Machine Learning

Structured Models:
Hidden Markov Models versus Conditional Random Fields

Eric Xing
Lecture 13, August 15, 2010

Reading:
From static to dynamic mixture models

Static mixture

Dynamic mixture

The underlying source:
Speech signal, dice,

The sequence:
Phonemes, sequence of rolls,
Hidden Markov Model

- **Observation space**
  - Alphabetic set: \( C = \{c_1, c_2, \ldots, c_K\} \)
  - Euclidean space: \( \mathbb{R}^d \)

- **Index set of hidden states**
  \( I = \{1, 2, \ldots, M\} \)

- **Transition probabilities** between any two states

\[
p(y_t | y_{t-1} = 1) = a_{i,j},
\]

or

\[
p(y_t | y_{t-1} = 1) \sim \text{Multinomial}(a_{i,1}, a_{i,2}, \ldots, a_{i,M}), \forall i \in I.
\]

- **Start probabilities**

\[
p(y_1) \sim \text{Multinomial}(\pi_1, \pi_2, \ldots, \pi_M).
\]

- **Emission probabilities** associated with each state

\[
p(x_t | y_t = 1) \sim \text{Multinomial}(b_{i,1}, b_{i,2}, \ldots, b_{i,K}), \forall i \in I.
\]

or in general:

\[
p(x_t | y_t = 1) \sim f(\cdot | \theta_i), \forall i \in I.
\]
Probability of a Parse

- Given a sequence \( x = x_1 \ldots x_T \)
  and a parse \( y = y_1, \ldots, y_T \),
- To find how likely is the parse:
  (given our HMM and the sequence)

\[
p(x, y) = p(x_1 \ldots x_T, y_1, \ldots, y_T) \quad \text{(Joint probability)}
\]

\[
= p(y_1) \ p(x_1 | y_1) \ p(y_2 | y_1) \ p(x_2 | y_2) \ldots \ p(y_T | y_{T-1}) \ p(x_T | y_T)
\]

\[
= p(y_1) \ p(y_2 | y_1) \ldots \ p(y_T | y_{T-1}) \times p(x_1 | y_1) \ p(x_2 | y_2) \ldots \ p(x_T | y_T)
\]

- Marginal probability:
  \[ p(x) = \sum_y p(x, y) = \sum_{y_1} \sum_{y_2} \ldots \sum_{y_N} \pi_{y_1} \prod_{t=2}^{T} a_{y_{t-1}, y_t} \prod_{t=1}^{T} p(x_t | y_t) \]

- Posterior probability:
  \[ p(y | x) = p(x, y) / p(x) \]
Shortcomings of Hidden Markov Model

- HMM models capture dependences between each state and only its corresponding observation
  - NLP example: In a sentence segmentation task, each segmental state may depend not just on a single word (and the adjacent segmental stages), but also on the (non-local) features of the whole line such as line length, indentation, amount of white space, etc.

- Mismatch between learning objective function and prediction objective function
  - HMM learns a joint distribution of states and observations $P(Y, X)$, but in a prediction task, we need the conditional probability $P(Y|X)$
Recall Generative vs. Discriminative Classifiers

- Goal: Wish to learn \( f: X \rightarrow Y \), e.g., \( P(Y|X) \)

- Generative classifiers (e.g., Naïve Bayes):
  - Assume some functional form for \( P(X|Y), P(Y) \)
  - This is a ‘generative’ model of the data!
  - Estimate parameters of \( P(X|Y), P(Y) \) directly from training data
  - Use Bayes rule to calculate \( P(Y|X=x) \)

- Discriminative classifiers (e.g., logistic regression)
  - Directly assume some functional form for \( P(Y|X) \)
  - This is a ‘discriminative’ model of the data!
  - Estimate parameters of \( P(Y|X) \) directly from training data
Structured Conditional Models

- Conditional probability $P(\text{label sequence } y \mid \text{observation sequence } x)$ rather than joint probability $P(y, x)$
  - Specify the probability of possible label sequences given an observation sequence

