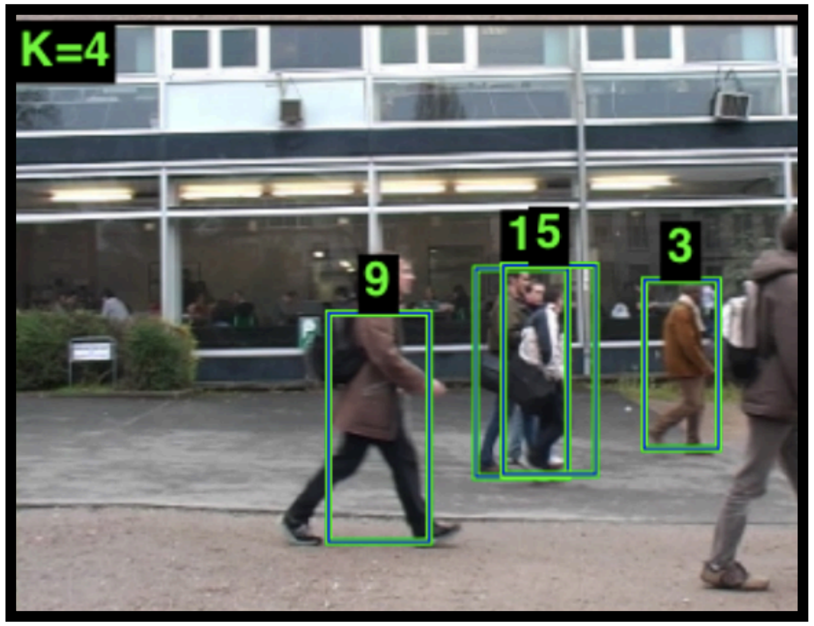
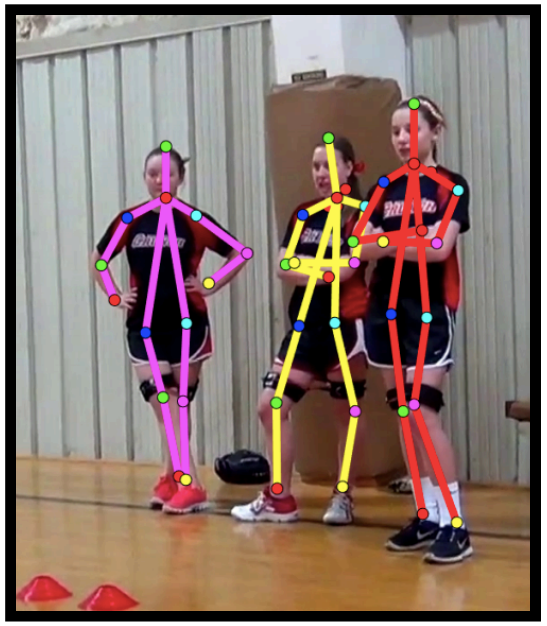
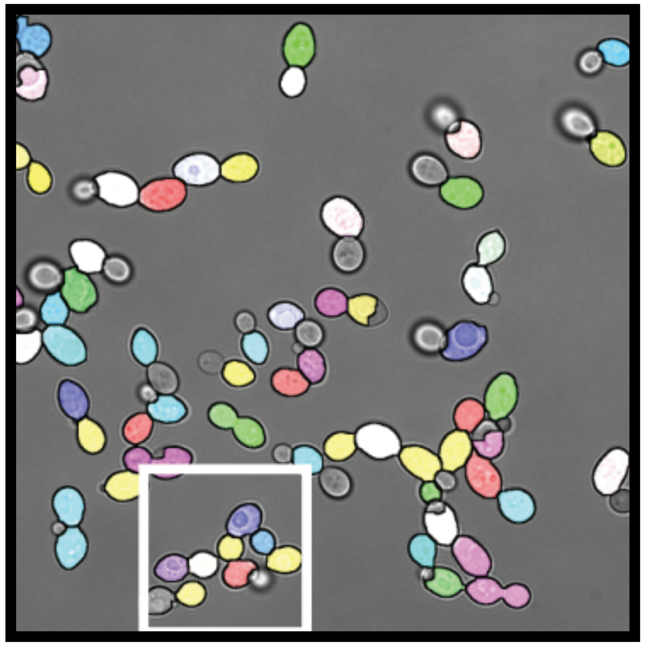
Column Generation for AI: Importing Methods in Operations Research for Problems in AI: Entity Resolution

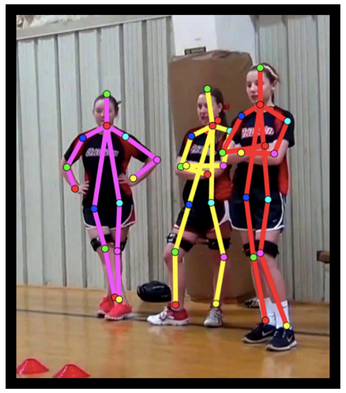




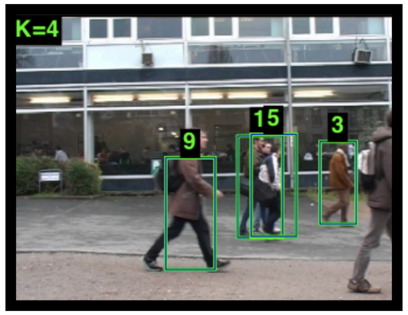


Minimum Weight Set Packing (MWSP)

