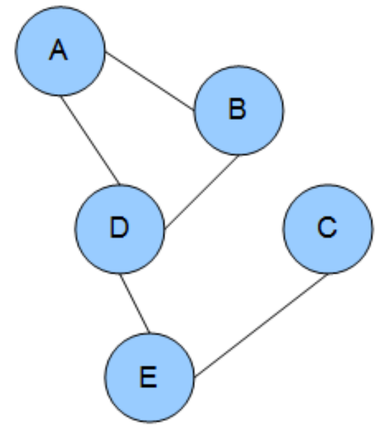
WIKIPEDIA

Markov random field

In the domain of physics and probability, a **Markov random field** (often abbreviated as **MRF**), **Markov network** or **undirected graphical model** is a set of random variables having a Markov property described by an undirected graph. In other words, a random field is said to be a Markov random field if it satisfies Markov properties.



An example of a Markov random field. Each edge represents dependency. In this example: A depends on B and D. B depends on A and D. D depends on A, B, and E. E depends on D and C. C depends on E.

A Markov network or MRF is similar to a Bayesian network in its representation of dependencies; the differences being that Bayesian networks are directed and acyclic, whereas Markov networks are undirected and may be cyclic. Thus, a Markov network can represent certain dependencies that a Bayesian network cannot (such as cyclic dependencies); on the other hand, it can't represent certain dependencies that a Bayesian network can (such as induced dependencies). The underlying graph of a Markov random field may be finite or infinite.

When the joint probability density of the random variables is strictly positive, it is also referred to as a **Gibbs random field**, because, according to the Hammersley–Clifford theorem, it can then be represented by a Gibbs measure for an appropriate (locally defined) energy function. The prototypical Markov random field is the Ising model; indeed, the Markov random field was introduced as the general setting for the Ising model.[1] In the domain of artificial intelligence, a Markov random field is used to model various low- to mid-level tasks in image processing and computer vision.[2]

Contents

Definition

Given an undirected graph $G=(V,E)$, a set of random variables $X=(X_v)_{v\in V}$ indexed by V form a Markov random field with respect to G if they satisfy the local Markov properties:

> **Pairwise Markov property**: Any two non-adjacent variables are conditionally independent given all other variables:
>
> $$X_u \perp\!\!\!\perp X_v \mid X_{V\setminus\{u,v\}}$$
>
> **Local Markov property**: A variable is conditionally independent of all other variables given its neighbors:
>
> $$X_v \perp\!\!\!\perp X_{V\setminus \mathrm{N}[v]} \mid X_{\mathrm{N}(v)}$$
>
> where $\mathrm{N}(v)$ is the set of neighbors of v, and $\mathrm{N}[v]=v\cup\mathrm{N}(v)$ is the closed neighbourhood of v.
>
> **Global Markov property**: Any two subsets of variables are conditionally independent given a separating subset:
>
> $$X_A \perp\!\!\!\perp X_B \mid X_S$$
>
> where every path from a node in A to a node in B passes through S.

The Global Markov property is stronger than the Local Markov property, which in turn is stronger than the Pairwise one.[3] However, the above three Markov properties are equivalent for a positive probability.[4]

Clique factorization

As the Markov property of an arbitrary probability distribution can be difficult to establish, a commonly used class of Markov random fields are those that can be factorized according to the cliques of the graph.

Given a set of random variables $X=(X_v)_{v\in V}$, let $P(X=x)$ be the probability of a particular field configuration x in X. That is, $P(X=x)$ is the probability of finding that the random variables X take on the particular value x. Because X is a set, the probability of x should be understood to be taken with respect to a *joint distribution* of the X_v.

If this joint density can be factorized over the cliques of G:

$$P(X=x)=\prod_{C\in \operatorname{cl}(G)} \phi_C(x_C)$$

then X forms a Markov random field with respect to G. Here, $\operatorname{cl}(G)$ is the set of cliques of G. The definition is equivalent if only maximal cliques are used. The functions φ_C are sometimes referred to as *factor potentials* or *clique potentials*. Note, however, conflicting terminology is in use: the word *potential* is often applied to the logarithm of φ_C. This is because, in statistical mechanics, $\log(\varphi_C)$ has a direct interpretation as the potential energy of a configuration x_C.

Some MRF's do not factorize: a simple example can be constructed on a cycle of 4 nodes with some infinite energies, i.e. configurations of zero probabilities,[5] even if one, more appropriately, allows the infinite energies to act on the complete graph on V.[6]

MRF's factorize if at least one of the following conditions is fulfilled:

- the density is positive (by the Hammersley–Clifford theorem)
- the graph is chordal (by equivalence to a Bayesian network)

When such a factorization does exist, it is possible to construct a factor graph for the network.

