This paper proposes conditional random fields (CNF), a probabilistic model for segmenting and labeling sequence data. For above tasks, Generative models (such as hidden Markov models and stochastic grammars) is limited to represent multiple interacting features or long-range dependencies of the observations. Conditional models (such as maximum entropy Markov models) work as alternatives, where the probability of a transition between labels depend on current, past and future observations. However, they usually confront the label bias problem (the transitions leaving a given state complete only against each other instead of against all other transitions in the model). There are at least two solutions: to change the state-transition structure of the model, and to start a with a fully-connected model and let the training procedure figure out a good structure. However, they bring new problems. In contrast, the proposed CNF model solves this problem in a principled way, which allows some transitions “vote” more strongly than others depending on the corresponding observations.

CNFs could be viewed as a finite state model with unnormalized transition probabilities. It assigns a well-defined probability distribution over possible labeling, trained by maximum likelihood or MAP estimation. Additionally, the convex loss function ensures convergence to global optimum. The parameter estimation problem is to determine the parameters from training data with empirical distribution over “random variable over data sequences to be labeled” and “random variable over corresponding label sequences”. And the log-likelihood objective function could be maximized via iterative scaling algorithms (Algorithm S using “slack feature” and Algorithm T keeping tracking of partial “total feature count” totals, both are based on the improved iterative scaling algorithm). Although CNF encompasses HMM-like models, it is more expressive because it allows arbitrary dependencies on the observation sequence.

The authors conducted several experiments. The first one verifies that CRFs solves the classical version of the label bias problem compared to MEMMs; the second one shows that CRFs give better performances than HMMs and MEMMs when true data distribution has higher-order dependencies than the model; the third one confirms above mentioned observations by comparing different models on a part-of-speech tagging task.