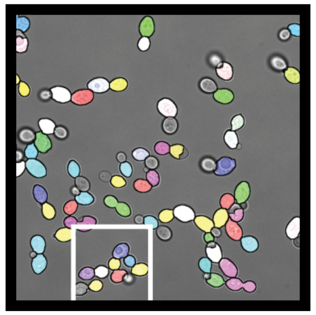
Explain observed data points by packing them, into separate coherent hypothesis



Detections of body parts
Some of which are false
People that respect the image data and human pose manifold



Detections of people
Some of which are false
Separate coherent tracks that respect how people look and move



Superpixels
Some of which are background
Separate cells that respect how cells look and the image data

Apply and adapt classic OR techniques for ILP problems in ML/CV

*Sentence color helps understand the category of example

© 2019 Verisk Analytics, Inc. All rights reserved.

MWSP: Entity Resolution (ER)

Explain observed data points by packing them, into separate coherent hypothesis

Name	Place	SSN
Geoff Hinton	Toronto	556
Yoshua Bengio	Montreal	223
Geoffrey Hinton	Ontario	556
Yann LeCun	NYC	361
Hinton	Canada	-
Y. Lecun	New York	367

Name	Place	SSN
Yoshua Bengio	Montreal,	223
Geoffrey	Toronto, Canada	556
Yann LeCun	NYC, USA	367

Observations of persons
Some of which are singletons
Individual people

© 2019 Verisk Analytics, Inc. All rights reserved.

Observations, Hypothesis, Cost Terms, Etc.

$d \in \mathscr{D}$: **Set of observations**

$g \in \mathscr{G}$: **Set of hypothesis= Power set of** $\mathscr{D}$

$G_{dg} = 1$ **if** g **includes** d **else** $G_{dg} = 0$

$\Gamma_g \in \mathbb{R}$: **Cost of** g

Common Form: $\mathscr{M}$ **is a finite set of models**

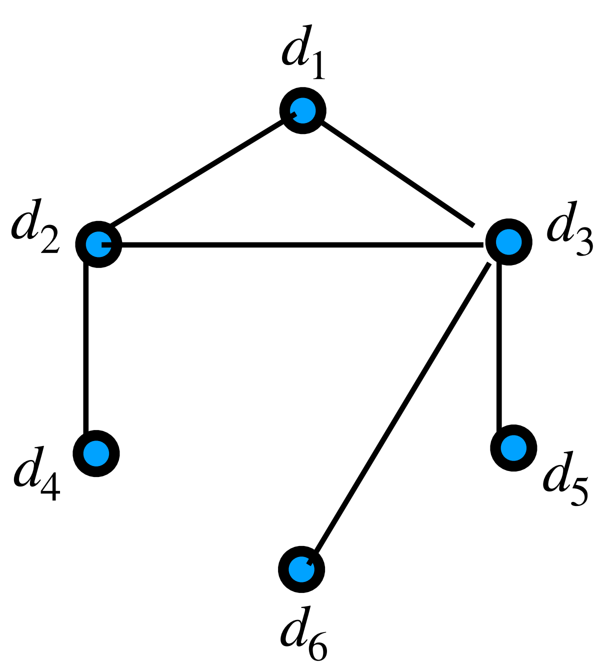
$$\Gamma_g = \min_{m \in \mathscr{M}} \theta_0^m + \sum_{d \in \mathscr{D}} \theta_d^m G_{dg} + \sum_{d_1, d_2 \in \mathscr{D}} \theta_{d_1 d_2}^m G_{d_1 g} G_{d_2 g}$$

© 2019 Verisk Analytics, Inc. All rights reserved.

ER: Observations, Hypothesis, Cost Terms

Set of Observations ($\mathscr{D}$)

- d_1
- d_2
- d_3
- d_4
- d_5
- d_6



Set of Hypothesis ($\mathscr{G}$)

