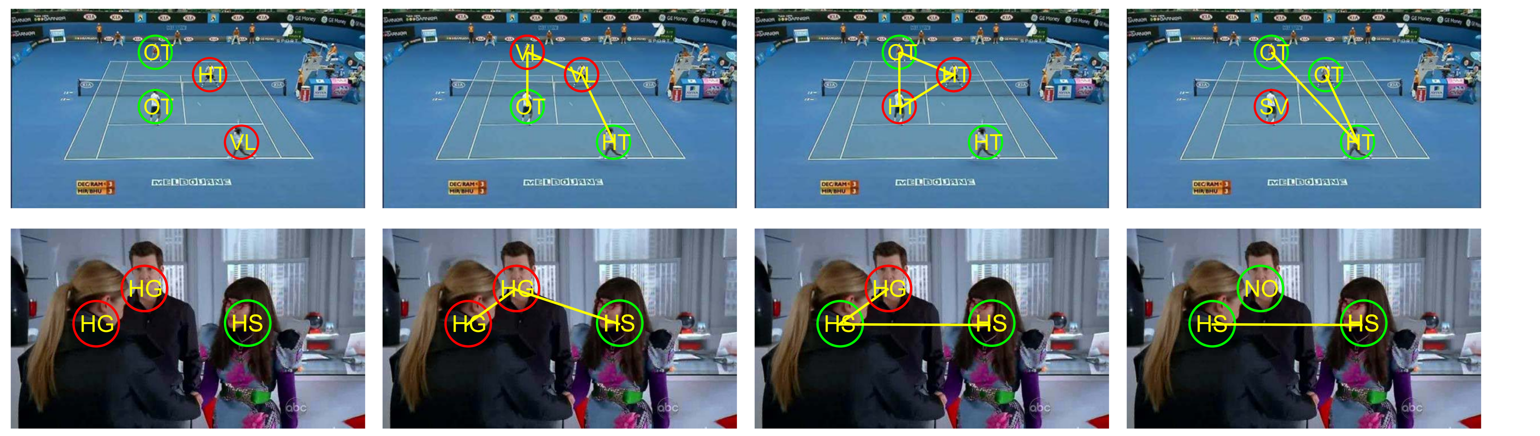

A comparison of the graph error (left) and the label error (right) by different methods on the synthetic data. The graph and label errors are plotted with error bars indicating the standard errors. Note the two subfigures share the same legend in the right diagram. Our method performs much better.

**Confusion matrices of the TVHI dataset.**

| Alg. | MCSVM | | | | | SSVM | | | | | Lan's | | | | | BLP | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A/A | **NO** | **HS** | **HF** | **HG** | **KS** | **NO** | **HS** | **HF** | **HG** | **KS** | **NO** | **HS** | **HF** | **HG** | **KS** | **NO** | **HS** | **HF** | **HG** | **KS** |
| **NO** | .37 | .07 | .21 | .11 | .24 | .20 | .40 | .27 | .06 | .06 | .11 | .36 | .19 | .20 | .13 | **.49** | .20 | .13 | .13 | .05 |
| **HS** | .01 | .55 | .06 | .17 | .21 | .10 | .51 | .21 | .11 | .06 | .09 | .52 | .14 | .15 | .10 | .18 | **.56** | .09 | .08 | .09 |
| **HF** | .09 | .03 | .52 | .21 | .14 | .08 | .11 | .61 | .08 | .12 | .02 | .14 | .58 | .18 | .08 | .11 | .09 | **.63** | .07 | .10 |
| **HG** | .02 | .14 | .20 | .49 | .15 | .05 | .15 | .11 | .58 | .11 | .03 | .06 | .11 | .55 | .26 | .03 | .10 | .10 | **.70** | .06 |
| **KS** | .07 | .11 | .09 | .05 | .67 | .02 | .26 | .14 | .12 | .46 | .01 | .07 | .15 | .11 | .67 | .06 | .08 | .04 | .13 | **.69** |

Prediction results on tennis data (1st row) and TVHI data (2nd rows). Both nodes and edges of the MRFs are shown. For the i.i.d. method, there are no edges. The green, red nodes denotes correct and incorrect predictions resp..

MCSVM SSVM Lan's BLP



## Conclusion

A MAP inference method for unknown graphs has been proposed. We formulated the problem as a bilinear program, which was solved by branch and bound. An LP relaxation was used as a lower bound for the bilinear program. The discriminant model was trained via a latent structured SVM technique.