- Allow arbitrary, non-independent features on the observation sequence $X$

- The probability of a transition between labels may depend on past and future observations

- Relax strong independence assumptions in generative models
Conditional Distribution

- If the graph $G = (V, E)$ of $Y$ is a tree, the conditional distribution over the label sequence $Y = y$, given $X = x$, by the Hammersley Clifford theorem of random fields is:

$$p_\theta(y | x) \propto \exp \left( \sum_{e \in E,k} \lambda_k f_k (e, y|_e, x) + \sum_{v \in V,k} \mu_k g_k (v, y|_v, x) \right)$$

  - $x$ is a data sequence
  - $y$ is a label sequence
  - $v$ is a vertex from vertex set $V = \text{set of label random variables}$
  - $e$ is an edge from edge set $E$ over $V$
  - $f_k$ and $g_k$ are given and fixed. $g_k$ is a Boolean vertex feature; $f_k$ is a Boolean edge feature
  - $k$ is the number of features
  - $\theta = (\lambda_1, \lambda_2, \cdots, \lambda_n; \mu_1, \mu_2, \cdots, \mu_n)$; $\lambda_k$ and $\mu_k$ are parameters to be estimated
  - $y|_e$ is the set of components of $y$ defined by edge $e$
  - $y|_v$ is the set of components of $y$ defined by vertex $v$
Conditional Random Fields

CRF is a partially directed model
- Discriminative model
- Usage of global normalizer $Z(x)$
- Models the dependence between each state and the entire observation sequence

$$P(y_{1:n} | x_{1:n}) = \frac{1}{Z(x_{1:n})} \prod_{i=1}^{n} \phi(y_{i}, y_{i-1}, x_{1:n}) = \frac{1}{Z(x_{1:n}, w)} \prod_{i=1}^{n} \exp(w^T f(y_{i}, y_{i-1}, x_{1:n}))$$
Conditional Random Fields

- General parametric form:

$$P(y|x) = \frac{1}{Z(x, \lambda, \mu)} \exp\left( \sum_{i=1}^{n} \left( \sum_{k} \lambda_k f_k(y_i, y_{i-1}, x) + \sum_{l} \mu_l g_l(y_i, x) \right) \right)$$

$$= \frac{1}{Z(x, \lambda, \mu)} \exp\left( \sum_{i=1}^{n} \left( \lambda^T f(y_i, y_{i-1}, x) + \mu^T g(y_i, x) \right) \right)$$

where $Z(x, \lambda, \mu) = \sum_{y} \exp\left( \sum_{i=1}^{n} \left( \lambda^T f(y_i, y_{i-1}, x) + \mu^T g(y_i, x) \right) \right)$
Conditional Random Fields

\[ p_\theta(y \mid x) = \frac{1}{Z(\theta, x)} \exp \left\{ \sum_c \theta_c f_c(x, y_c) \right\} \]

- Allow arbitrary dependencies on input
- Clique dependencies on labels
- Use approximate inference for general graphs
CRFs: Inference

- Given CRF parameters $\lambda$ and $\mu$, find the $y^*$ that maximizes $P(y|x)$

$$y^* = \arg \max_y \exp \left( \sum_{i=1}^{n} (\lambda^T f(y_i, y_{i-1}, x) + \mu^T g(y_i, x)) \right)$$

- Can ignore $Z(x)$ because it is not a function of $y$

- Run the max-product algorithm on the junction-tree of CRF:

Same as Viterbi decoding used in HMMs!
**CRF learning**

- Given \( \{(x_d, y_d)\}_{d=1}^N \), find \( \lambda^*, \mu^* \) such that

\[
\lambda^*, \mu^* = \arg \max_{\lambda, \mu} L(\lambda, \mu) = \arg \max_{\lambda, \mu} \prod_{d=1}^N P(y_d|x_d, \lambda, \mu) \\
= \arg \max_{\lambda, \mu} \prod_{d=1}^N \frac{1}{Z(x_d, \lambda, \mu)} \exp\left( \sum_{i=1}^n (\lambda^T f(y_{d,i}, y_{d,i-1}, x_d) + \mu^T g(y_{d,i}, x_d)) \right) \\
= \arg \max_{\lambda, \mu} \sum_{d=1}^N \left( \sum_{i=1}^n (\lambda^T f(y_{d,i}, y_{d,i-1}, x_d) + \mu^T g(y_{d,i}, x_d)) - \log Z(x_d, \lambda, \mu) \right)
\]