Exponential family

Any Markov random field can be written as exponential family in canonical form with feature functions f_k such that the full-joint distribution can be written as

$$P(X=x) = \frac{1}{Z}\exp\left(\sum_k w_k^\top f_k(x_{\{k\}})\right)$$

where the notation

$$w_k^\top f_k(x_{\{k\}}) = \sum_{i=1}^{N_k} w_{k,i} \cdot f_{k,i}(x_{\{k\}})$$

is simply a dot product over field configurations, and Z is the partition function:

$$Z = \sum_{x \in \mathcal{X}} \exp\left(\sum_k w_k^\top f_k(x_{\{k\}})\right).$$

Here, $\mathcal{X}$ denotes the set of all possible assignments of values to all the network's random variables. Usually, the feature functions $f_{k,i}$ are defined such that they are indicators of the clique's configuration, *i.e.* $f_{k,i}(x_{\{k\}}) = 1$ if $x_{\{k\}}$ corresponds to the i-th possible configuration of the k-th clique and 0 otherwise. This model is equivalent to the clique factorization model given above, if $N_k = |\operatorname{dom}(C_k)|$ is the cardinality of the clique, and the weight of a feature $f_{k,i}$ corresponds to the logarithm of the corresponding clique factor, *i.e.* $w_{k,i} = \log \phi(c_{k,i})$, where $c_{k,i}$ is the i-th possible configuration of the k-th clique, *i.e.* the i-th value in the domain of the clique C_k.

The probability P is often called the Gibbs measure. This expression of a Markov field as a logistic model is only possible if all clique factors are non-zero, *i.e.* if none of the elements of $\mathcal{X}$ are assigned a probability of 0. This allows techniques from matrix algebra to be applied, *e.g.* that the trace of a matrix is log of the determinant, with the matrix representation of a graph arising from the graph's incidence matrix.

The importance of the partition function Z is that many concepts from statistical mechanics, such as entropy, directly generalize to the case of Markov networks, and an *intuitive* understanding can thereby be gained. In addition, the partition function allows variational methods to be applied to the solution of the problem: one can attach a driving force to one or more of the random variables, and explore the reaction of the network in response to this perturbation. Thus, for example, one may add a driving term J_v, for each vertex v of the graph, to the partition function to get:

$$Z[J] = \sum_{x \in \mathcal{X}} \exp\left(\sum_k w_k^\top f_k(x_{\{k\}}) + \sum_v J_v x_v\right)$$

Formally differentiating with respect to J_v gives the expectation value of the random variable X_v associated with the vertex v:

$$E[X_v] = \frac{1}{Z} \frac{\partial Z[J]}{\partial J_v}\bigg|_{J_v=0}.$$

Correlation functions are computed likewise; the two-point correlation is:

$$C[X_u, X_v] = \frac{1}{Z} \frac{\partial^2 Z[J]}{\partial J_u \partial J_v}\bigg|_{J_u=0, J_v=0}.$$

Unfortunately, though the likelihood of a logistic Markov network is convex, evaluating the likelihood or gradient of the likelihood of a model requires inference in the model, which is generally computationally infeasible (see 'Inference' below).

Examples

Gaussian

A multivariate normal distribution forms a Markov random field with respect to a graph $G = (V, E)$ if the missing edges correspond to zeros on the precision matrix (the inverse covariance matrix):

$$X = (X_v)_{v \in V} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

such that

$$(\Sigma^{-1})_{uv} = 0 \quad \text{iff} \quad \{u, v\} \notin E.$$ [7]

Inference

As in a Bayesian network, one may calculate the conditional distribution of a set of nodes $V' = \{v_1, \ldots, v_i\}$ given values to another set of nodes $W' = \{w_1, \ldots, w_j\}$ in the Markov random field by summing over all possible assignments to $u \notin V', W'$; this is called exact inference. However, exact inference is a #P-complete problem, and thus computationally intractable in the general case. Approximation techniques such as Markov chain Monte Carlo and loopy belief propagation are often more feasible in practice. Some particular subclasses of MRFs, such as trees (see Chow–Liu tree), have polynomial-time inference algorithms; discovering such subclasses is an active research topic. There are also subclasses of

MRFs that permit efficient MAP, or most likely assignment, inference; examples of these include associative networks.[8][9] Another interesting sub-class is the one of decomposable models (when the graph is chordal): having a closed-form for the MLE, it is possible to discover a consistent structure for hundreds of variables.[10]

Conditional random fields

One notable variant of a Markov random field is a **conditional random field**, in which each random variable may also be conditioned upon a set of global observations o. In this model, each function ϕ_k is a mapping from all assignments to both the clique k and the observations o to the nonnegative real numbers. This form of the Markov network may be more appropriate for producing discriminative classifiers, which do not model the distribution over the observations. CRFs were proposed by John D. Lafferty, Andrew McCallum and Fernando C.N. Pereira in 2001.[11]

