Random Field Models
for Applications in Computer Vision

Nazre Batool
Post-doctorate Fellow, Team AYIN,
INRIA Sophia Antipolis

Outline

- Graphical Models
- Generative vs. Discriminative Classifiers
- Markov Random Field Model
- Conditional Random Field Model
- Discriminative Random Field Model - *An Example*

Graphical Models[1]

- Graphical Models – **Graph** Theory + Probability **Models**

A graphical model is a family of probability distributions that factorizes according to an underlying graph

Factorization -> incorporate contextual constraints -> simplify computation

A distribution over a large number of random variables -> product of local functions of small number of variables only

$$P(Y) = \prod_{c} V_c(Y)$$

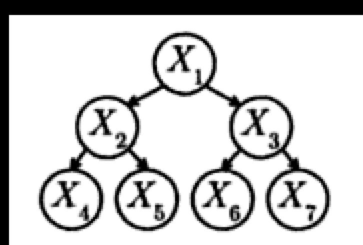
[1] Christopher M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006

Graphical Models[1]

How is Factorization achieved in a graphical model?

Conditional Independence in a graphical model

- The absence of an edge between two nodes shows the conditional independence of corresponding random variables
- Conditional independence/Absence of edges allows a complex probability distribution to be decomposed into factors
- Example : Directed Graphical Model (Bayesian Network)



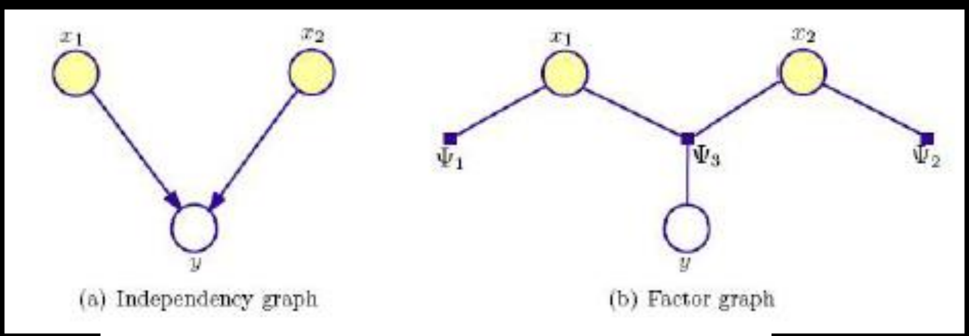
$$P(X_1, \ldots, X_n) = \Pi_{i=1}^{n} P(X_i | X_{\pi_i}).$$

[1] Christopher M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006

Graphical Models[1]

- Factor Graphs

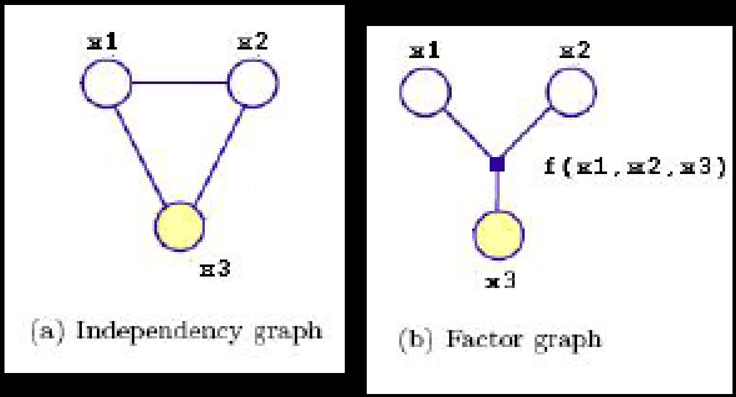
A bipartite graph to represent factorization /conditional independence



(a) Independency graph (b) Factor graph

$$p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2)$$
$$\Psi_1(x_1) = p(x_1)$$
$$\Psi_2(x_2) = p(x_2)$$
$$\Psi_3(x_3) = p(y|x_1, x_2)$$