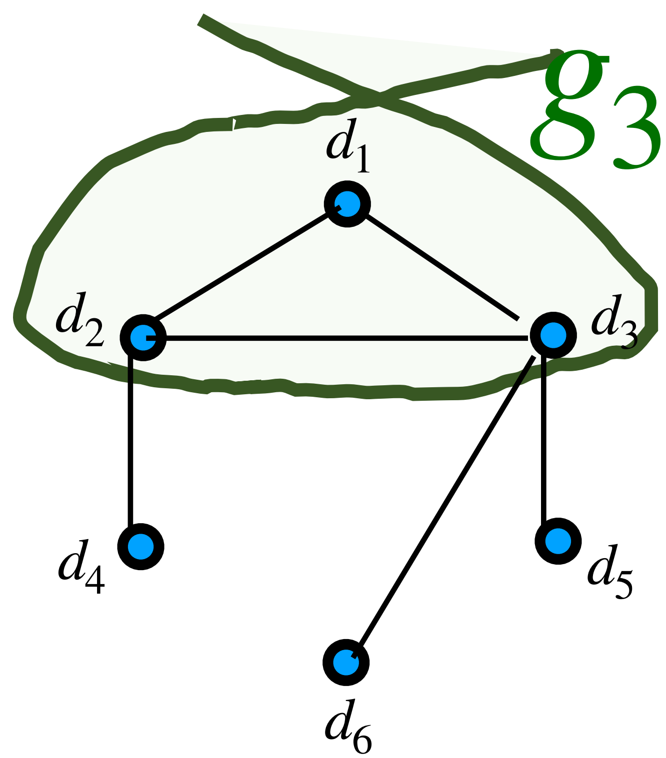
$g_1 = \{d_1, d_2\}$

$g_2 = \{d_3, d_4\}$

$\boxed{g_3 = \{d_1, d_2, d_3\}}$

$\vdots \qquad \vdots$

$g_{2^6} = \{d_1, d_2, d_3, d_4, d_5, d_6\}$



Cost Terms

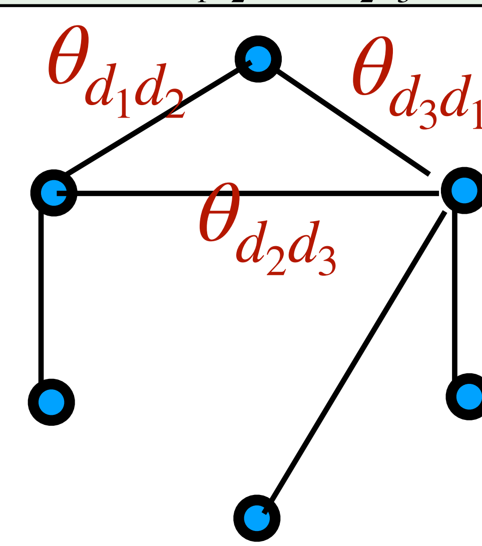
$\Gamma_{g_1} = 2\theta_{d_1 d_2}$

$\Gamma_{g_2} = 2\theta_{d_3 d_4}$

$\boxed{\Gamma_{g_3} = 2(\theta_{d_1 d_2} + \theta_{d_2 d_3} + \theta_{d_3 d_1})}$

$\vdots \qquad \vdots$

$\Gamma_{g_{2^6}} = 2(\theta_{d_1 d_2} + \theta_{d_2 d_3} + \ldots + \theta_{d_6 d_1})$



$\theta_{d_1 d_2}$:= **-Similarity of d_1 and d_2**

$\theta_{d_1 d_2} = \infty$ **almost always**

$$\Gamma_g = \sum_{d_1, d_2} \theta_{d_1 d_2} G_{d_1 g} G_{d_2 g}$$

© 2019 Verisk Analytics, Inc. All rights reserved.

MWSP and Column Generation

Select the lowest cost set of non-overlapping hypotheses

$\gamma_g = 1$ **iff g is selected**

$$\gamma \geq 0 \quad \min_{\cancel{\gamma \in \{0,1\}}} \sum_{g \in \hat{\mathscr{G}}} \Gamma_g \gamma_g$$

1. MWSP is NP-Hard (Karp 1972)

$$g \in \hat{\mathscr{G}} \quad \sum_{\cancel{g \in \mathscr{G}}} G_{dg} \gamma_g \leq 1 \ \forall d \in \mathscr{D}$$

2. $\mathscr{G}$ **is too large to enumerate**

Column Generation

Construct a small sufficient subset $\hat{\mathscr{G}}$ such that an optimal solution exists using only the hypothesis in $\hat{\mathscr{G}}$

© 2019 Verisk Analytics, Inc. All rights reserved.

The Column Algorithm: Initalize $\hat{\mathscr{G}} \leftarrow \{\,\}$

ITERATE

STEP 1 (SOLVE) over $\hat{\mathscr{G}} \subset \mathscr{G}$

Restricted Primal

$$\min_{\gamma \geq 0} \sum_{g \in \hat{\mathscr{G}}} \Gamma_g \gamma_g$$

$$\sum_{g \in \hat{\mathscr{G}}} G_{dg} \gamma_g \leq 1 \quad \forall d \in \mathscr{D}$$

Restricted Dual

$$\max_{\lambda \leq 0} \sum_{d \in \mathscr{D}} \lambda_d$$

$$\Gamma_g - \sum_{d \in \mathscr{D}} G_{dg} \lambda_d \geq 0 \quad \forall g \in \hat{\mathscr{G}}$$

STEP 2 (GROW): $\hat{\mathscr{G}} \leftarrow \hat{\mathscr{G}} \cup g^*$

The Pricing Problem

$$g^* \leftarrow \min_{g \in \mathscr{G}} \Gamma_g - \sum_{d \in \mathscr{D}} \lambda_d G_{dg}$$

© 2019 Verisk Analytics, Inc. All rights reserved.