- Computing the gradient w.r.t \( \lambda \):

\[
\nabla_\lambda L(\lambda, \mu) = \sum_{d=1}^N \sum_{i=1}^n f(y_{d,i}, y_{d,i-1}, x_d) - \sum_y (P(y|x_d) \sum_{i=1}^n f(y_{d,i}, y_{d,i-1}, x_d))
\]

Gradient of the log-partition function in an exponential family is the expectation of the sufficient statistics.
CRFs: some empirical results

- Comparison of error rates on synthetic data

Data is increasingly higher order in the direction of arrow

CRFs achieve the lowest error rate for higher order data
CRFs: some empirical results

- Parts of Speech tagging

<table>
<thead>
<tr>
<th>model</th>
<th>error</th>
<th>oov error</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>5.69%</td>
<td>45.99%</td>
</tr>
<tr>
<td>MEMM</td>
<td>6.37%</td>
<td>54.61%</td>
</tr>
<tr>
<td>CRF</td>
<td>5.55%</td>
<td>48.05%</td>
</tr>
<tr>
<td>MEMM+</td>
<td>4.81%</td>
<td>26.99%</td>
</tr>
<tr>
<td>CRF+</td>
<td>4.27%</td>
<td>23.76%</td>
</tr>
</tbody>
</table>

+ Using spelling features

- Using same set of features: HMM >= CRF > MEMM
- Using additional overlapping features: CRF+ > MEMM+ >> HMM
Summary

- Conditional Random Fields is a discriminative Structured Input Output model!

- HMM is a generative structured I/O model

- Complementary strength and weakness:
  1. 
  2. 
  3. 
  ...

Condition Random Fields: applications in vision

L. Fei-Fei
Computer Science Dept.
Stanford University
Machine learning in computer vision

- Aug 15, Lecture 13: Conditional Random Field
  - Applications in computer vision
  - Image labeling, segmentation, object recognition & image annotation
Image labeling
(slides courtesy to Sanjiv Kumar (Google))

Reference:
Context helps visual recognition

Context from Larger Neighborhoods!
Context helps visual recognition

Context from Whole Image!
Contextual Interactions

Pixel-Pixel
8 August 2010

Region-Region

Block-Block
22
Contextual Interactions

Part-Part

Object-Object

Region-Object

8 August 2010
Modeling Contextual Interactions

• Framework to learn all relevant contextual interactions in a single model automatically from training data

• Probabilistic models
  – Principled way to deal with ambiguities

• Graphical models
  – Powerful framework for ensuring global consistency using relatively local constraints

Undirected graphs = Random Fields
No context  

With context

No context  

With context

8 August 2010

[ Kumar '05 ]

25
Image Labeling Outline

- **Background**
  - Markov Random Fields (MRFs)

- **Conditional Random Fields (CRFs)**

- **Multiclass and Hierarchical Interactions**

- **Hidden Conditional Random Fields**

- **Semi-supervised Learning**

- **Open Issues**
Markov Random Field (MRF)

Data $y_i$

Label $x_i$

\{-1, 1\}

[ Geman & Geman '84 ]
Markov Random Field (MRF)

Data $y_i$
Label $x_i$
\{-1, 1\}

[ Geman & Geman '84 ]

8 August 2010
Markov Random Field (MRF)

Data \( y_i \)

Label \( x_i \)

\( x = \{x_i\}_{i \in S} \)

\( y = \{y_i\}_{i \in S} \)

Generative Framework

\[ P(x|y) \propto P(x, y) = P(y|x)P(x) \]

We want

Observation Model

Prior Model (MRF)

[ Geman & Geman '84 ]
Markov Random Field (MRF)

Data $y_i$
Label $x_i$

$x = \{x_i\}_{i \in S}$
$y = \{y_i\}_{i \in S}$

Generative Framework

\[
P(x|y) \propto P(x, y) = P(y|x)P(x)
\]

Observation Model
Prior Model (MRF)

We want

\[
P(y|x) \approx \prod_{i \in S} P(y_i|x_i)
\]

Too restrictive!