Directed Graph



(a) Independency graph (b) Factor graph

$$P(x) \propto \prod_{c \in C} \psi_c(x_c)$$

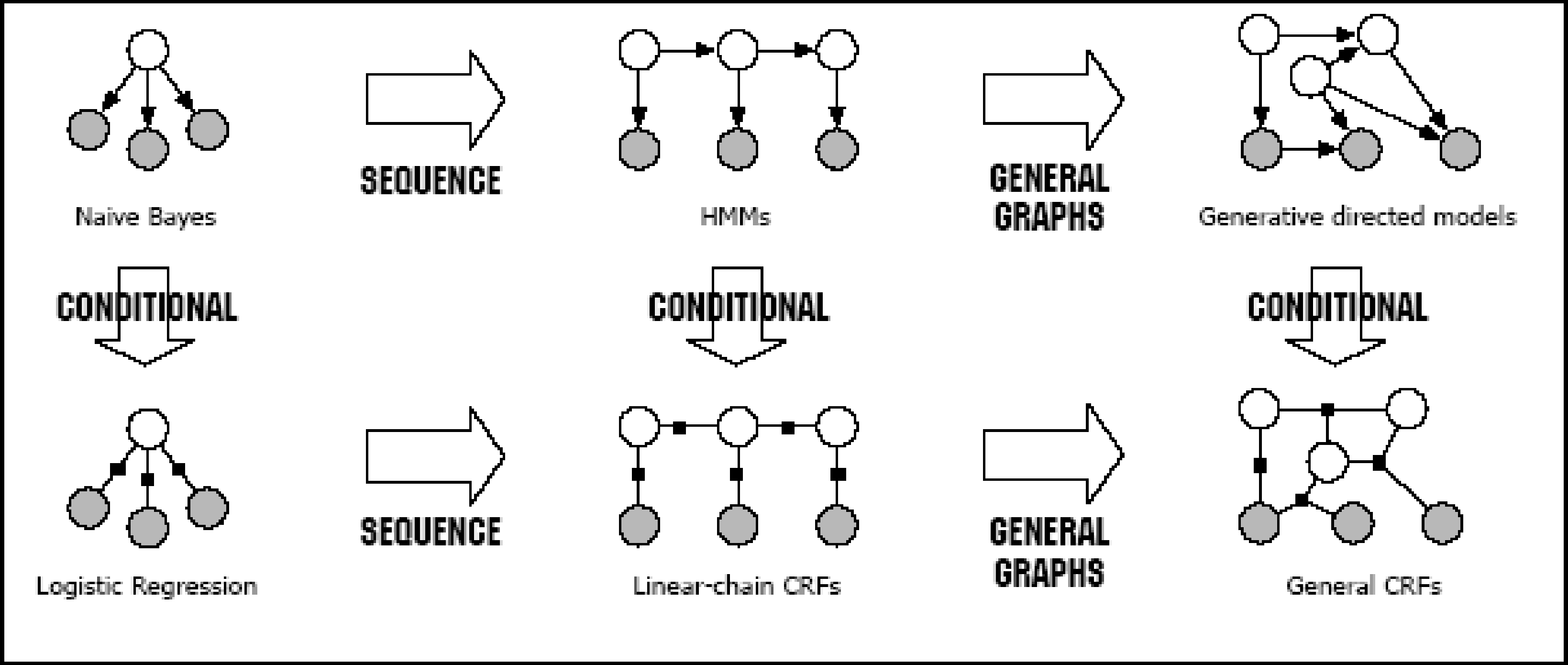
$$\psi(x_1, x_2, x_3)$$

Undirected Graph

[1] Christopher M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006

Generative vs. Discriminative Models for Classification[1]

Overview of relationship between Generative and Conditional/Discriminative Models [1]



[1] Charles Sutton, Andrew McCallum, An Introduction to Conditional Random Fields for Relational Learning. Introduction to Statistical Rela

Generative vs. Discriminative Models for Classification[1]

- Generative Models
  - Separately estimate a probability model for each class – generative term
  - Classify by comparing probabilities under these models to find the class with highest probability
  - Requires modeling of joint probability distribution of all variables
  - The joint distribution can be used for classification by conditioning and using marginalization, Bayes' rule, etc.

$$P(X_j) = \sum_{\forall X_i, i \neq j} P(X_1, \ldots, X_n)$$

$$P(X_j|X_k) = \frac{P(X_j, X_k)}{P(X_k)} = \frac{P(X_k|X_j)P(X_j)}{P(X_k)}$$

  - Examples: Gaussian, naïve Bayes, hidden Markov Model, Bayesian Network, **MRF**

[1] T. Jebara, *Machine Learning :* Discriminative and Generative (The Kluwer International Series in Engineering and Computer Science). Springer, December 2003.