Easy Pricing by Conditioning on Unique Rank

The Pricing Problem

$$\min_{g \in \mathcal{G}} \Gamma_g - \sum_{d \in \mathcal{D}} \lambda_d G_{dg}$$

$$g_{d*} \leftarrow \min_{g \in \mathcal{G}} \Gamma_g - \sum_{d \in \mathcal{D}} \lambda_d G_{dg}$$

$$G_{dg} = 0 \quad \forall d \notin \mathcal{D}_{d*}$$

$$G_{d_* g} = 1$$

r_d: Rank by $\sum_{\hat{d}, \theta_{d\hat{d}} < \infty} 1$ (low to high)

Condition on lowest r_{d*} s.t. $G_{d*g} = 1$

$$\mathcal{D}^*_{d*} = \{d \in \mathcal{D}, \theta_{dd_*} < \infty, r_d \geq r_{d*}\}$$

$$\hat{\mathcal{G}} \leftarrow \{\hat{\mathcal{G}} \cup g_{d_1} \cup g_{d_2} \cup g_{d_3} \ldots\}$$

Decrease # Sub-Problems:

1. Partial Pricing: Solve till find $\Gamma_g - \sum_{d \in \mathcal{D}} \lambda_d G_{dg} < 0$
2. Let $G_{d*g} \in \{0,1\}$ Ignore dominated $\mathcal{D}^*_{d*}$ sub-problems

© 2019 Verisk Analytics, Inc. All rights reserved.

Pricing as ILP: on $\mathscr{D}^*_{d^*}$

TINY BINARY PAIRWISE MRF:

$$G_{dg} \leftarrow x_d \qquad y_{d_1 d_2} = x_{d_1} x_{d_2}$$

Solve exactly or heuristically

$$\min_{x \in \{0,1\}, y \geq 0} \sum_{d \in \mathscr{D}^*_{d^*}} \lambda_d x_d + \sum_{d_1, d_2 \in \mathscr{D}^*_{d^*}} \theta_{d_1 d_2} y_{d_1 d_2}$$

$$y_{d_1 d_2} \leq x_{d_1}$$

$$y_{d_1 d_2} \leq x_{d_2}$$

$$-y_{d_1 d_2} + x_{d_1} + x_{d_2} \leq 1$$

© 2019 Verisk Analytics, Inc. All rights reserved.

Invariant Dual Optimal Inequalities

Dual

$$-\Xi_d \le \quad \max_{\lambda_d \le 0} \sum_{d\in\mathcal{D}} \lambda_d$$

$$\Gamma_g - \sum_{d\in\mathcal{D}} G_{dg}\lambda_d \ge 0 \ \forall g \in \hat{\mathcal{G}}$$

Reduce feasible region while preserving an optimal solution provably

Primal

$$\min_{\substack{\gamma \ge 0 \\ \xi \ge 0}} \sum_{g\in\hat{\mathcal{G}}} \Gamma_g \gamma_g + \sum_{d\in\mathcal{D}} \Xi_d \xi_d$$

$$-\xi_d + \sum_{g\in\hat{\mathcal{G}}} G_{dg}\gamma_g \le 1 \quad \forall d \in \mathcal{D}$$

Relax constraints and add a soft penalty

Penalty provably inactive at termination

$\xi_d :=$ **# violation** d

$$\Xi_d \ge \Gamma_g - \Gamma_{g-d} \quad \forall g \in \mathcal{G}$$

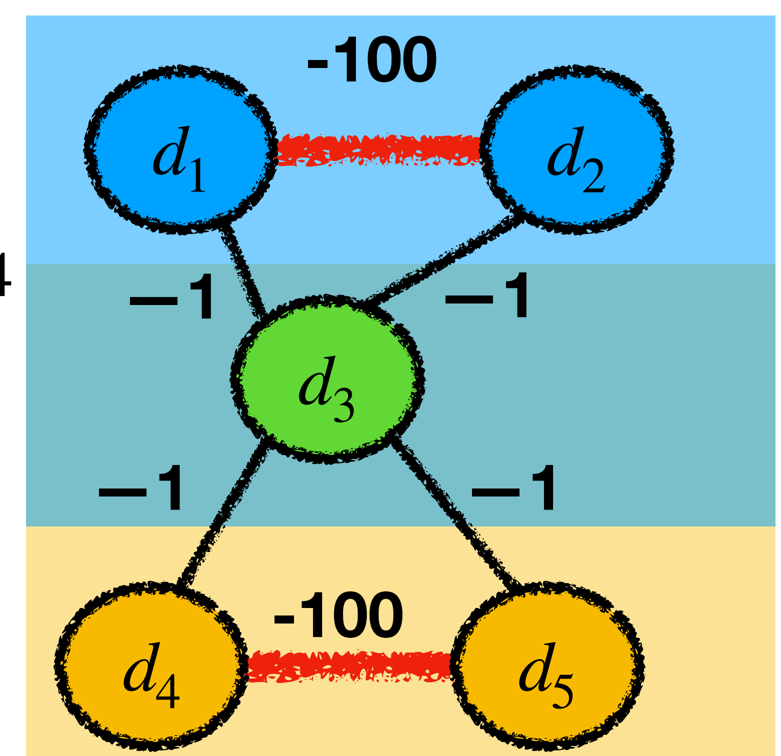
$$\Xi_d = -\sum_{d_1\in\mathcal{D}} 2\theta^-_{dd_1}$$