8 August 2010

[ Geman & Geman ’84 ]
Markov Random Field (MRF)

Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i | x_{S \setminus \{i\}}) = P(x_i | x_{N_i})$
Markov Random Field (MRF)

Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i | x_{S\setminus\{i\}}) = P(x_i | x_{N_i})$

Positivity $P(x) > 0 \quad \forall x$

$P(x) \propto \prod_{e \in E} \Psi_e(x_i, x_j)$

[ Hammersley & Clifford ’71 ]
Markov Random Field (MRF)

Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i | x_{S \setminus \{i\}}) = P(x_i | x_{N_i})$

Positivity $P(x) > 0 \quad \forall x$

$P(x) \propto \prod_{e \in E} \Psi_e(x_i, x_j)$

[ Hammersley & Clifford ’71 ]

Ising Model $P(x) \propto \exp(\sum_{i \in S} \sum_{j \in N_i} \beta x_i x_j) \quad \beta > 0$

[ Ising ’25 ]

8 August 2010
Markov Random Field (MRF)

Label Prior $P(x)$ is modeled as MRF

Markovianity $P(x_i | x_{S \setminus \{i\}}) = P(x_i | x_{N_i})$

Positivity $P(x) > 0 \quad \forall x$

$$P(x) \propto \prod_{e \in E} \Psi_e(x_i, x_j)$$

[ Hammersley & Clifford ’71 ]

Ising Model $P(x) \propto \exp\left( \sum_{i \in S} \sum_{j \in N_i} \beta x_i x_j \right) \quad \beta > 0$

[ Ising ’25 ]

Traditional MRF

$$P(x | y) = \frac{1}{Z} \exp\left( \sum_{i \in S} \log P(y_i | x_i) + \sum_{i \in S} \sum_{j \in N_i} \beta x_i x_j \right)$$

8 August 2010
Generative vs. Discriminative

We want: $P(x|y)$

• Generative Framework
  – Models $P(x, y)$ to get $P(x|y)$
  – Implicit modeling of observations

• Discriminative Framework
  – Models conditional distribution $P(x|y)$ directly
Conditional Random Field (CRF)

[Lafferty et al. ’01]

- Conditional distribution $P(x|y)$ is modeled as MRF

$$P(x_i|x_{S\setminus\{i\}},y) = P(x_i|x_{N_i},y)$$

$$P(x|y) > 0 \quad \forall \ x$$

$$P(x|y) \propto \prod_{e\in E} \Psi_e(x_i, x_j, y)$$

[Hammersley & Clifford ’71]

- Segmentation and labeling of 1D text sequences
Conditional Random Field (CRF)

- Graphs on images with loops
  - Grid or irregular topology

- Potentials using arbitrary discriminative classifiers
  - Discriminative Random Field (DRF)

[Kumar & Hebert, ICCV'03]
Conditional Random Field (CRF)

- Graphs on images with loops
  - Grid or irregular topology
- Potentials using arbitrary discriminative classifiers
- If only unary and pairwise potentials are non-zero

\[
P(x|y) = \frac{1}{Z} \exp \left( \sum_{i \in S} A_i(x_i, y) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(x_i, x_j, y) \right)
\]

[Association Potential]

[Interaction Potential]

[Kumar & Hebert, ICCV’03]

8 August 2010
Comparison with MRF framework

Traditional MRF

\[ P(x|y) = \frac{1}{Z} \exp \left( \sum_{i \in S} \log P(y_i|x_i) + \sum_{i \in S} \sum_{j \in N_i} \beta x_i x_j \right) \]

CRF

\[ P(x|y) = \frac{1}{Z} \exp \left( \sum_{i \in S} A(x_i, y) + \sum_{i \in S} \sum_{j \in N_i} I(x_i, x_j, y) \right) \]