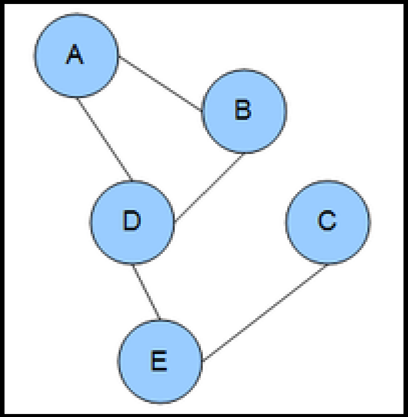
Generative vs. Discriminative Models for Classification[1]

- Discriminative Models
  - No modeling of underlying probability distribution of classes
  - Try to learn a classifier that performs well on the training data
  - Process examples from several classes to find the optimal boundaries among classes
  - Examples: Logistic Regression, Neural Networks, Support Vector Machines, **Conditional Random Field**

[1] T. Jebara, *Machine Learning :* Discriminative and Generative (The Kluwer International Series in Engineering and Computer Science). Springer, December 2003.

Markov Random Field

- Markov random field is defined on an undirected graph
- Nodes of graph are random variables having Markov property
- A random field is a Markov random field when:
  - The joint probability factorizes because of Markovianity property and
  - can be represented as a Gibbs distribution (Hammersely-Clifford theorem) i.e. in terms of cliques, normalization constant and temperature parameter.



$$P(Y) = \frac{1}{Z}\exp(-U(Y)) = \frac{1}{Z}\exp\sum_{c\in C}(-V_c(Y))$$

Markov Random Field

- How is MRF a generative model?[1]

Likelihood/Generative Term

$$P(labels \mid observations) = \frac{P(observations \mid labels) P(labels)}{P(observations)} \propto P(observations \mid labels) P(labels)$$

Prior probability model on labels

An MRF:

$$P(labels \mid observations) \propto -U(observations \mid labels) - U(labels)$$

An MRF:

$$P(labels) = \frac{1}{Z} \exp \sum_{c \in C} (-V_c(labels)) \propto -U(labels)$$

$$P(labels \mid observations) \propto P(observations \mid labels) P(labels)$$

Any exponential probability function:

$$P(observations \mid labels) = \frac{1}{Z} \exp - U(observations \mid labels)$$

[1] S. Z. Li, Markov Random Field Modeling in Image Analysis, Springer 2009.

Generative vs. Discriminative Models for Classification

Both types of classifiers have pros and cons

- Why Conditional Random Field (instead of MRF)?

- Modeling of joint probability distribution (Generative Model) is not always tractable/efficient
  - Independence assumptions are made for conditional, easy to compute, distributions/likelihood
  - Prior distributions are incorporated to constrain the joint distribution

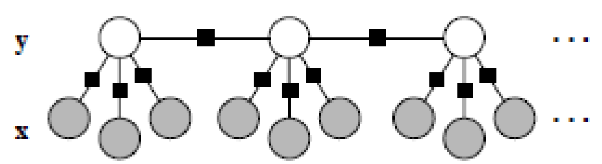
- Modeling joint probability, *P(observations|labels)P(labels),* is only an intermediate step towards the ultimate goal of classification

- Why not use discriminative classifiers directly?

Conditional Random Field[1,2]

- Definition: Let G = (S, E) be a graph such that x is indexed by the vertices of G. Then (x, y) is said to be a conditional random field if, when conditioned on x, the random variables {y} obey the Markov property with respect to the graph.

- Markovianity only after labels have been conditioned on observations in some features
- No Markovianity on label priors
- No generative terms/likelihood



Graphical model of an HMM-like linear-chain CRF.