Ξ_d **is -sum rewards of d**

NO DOI: $\gamma_g = 1 \quad \gamma_{\hat{g}} = 0,$ **OBJ** $= -204$

DOI $\gamma_g = \gamma_{\hat{g}} = \xi_{d_3} = 1,$ **OBJ** $= -400$

True ILP Opt over all $\mathcal{G}$ **= -404**



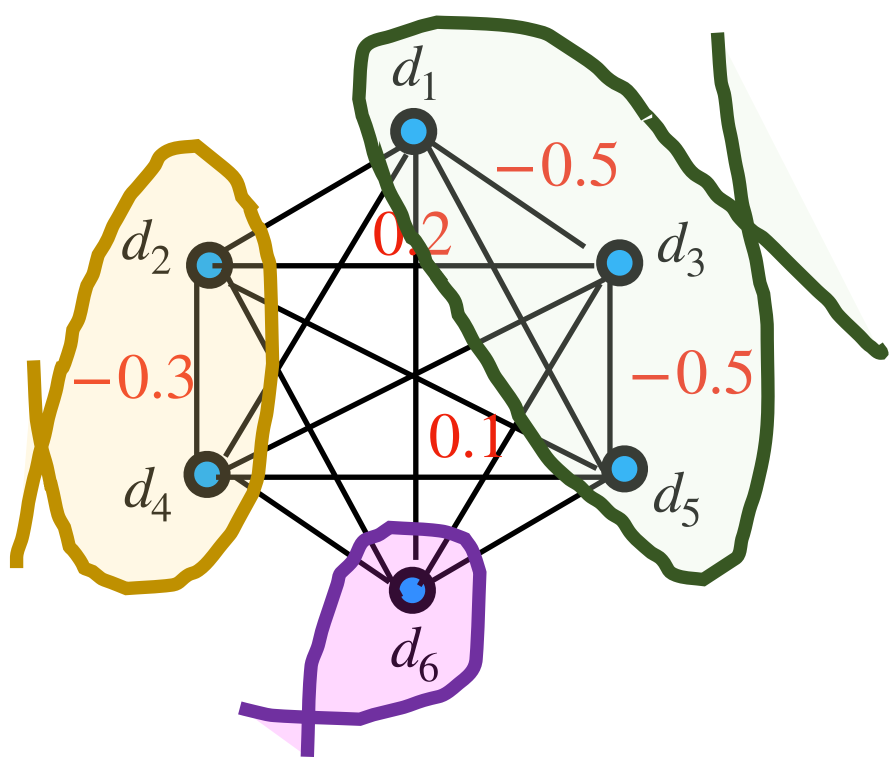
SERVE | ADD VALUE | INNOVATE

© 2019 Verisk Analytics, Inc. All rights reserved.

The Complete Pipeline

Input Dataset → Blocking → Scoring → Clustering → Output Dataset

Name	Place	Age	
Geoff Hinton	Toronto	56	d_1
Yoshua	Montreal	55	d_2
Geoffrey	Ontario	69	d_3
Yann LeCun	NYC	61	d_4
Hinton	Canada	78	d_5
Y. Lecun	New York	94	d_6



Name	Place	Ag
Yoshua	Montreal,	55
Geoffrey	Toronto,	69
Yann LeCun	NYC, USA	56

Blocking: Remove obvious non-matches by comparing a certain key attribute
Scoring: A score represents how close one observation is to another.
Clustering: With MWSP and F-DOIs

© 2019 Verisk Analytics, Inc. All rights reserved.

Results

Method	Settlements	Music 20K
ConCom	0.65	0.26
CCPivot	0.90	0.74
Center	0.88	0.66
MergCenter	0.68	0.39
Star1	0.82	0.62
Star2	0.92	0.69
F-MWSP	**0.96**	**0.81**