Varied applications

Markov random fields find application in a variety of fields, ranging from computer graphics to computer vision and machine learning.[12] MRFs are used in image processing to generate textures as they can be used to generate flexible and stochastic image models. In image modelling, the task is to find a suitable intensity distribution of a given image, where suitability depends on the kind of task and MRFs are flexible enough to be used for image and texture synthesis, image compression and restoration, image segmentation, 3D image inference from 2D images, image registration, texture synthesis, super-resolution, stereo matching and information retrieval. They can be used to solve various computer vision problems which can be posed as energy minimization problems or problems where different regions have to be distinguished using a set of discriminating features, within a Markov random field framework, to predict the category of the region.[13] Markov random fields were a generalization over the Ising model and have, since then, been used widely in combinatorial optimizations and networks.

See also

- Constraint Composite Graph
- Graphical model
- Hammersley–Clifford theorem
- Hopfield network
- Interacting particle system
- Ising model
- Log-linear analysis
- Markov chain
- Markov logic network
- Maximum entropy method
- Probabilistic cellular automata

References

1. Kindermann, Ross; Snell, J. Laurie (1980). *Markov Random Fields and Their Applications* (http://www.cmap.polytechnique.fr/~rama/ehess/mrfbook.pdf) (PDF). American Mathematical Society. ISBN 978-0-8218-5001-5. MR 0620955 (https://www.ams.org/mathscinet-getitem?mr=0620955).

2. Li, S. Z. (2009). *Markov Random Field Modeling in Image Analysis* (https://books.google.com/books?id=rDsObhDkCIAC&printsec=frontcover#v=onepage&q&f=false). Springer. ISBN 9781848002791.
3. Lauritzen, Steffen (1996). *Graphical models*. Oxford: Clarendon Press. p. 33. ISBN 978-0198522195.
4. *Probabilistic Graphical Models*.
5. Moussouris, John (1974). "Gibbs and Markov random systems with constraints". *Journal of Statistical Physics*. **10** (1): 11–33. doi:10.1007/BF01011714 (https://doi.org/10.1007%2FBF01011714). hdl:10338.dmlcz/135184 (https://hdl.handle.net/10338.dmlcz%2F135184). MR 0432132 (https://www.ams.org/mathscinet-getitem?mr=0432132).
6. Gandolfi, Alberto; Lenarda, Pietro (2016). "A note on Gibbs and Markov Random Fields with constraints and their moments". *Mathematics and Mechanics of Complex Systems*. **4** (3–4): 407–422. doi:10.2140/memocs.2016.4.407 (https://doi.org/10.2140%2Fmemocs.2016.4.407).
7. Rue, Håvard; Held, Leonhard (2005). *Gaussian Markov random fields: theory and applications*. CRC Press. ISBN 978-1-58488-432-3.
8. Taskar, Benjamin; Chatalbashev, Vassil; Koller, Daphne (2004), "Learning associative Markov networks", in Brodley, Carla E. (ed.), *Proceedings of the Twenty-first International Conference on Machine Learning (ICML 2004), Banff, Alberta, Canada, July 4-8, 2004*, ACM International Conference Proceeding Series, **69**, Association for Computing Machinery, p. 102, CiteSeerX 10.1.1.157.329 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.157.329), doi:10.1145/1015330.1015444 (https://doi.org/10.1145%2F1015330.1015444), ISBN 978-1581138283.
9. Duchi, John C.; Tarlow, Daniel; Elidan, Gal; Koller, Daphne (2006), "Using Combinatorial Optimization within Max-Product Belief Propagation" (http://papers.nips.cc/paper/3117-using-combinatorial-optimization-within-max-product-belief-propagation), in Schölkopf, Bernhard; Platt, John C.; Hoffman, Thomas (eds.), *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006*, Advances in Neural Information Processing Systems, **19**, MIT Press, pp. 369–376.
10. Petitjean, F.; Webb, G.I.; Nicholson, A.E. (2013). *Scaling log-linear analysis to high-dimensional data* (http://www.tiny-clues.eu/Research/Petitjean2013-ICDM.pdf) (PDF). International Conference on Data Mining. Dallas, TX, USA: IEEE.
11. "Two classic paper prizes for papers that appeared at ICML 2013" (http://icml.cc/2013/?page_id=21). *ICML*. 2013. Retrieved 15 December 2014.
12. Kindermann & Snell, Ross & Laurie (1980). *Markov Random Fields and their Applications*. Rhode Island: American Mathematical Society. ISBN 978-0-8218-5001-5.
13. Zhang & Zakhor, Richard & Avideh (2014). "Automatic Identification of Window Regions on Indoor Point Clouds Using LiDAR and Cameras". *VIP Lab Publications*. CiteSeerX 10.1.1.649.303 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.649.303).

External links

- MRF implementation in C++ for regular 2D lattices (https://bitbucket.org/rukletsov/b)