- Data from multiple sites
- Data-dependent label interactions
Association Potential $A(x_i, y)$
Association Potential $A(x_i, y)$

$$A(x_i, y) = \log P(x_i \mid f_i(y))$$
Association Potential $A(x_i, y)$

$A(x_i, y) = \log P(x_i | f_i(y))$

$= \log \sigma(x_i w^T f_i(y))$

Other classifier choices: [Szummer et al. ‘04] [He et al. ‘04] [Torralba et al. ‘05]
Association Potential $A(x_i, y)$

$$A(x_i, y) = \log P(x_i | f_i(y))$$

$$= \log \sigma(x_i w^T f_i(y))$$

$$f_i(y) \xrightarrow{\phi(.)} \phi_i(y)$$

$$A(x_i, y) = \log \sigma(x_i w^T \phi_i(y))$$

Other classifier choices: [Szummer et al. ‘04] [He et al. ‘04] [Torralba et al. ‘05]
Interaction Potential $I(x_i, x_j, y)$

[Image: Diagram illustrating the interaction potential with nodes $x_i$ and $x_j$ connected by an edge, and functions $\psi_i(y)$ and $\psi_j(y)$ connected to $y$.]
Interaction Potential $I(x_i, x_j, y)$

Pairwise discriminative classifier

$$I(x_i, x_j, y) = \log P(x_i, x_j | \psi_i, \psi_j)$$

$$= x_i x_j v^T \mu_{ij}(y)$$

$\beta$

[8 August 2010]

[Kumar & Hebert, NIPS '04]
**Learning and Inference**

Given input image $y$ and $P(x|y)$, get the optimal labels $x$

### Inference Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>MAP (min-cut)</th>
<th>MPM (BP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA (min-cut)</td>
<td>5.82</td>
<td>19.19</td>
</tr>
<tr>
<td>MMA (BP)</td>
<td>26.53</td>
<td>5.70</td>
</tr>
<tr>
<td>Contrastive Divergence</td>
<td>8.88</td>
<td>6.29</td>
</tr>
<tr>
<td>Pseudo-Likelihood</td>
<td>17.69</td>
<td>7.31</td>
</tr>
</tbody>
</table>

Pixelwise error (%) on 200 test images

Parameter Learning Methods

![Diagram with coupling arrows between SPA (min-cut), MMA (BP), MAP (min-cut), and MPM (BP)]

[Kumar et al. EMMCVPR ‘05]

8 August 2010
Man-Made Structure Detection

• Training Set – 130 images (256x384 pixels)
• Test Set – 108 images

- Scale variations
- Illumination variations
- Pose variations
- Non-linear structures
- Negative samples

- Gradient magnitude and orientation features

14 dim \rightarrow 119 dim

[8 August 2010]

[Kumar & Hebert CVPR ‘03]
Traditional MRF

CRF

[Kumar & Hebert
ICCV ‘03]

8 August 2010
Real-time Structure Detection
Hierarchical Interactions

Layer-2: Labels $x^{(2)}$

Layer-1: Labels $x^{(1)}$

Observed Image $y$

Partition $h$

Classification

[Kumar & Hebert, ICCV ’05]
Region Classification

![Region Classification Diagram]

- **Sky**
- **Water**
- **Sand**
- **Skin**
- **Grass**
- **Other**

Input image: Softmax (no context) 62.3% Layer-1 (label smoothing only) 63.8% Layer-2 (full model) 74% (~2 Sec)

[8 August 2010] [Kumar & Hebert ICCV '05]
Object-Region Interactions

MIT Dataset
[Torralba et al., ‘05]

8 August 2010
[Kumar & Hebert ICCV ‘05]
Object-Object Interactions

- MIT context database: 164 images (100x100 pixels)
- Very small objects (8x5 pixels) → High false positives
- Initial object detectors trained with gentle boosting
Applications of Conditional Random Fields in Computer Vision

Segmentation, Object recognition, Scene recognition, Human-object interaction
**ObjCut – CRF for image segmentation**

\[ \theta \text{ (shape parameter)} \]