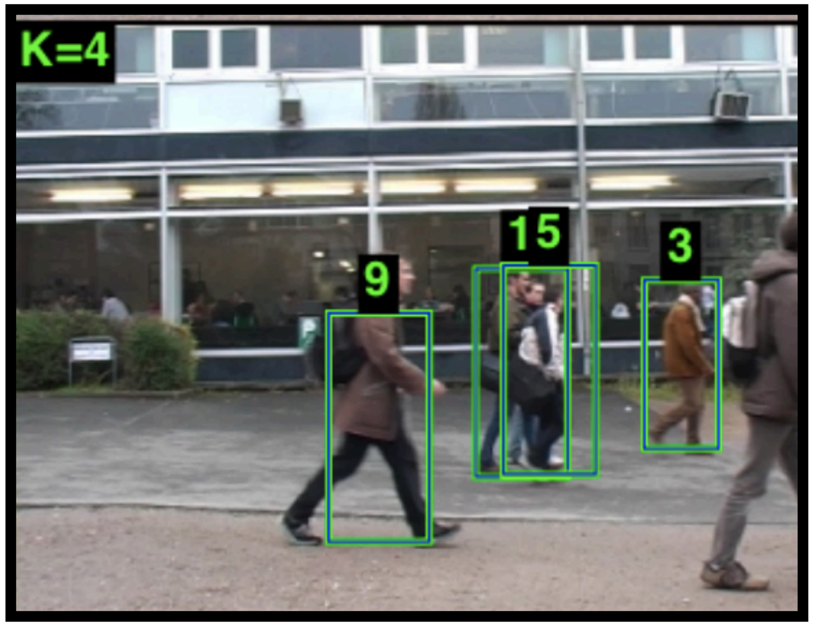
Datasets

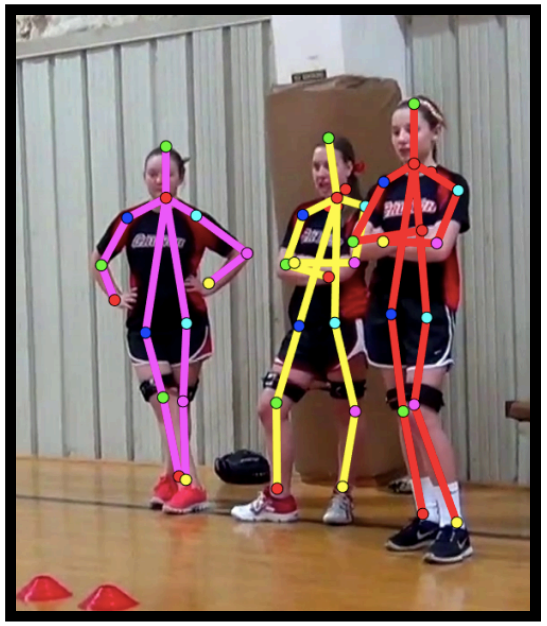
Settlements - 3054 observations - 820 clusters

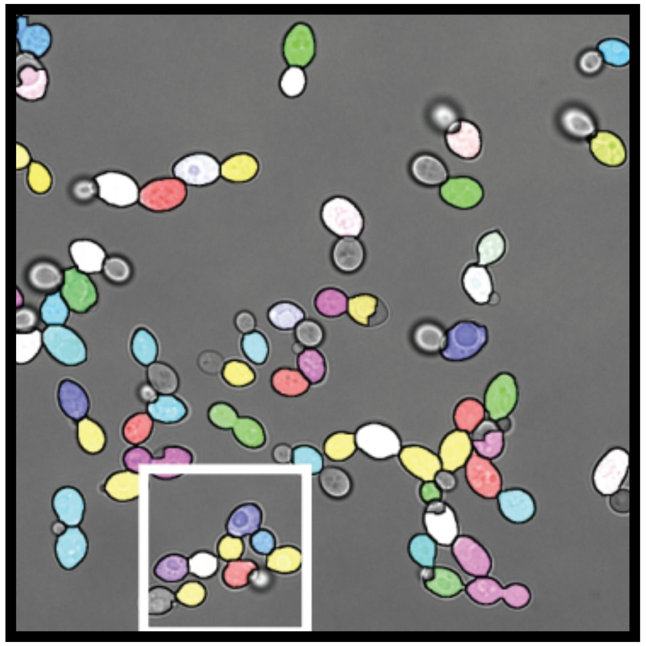
Music 20K - 19375 observations - 10000 clusters

© 2019 Verisk Analytics, Inc. All rights reserved.

APPENDIX FOR CONTENT BASED ON AUDIENCE INTEREST







Multi-Person Pose Estimation (MPPE)

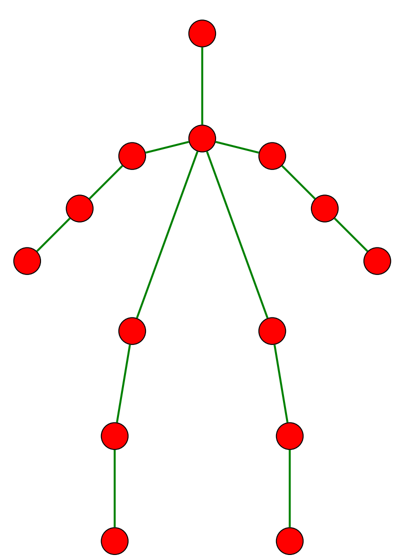
$$\Gamma_g = \min_{m \in \mathscr{M}} \theta_0^m + \sum_{d \in \mathscr{D}} \theta_d^m G_{dg} + \sum_{d_1, d_2 \in \mathscr{D}} \theta_{d_1 d_2}^m G_{d_1 g} G_{d_2 g}$$

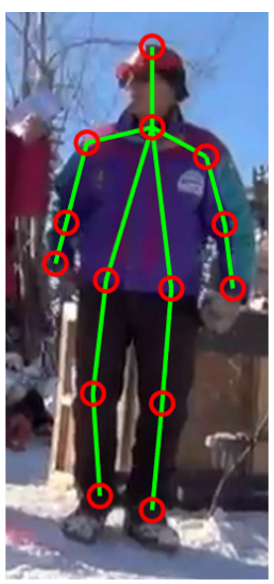
$r \in \mathscr{R}$: **Set of body parts. Each** m **is associated with a tree over** $\mathscr{R}$

$d \in \mathscr{D}^r$: **Set of detections of part** r.

