

Title: Conditional random field

URL: [https://en.wikipedia.org/wiki/Conditional\\_random\\_field](https://en.wikipedia.org/wiki/Conditional_random_field)

PageID: 4118276

Categories: Category:Graphical models, Category:Machine learning

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

-----

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning

Decision trees

Ensembles Bagging Boosting Random forest

Bagging

Boosting

Random forest

k -NN

Linear regression

Naive Bayes

Artificial neural networks

Logistic regression

Perceptron

Relevance vector machine (RVM)

Support vector machine (SVM)

BIRCH

CURE

Hierarchical

k -means

Fuzzy

Expectation–maximization (EM)

DBSCAN

OPTICS

Mean shift

Factor analysis

CCA

ICA

LDA

NMF

PCA

PGD

t-SNE

SDL

Graphical models Bayes net Conditional random field Hidden Markov

Bayes net

Conditional random field

Hidden Markov

RANSAC

k -NN

Local outlier factor  
Isolation forest  
Autoencoder  
Deep learning  
Feedforward neural network  
Recurrent neural network LSTM GRU ESN reservoir computing  
LSTM  
GRU  
ESN  
reservoir computing  
Boltzmann machine Restricted  
Restricted  
GAN  
Diffusion model  
SOM  
Convolutional neural network U-Net LeNet AlexNet DeepDream  
U-Net  
LeNet  
AlexNet  
DeepDream  
Neural field Neural radiance field Physics-informed neural networks  
Neural radiance field  
Physics-informed neural networks  
Transformer Vision  
Vision  
Mamba  
Spiking neural network  
Memtransistor  
Electrochemical RAM (ECRAM)  
Q-learning  
Policy gradient  
SARSA  
Temporal difference (TD)  
Multi-agent Self-play  
Self-play  
Active learning  
Crowdsourcing  
Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

Conditional random fields ( CRFs ) are a class of statistical modeling methods often applied in pattern recognition and machine learning and used for structured prediction . Whereas a classifier predicts a label for a single sample without considering "neighbouring" samples, a CRF can take context into account. To do so, the predictions are modelled as a graphical model , which represents the presence of dependencies between the predictions. The kind of graph used depends on the application. For example, in natural language processing , "linear chain" CRFs are popular, for which each prediction is dependent only on its immediate neighbours. In image processing, the graph typically connects locations to nearby and/or similar locations to enforce that they receive similar predictions.

Other examples where CRFs are used are: labeling or parsing of sequential data for natural language processing or biological sequences , part-of-speech tagging , shallow parsing , named entity recognition , gene finding , peptide critical functional region finding, and object recognition and image segmentation in computer vision .

#### Description

CRFs are a type of discriminative undirected probabilistic graphical model .

Lafferty , McCallum and Pereira define a CRF on observations  $X$  and random variables  $Y$  as follows:

Let  $G = (V, E)$  be a graph such that  $Y = (Y_v)_{v \in V}$ , so that  $Y$  is indexed by the vertices of  $G$  .

Then  $(X, Y)$  is a conditional random field when each random variable  $Y_v$ , conditioned on  $X$ , obeys the Markov property with respect to the graph; that is, its probability is dependent only on its neighbours in  $G$  and not its past states:

$$P(Y_v | X, \{Y_w : w \neq v\}) = P(Y_v | X, \{Y_w : w \sim v\})$$

that  $w$  and  $v$  are neighbors in  $G$  .

What this means is that a CRF is an undirected graphical model whose nodes can be divided into exactly two disjoint sets  $X$  and  $Y$ , the observed and output variables, respectively; the conditional distribution  $p(Y | X)$  is then modeled.

#### Inference

For general graphs, the problem of exact inference in CRFs is intractable. The inference problem for a CRF is basically the same as for an MRF and the same arguments hold. However, there exist special cases for which exact inference is feasible:

If the graph is a chain or a tree, message passing algorithms yield exact solutions. The algorithms used in these cases are analogous to the forward-backward and Viterbi algorithm for the case of HMMs.

If the CRF only contains pair-wise potentials and the energy is submodular , combinatorial min cut/max flow algorithms yield exact solutions.

If exact inference is impossible, several algorithms can be used to obtain approximate solutions. These include:

Loopy belief propagation

Alpha expansion

Mean field inference

Linear programming relaxations

Parameter learning

Learning the parameters  $\theta$  is usually done by maximum likelihood learning for  $p(Y_i | X_i; \theta)$ . If all nodes have exponential family distributions and all nodes are observed during training, this optimization is convex. It can be solved for example using gradient descent algorithms, or Quasi-Newton methods such as the L-BFGS algorithm. On the other hand, if some variables are unobserved, the inference problem has to be solved for these variables. Exact inference is intractable in general graphs, so approximations have to be used.