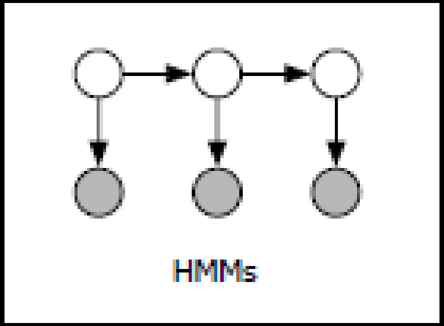
[1] Charles Sutton, Andrew McCallum, An Introduction to Conditional Random Fields for Relational Learning. Introduction to Statistical Relational Learning, MIT Press, 2006.
[2] J. Lafferty, A. McCallum, F. Pereira, Conditional random elds: Probablistic models for segmenting and labeling sequence data.Proc. Int. Conf. on Machinge Learning, 2001.

Conditional Random Field

A discriminative counterpart to generative Hidden Markov Model (HMM)

HMM: Independence assumptions are made for tractability

- Each observation variable depends only on the current state
- Each state depends only on its immediate state



Likelihood/Generative Term

$$p(\mathrm{y}, \mathrm{x}) = \prod_{t=1}^{T} p(y_t | y_{t-1}) p(x_t | y_t)$$

Conditional Random Field

- CRF: Condition HMM on observations (instead of the generative term)

$$p(\mathbf{y},\mathbf{x}) = \prod_{t=1}^{T} p(y_t|y_{t-1})p(x_t|y_t)$$

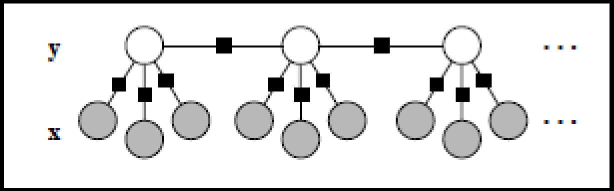
$$p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{y},\mathbf{x})}{\sum_{\mathbf{y}'} p(\mathbf{y}',\mathbf{x})} = \frac{\exp\left\{\sum_{k=1}^{K} \lambda_k f_k(y_t, y_{t-1}, x_t)\right\}}{\sum_{\mathbf{y}'} \exp\left\{\sum_{k=1}^{K} \lambda_k f_k(y'_t, y'_{t-1}, x_t)\right\}}.$$

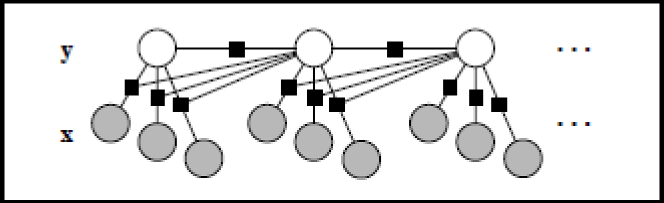
- Introduce feature functions (flexible – may not be log probabilities) and Normalization constant to sum to 1

Conditional Random Field

CRF: Contrary to HMM, independence assumption between observations/hidden variables is relaxed

$$p(\mathrm{y}|\mathrm{x}) = \frac{1}{Z(\mathrm{x})} \exp\left\{ \sum_{k=1}^{K} \lambda_k f_k(y_t, y_{t-1}, \mathrm{x}_t) \right\}$$





Discriminative Random Field (DRF)[1]

An example of CRF

- 2D extension of CRF – discriminate model in 2D

- Globally conditioned on observation data
- Not linear-chain -> has neighborhood relationships among labels

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

[1] S. Kumar and M. Herbert, "Discriminative Fields for Modeling Spatial Dependencies in Natural Images," Proc. 18th Ann. Conf. Neural Information Processing Systems, 2004

DRF for Object Detection

- DRF for detection of Man-made structures in images
  - Features $\boldsymbol{h_i(y)}$ and $\boldsymbol{\mu_{ij}(y)}$ are based on Histograms of gradients



  - *Please look at cited work for MAP optimization algorithms!*

Conclusion

- Take-home Messages ☺
  - Graphical Models
  - conditional independence -> Factorization in graphs -> tractability
  - Generative vs. Discriminative Classifiers/Models
  - Markovianity is more general than characteristic of being generative
  - MRF vs. CRF

Thank You!