$\theta_{\hat{d}d}^m$ **is zero if** $\hat{d}, d$ **correspond to non-adjacent parts in the tree of model** m

Due to occlusion and non-max suppression body parts can be replicated in a person



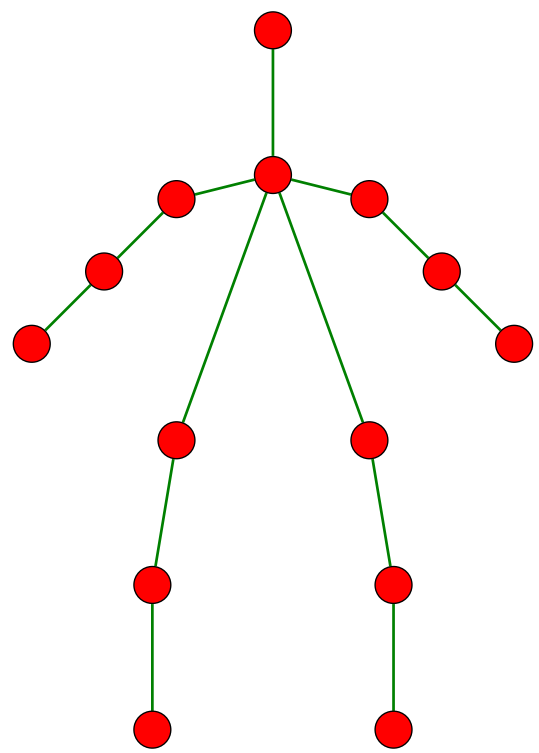


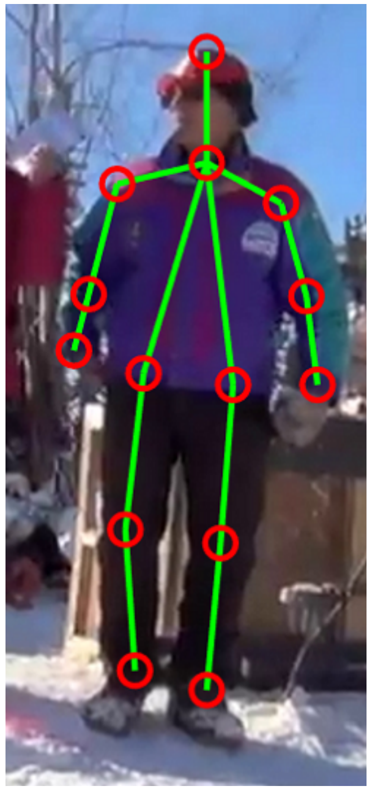
© 2019 Verisk Analytics, Inc. All rights reserved.

Pricing MPPE: Condition on m. $G_{dg} \leftarrow x_d$

$$\min_{x_d \in \{0,1\}} \theta_0^m + \sum_{d \in \mathcal{D}} \theta_d^m x_d + \sum_{d_1, d_2 \in \mathcal{D}} \theta_{d_1 d_2}^m x_{d_1} x_{d_2}$$

Solve as a dynamic program. Nodes correspond to body parts
Nodes have high state space





© 2019 Verisk Analytics, Inc. All rights reserved.

Why In Practice MWSP Tight in Practice?

MWSP LP >= local + cycle + odd wheel + bicycle odd wheel

Where local polytope: observations take labels in {1, 2, 3, … $|\mathscr{D}|$ }

Integrality Property.

More non-integrality in the LP of the pricing problem → More Tightening.

We can make even tighter with subset row inequalities.

© 2019 Verisk Analytics, Inc. All rights reserved.

Branch+Price: Using B&B in CG

Optima solution need not lie in $\hat{\mathscr{G}}$ unless γ is binary.

Branch on $\sum_{g \in \mathscr{G}} \gamma_g G_{d_1 g} G_{d_2 g} \notin \{0,1\}$ **Easy to enforce in pricing**

Branch 1: $\sum_{g \in \mathscr{G}} \gamma_g G_{d_1 g} G_{d_2 g} = 1 :\ d_1, d_2$ **must be grouped together.**

Branch 2: $\sum_{g \in \mathscr{G}} \gamma_g G_{d_1 g} G_{d_2 g} = 0 : d_1, d_2$ **must never be grouped together.**

© 2019 Verisk Analytics, Inc. All rights reserved.

Subset Row Inequalities: Size Three

Problematic Example

$$g_1 = \{d_1, d_2\}, g_2 = \{d_1, d_3\}, g_3 = \{d_2, d_3\}, g_4 = \{d_1, d_2, d_3\}$$

$$\Gamma_{g_1} = \Gamma_{g_2} = \Gamma_{g_3} = -1 \qquad \Gamma_{g_4} = -1.1$$

Optimal Fractional Solution: $\gamma_{g_1} = \gamma_{g_2} = \gamma_{g_3} = \frac{1}{2} \qquad \gamma_{g_4} = 0$ **Objective is** -1.5

Optimal Binary Solution: $\gamma_{g_1} = \gamma_{g_2} = \gamma_{g_3} = 0 \qquad \gamma_{g_4} = 1$ **Objective is** -1.1

For any group of 3 observations the number of hypothesis including 2 or more observations is bounded by 1

$$\sum_{g \in \mathcal{G}} \gamma_g \lfloor \frac{G_{d_1 g} + G_{d_2 g} + G_{d_3 g}}{2} \rfloor \leq 1$$

Fit elegantly into pricing
Can be easily separated
General Case Exists

© 2019 Verisk Analytics, Inc. All rights reserved.