#### Examples

In sequence modeling, the graph of interest is usually a chain graph. An input sequence of observed variables  $X$  represents a sequence of observations and  $Y$  represents a hidden (or unknown) state variable that needs to be inferred given the observations. The  $Y_i$  are structured to form a chain, with an edge between each  $Y_{i-1}$  and  $Y_i$ . As well as having a simple interpretation of the  $Y_i$  as "labels" for each element in the input sequence, this layout admits efficient algorithms for:

model training, learning the conditional distributions between the  $Y_i$  and feature functions from some corpus of training data.

decoding, determining the probability of a given label sequence  $Y$  given  $X$ .

inference, determining the most likely label sequence  $Y$  given  $X$ .

The conditional dependency of each  $Y_i$  on  $X$  is defined through a fixed set of feature functions of the form  $f(i, Y_{i-1}, Y_i, X)$ , which can be thought of as measurements on the input sequence that partially determine the likelihood of each possible value for  $Y_i$ . The model assigns each feature a numerical weight and combines them to determine the probability of a certain value for  $Y_i$ .

Linear-chain CRFs have many of the same applications as conceptually simpler hidden Markov models (HMMs), but relax certain assumptions about the input and output sequence distributions. An HMM can loosely be understood as a CRF with very specific feature functions that use constant probabilities to model state transitions and emissions. Conversely, a CRF can loosely be understood as a generalization of an HMM that makes the constant transition probabilities into arbitrary functions that vary across the positions in the sequence of hidden states, depending on the input sequence.

Notably, in contrast to HMMs, CRFs can contain any number of feature functions, the feature functions can inspect the entire input sequence  $X$  at any point during inference, and the range of the feature functions need not have a probabilistic interpretation.

Variants

Higher-order CRFs and semi-Markov CRFs

CRFs can be extended into higher order models by making each  $Y_i$  dependent on a fixed number  $k$  of previous variables  $Y_{i-k}, \dots, Y_{i-1}$ . In conventional formulations of higher order CRFs, training and inference are only practical for small values of  $k$  (such as  $k \leq 5$ ), since their computational cost increases exponentially with  $k$ .

However, another recent advance has managed to ameliorate these issues by leveraging concepts and tools from the field of Bayesian nonparametrics. Specifically, the CRF-infinity approach constitutes a CRF-type model that is capable of learning infinitely-long temporal dynamics in a scalable fashion. This is effected by introducing a novel potential function for CRFs that is based on the Sequence Memoizer (SM), a nonparametric Bayesian model for learning infinitely-long dynamics in sequential observations. To render such a model computationally tractable, CRF-infinity employs a mean-field approximation of the postulated novel potential functions (which are driven by an SM). This allows for devising efficient approximate training and inference algorithms for the model, without undermining its capability to capture and model temporal dependencies of arbitrary length.

There exists another generalization of CRFs, the semi-Markov conditional random field (semi-CRF), which models variable-length segmentations of the label sequence  $Y$ . This provides much of the power of higher-order CRFs to model long-range dependencies of the  $Y_i$ , at a reasonable computational cost.

Finally, large-margin models for structured prediction, such as the structured Support Vector Machine can be seen as an alternative training procedure to CRFs.

## Latent-dynamic conditional random field

Latent-dynamic conditional random fields ( LDCRF ) or discriminative probabilistic latent variable models ( DPLVM ) are a type of CRFs for sequence tagging tasks. They are latent variable models that are trained discriminatively.

In an LDCRF, like in any sequence tagging task, given a sequence of observations  $x = x_1, \dots, x_n$ , the main problem the model must solve is how to assign a sequence of labels  $y = y_1, \dots, y_n$  from one finite set of labels  $Y$ . Instead of directly modeling  $P(y | x)$  as an ordinary linear-chain CRF would do, a set of latent variables  $h$  is "inserted" between  $x$  and  $y$  using the chain rule of probability :

This allows capturing latent structure between the observations and labels. While LDCRFs can be trained using quasi-Newton methods, a specialized version of the perceptron algorithm called the latent-variable perceptron has been developed for them as well, based on Collins' structured perceptron algorithm. These models find applications in computer vision , specifically gesture recognition from video streams and shallow parsing .

See also

Hammersley–Clifford theorem

Maximum entropy Markov model (MEMM)

## References

### Further reading

McCallum, A.: Efficiently inducing features of conditional random fields . In: Proc. 19th Conference on Uncertainty in Artificial Intelligence . (2003)

Wallach, H.M. : Conditional random fields: An introduction . Technical report MS-CIS-04-21, University of Pennsylvania (2004)

Sutton, C., McCallum, A.: An Introduction to Conditional Random Fields for Relational Learning. In "Introduction to Statistical Relational Learning". Edited by Lise Getoor and Ben Taskar. MIT Press. (2006) Online PDF

Klinger, R., Tomanek, K.: Classical Probabilistic Models and Conditional Random Fields. Algorithm Engineering Report TR07-2-013, Department of Computer Science, Dortmund University of Technology, December 2007. ISSN 1864-4503. Online PDF