Unary Potential: \( \phi_x(m_x | \theta) \)

Pairwise Potential: \( \phi_{xy}(m_x, m_y) \)

Contrast Term: \( \phi(D | m_x, m_y) \)

Unary Potential: \( \phi_x(D | m_x) \)

\( m \) (labels)

\( m_x \)

\( m_y \)

\( \theta \)

\( m \)

\( D \) (pixels)

Image Plane

[Kumar et al, CVPR 2005]
Segmentation Results

Shape only

Appearance only

Shape + Appearance

Without $\phi_x(D|m_x)$

Without $\phi_x(m_x|\theta)$

[Kumar et al, CVPR 2005]
Objects in Context – CRF for Object Recognition

[Original image] → [Image segmentation] → [Semantic context (Interaction terms)] → [Segment label tuning] (Unary terms) → [Rabinovich et al, CVPR 2005]
Object recognition results

Original image | Without context | With context
---|---|---

[Rabinovich et al, CVPR 2005]
Region-based Model – CRF for Scene recognition

\[ E(R, A, S, G, v^hz, K | I, \theta) = \]

- \( \psi_{\text{horizon}}(v^hz) \)
- \( \psi_{\text{region}}(S_r, G_r, A_r, v^hz) \)
- \( \psi_{\text{boundary}}(A_r, A_s) \)
- \( \psi_{\text{pair}}(S_r, S_s, G_r, G_s) \)

**Variables**
- \( \alpha_p \): pixel appearance
- \( R_p \): pixel-to-region correspondence
- \( A_r \): region appearance
- \( S_r \): region semantic class
- \( G_r \): region geometry
- \( v^hz \): location of horizon

[Gould et al, CVPR 2009]
Scene recognition results

[Scene recognition results image]

[8 August 2010 L. Fei-Fei, Dragon Star 2010, Stanford]

[Gould et al, CVPR 2009]
Mutual Context Model – CRF for human-object interaction

HOI activity: Tennis Forehand

[Yao and Fei-Fei, CVPR 2010]
Object Detection Results

Sliding window  Pedestrian context  Our Method
[Andriluka et al, 2009] [Dalal & Triggs, 2006]

Cricket bat

Cricket ball

Croquet mallet

Tennis racket

Volleyball

[Yao and Fei-Fei, CVPR 2010]
## Human Pose Estimation Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Torso</th>
<th>Upper Leg</th>
<th>Lower Leg</th>
<th>Upper Arm</th>
<th>Lower Arm</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramanan, 2006</td>
<td>.52</td>
<td>.22</td>
<td>.22</td>
<td>.21</td>
<td>.28</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
<td>.28</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.17</td>
<td>.14</td>
<td>.42</td>
</tr>
<tr>
<td>Andriluka et al, 2009</td>
<td>.50</td>
<td>.31</td>
<td>.30</td>
<td>.31</td>
<td>.27</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.18</td>
<td>.19</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.11</td>
<td>.45</td>
</tr>
<tr>
<td>Our model</td>
<td>.66</td>
<td>.43</td>
<td>.39</td>
<td>.44</td>
<td>.34</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.44</td>
<td>.40</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.27</td>
<td>.29</td>
<td>.58</td>
</tr>
</tbody>
</table>

## Activity Recognition Results

![Activity Recognition Results Graph]

- **Our model**: 83.3%
- **Gupta et al, 2009**: 78.9%
- **Bag-of-words SIFT+SVM**: 52.5%

[Yao and Fei-Fei, CVPR 2010]
Summary

• CRF-based discriminative models in Vision
  – Principled approach to model interactions at pixel, patch, region or object level for robust classification

• Combine local discriminative classifiers with data-dependent label interactions
  – Alternative to traditional MRFs

• Hierarchy of fields to capture different contexts

• Several computer vision tasks in the same framework
  – Denoising, region classification, texture recognition, object detection, gesture recognition, …
Open Questions

Features ↔ Model

Convolutional Networks
[Lecun & Bengio ‘98]

Ambiguity in recognition!