Pricing with Dual Vars of Subset-row Ineq = ψ.

TINY BINARY PAIRWISE MRF:

$$G_{dg} \leftarrow x_d \qquad y_{d_1 d_2} = x_{d_1} x_{d_2} \qquad z_{d_1 d_2 d_3} = \lfloor \frac{G_{d_1 g} + G_{d_2 g} + G_{d_3 g}}{2} \rfloor$$

Solve exactly or heuristically

$$\min_{\substack{x \in \{0,1\}, y \geq 0 \\ z \geq 0}} \sum_{d \in \mathscr{D}^*_{d*}} -\lambda_d x_d + \sum_{d_1, d_2 \in \mathscr{D}^*_{d*}} \theta_{d_1 d_2} y_{d_1 d_2} - \sum_{d_1 d_2 d_3 \in \hat{C}} \psi_{d_1 d_2 d_3} z_{d_1 d_2 d_3}$$

$$y_{d_1 d_2} \leq x_{d_1} \qquad y_{d_1 d_2} \leq z_{d_1 d_2 d_3}$$

$$y_{d_1 d_2} \leq x_{d_2} \qquad y_{d_1 d_3} \leq z_{d_1 d_2 d_3}$$

$$-y_{d_1 d_2} + x_{d_1} + x_{d_2} \leq 1 \qquad y_{d_2 d_3} \leq z_{d_1 d_2 d_3}$$

© 2019 Verisk Analytics, Inc. All rights reserved.

Varying DOI: Tighter DOI since $\hat{\mathscr{G}} \subset \mathscr{G}$

$$\sum_{d \in \hat{\mathscr{D}}} \Xi_{dg} \geq \Gamma_{g-\hat{\mathscr{D}}} - \Gamma_g$$ **Increase in cost of removing $\hat{\mathscr{D}}$ from g**

For ER: $$\Xi_{dg} = -G_{dg}\Big(\sum_{\hat{d} \in \mathscr{D}} G_{\hat{d}g}(\theta_{d\hat{d}} - \theta^{-}_{d\hat{d}})\Big)^{-}$$

Include all potential penalties with weight 0.5 and all potential rewards once

© 2019 Verisk Analytics, Inc. All rights reserved.

Flexible Dual Optimal Inequalities (F-DOIs)

Model where removal occurs → Lower Obj given $\hat{\mathscr{G}}$ → Faster CG convergence

$$\min_{\substack{\gamma \geq 0 \\ \xi \geq 0}} \sum_{g \in \hat{\mathscr{G}}} \Gamma_g \gamma_g + \sum_{d \in \mathscr{D}, z \in \mathscr{Z}_d} z \xi_{dz}$$

$$\sum_{z \in \mathscr{Z}_d} -\xi_{dz} + \sum_{g \in \hat{\mathscr{G}}} G_{dg} \gamma_g \leq 1 \quad \forall d \in \mathscr{D}$$

$$\xi_{dz} \leq \sum_{g \in \hat{\mathscr{G}}} Z_{dzg} \gamma_g \qquad \forall d \in \mathscr{D}, z \in \mathscr{Z}_d$$

Quantize and round up Ξ_{dg}

$\lceil \Xi_{dg} \rceil = z \leftrightarrow Z_{dgz} = 1$

$\mathscr{Z}_d =$ **Unique**$(\Xi_{dg}, \forall g \in \hat{\mathscr{G}})$

$\xi_{dg} :=$ **# violation for d using z**

Price Using λ as usual

© 2019 Verisk Analytics, Inc. All rights reserved.

Price Using λ !!!! WHY

$$\max_{\lambda \le 0, \pi \le 0} \sum_{d \in \mathscr{D}} \lambda_d$$

$$\Gamma_g - \sum_{g \in \hat{\mathscr{G}}} G_{dg}\lambda_d + \sum_{d \in \mathscr{D}, z \in \mathscr{Z}_d} Z_{dgz}\pi_{zd} \ge 0 \qquad \forall g \in \hat{\mathscr{G}}$$

$$z + \lambda_d \ge \pi_{dz} \quad \forall d \in \mathscr{D}, z \in \mathscr{Z}_d$$

$$0 \le \Gamma_g - \sum_{d \in \mathscr{D}} G_{dg}\lambda_d + \sum_{d \in \mathscr{D}, z \in \mathscr{Z}_d} Z_{dgz}\pi_{zd} \le \Gamma_g - \sum_{d \in \mathscr{D}} G_{dg}\lambda_d \qquad \forall g \in \hat{\mathscr{G}}$$

© 2019 Verisk Analytics, Inc. All rights reserved.