None

-----------------------

Many tasks in computer vision can be framed as mak- ing predictions about complex, structured objects. For ex- ample, image labeling problems like stereo depth estima- tion, optical flow, and image segmentation can all be cast as making predictions jointly over many correlated outputs. The modeling frameworks that have found the most success for this type of problems are those like Conditional Ran- dom Fields (CRFs) and Structural Support Vector Machines (SSVMs), which explicitly model the correlations over the outputs and make test-time predictions by either exactly or approximately solving a joint inference task. These formu- lations are collectively known as structured output learning, or structured prediction, and are the focus of this work.

A key research issue that arises when working with struc- tured output problems is how to best tradeoff expressivity of the model with the ability to efficiently learn and per- form inference (make predictions). Traditionally, these con- cerns have led to the use of overly simplistic models over la- belings that make unrealistic conditional independence as- sumptions, such as pairwise models with grid-structured topology. Recently, there have been successful efforts that weaken these assumptions, either by moving to densely connected pairwise models [13] or by enforcing smooth- ness in higher order neighborhoods [10]. However, while these approaches can lead to improved performance, they do not capture much higher level structure in the data, such as information about shape. As we look to build models that more faithfully represent structure present in the world, it is desirable to explore the use of models capable of represent- ing this higher level structure.

++++++++++++++++++

One interesting task in machine learning is labeling over complex and structured objects. Many applications such as image segmentation, motif finding and noun-phrase parsing involved with representing jointly correlated variables. Encoding consistency constraints over large number of cliques, for example, is central to the problem of image segmentation. Algorithm frameworks like Markov Random Field (MRF) containing higher order energy functions and max margin method for Structural SVM are raising interests recently due to their capability of representing structural dependencies of variables and ensuring computationally feasible approximate or exact inference.

Higher order potentials can be optimized efficiently using modern Linear Programming frameworks. For example, Kohli and Kumar\cite{kohli2009robust} developed lower linear envelope potentials by introducing auxiliary variables and reduce the potential function to a pairwise formulation. Thus the formulation can be optimized using LP programming. However, they only provide an approximate routine for the problem.\\ Another powerful algorithm called “graph-cuts” which can do exact inference on binary MRFs with internal submodular potentials was well studied in past years. Gould\cite{gouldlearning} extended the previous research via graph-cuts to do an exact inference using weighted lower linear envelope potentials and optimized the model parameters within a max-margin learning framework. However, the approximation used in their work leading to a fixed spaced break-points lower linear envelope. If the equally spaced constraint can be removed, their formulation will result in a latent SVM.\\

Intro&rela

-----------------------

Probabilistic graphical models are widely used in ma- chine learning due to their representational power and the existence of efficient algorithms for inference and learning. Typically, however, the model structure must be restricted to ensure tractability. To enable efficient exact inference, the most common restriction is that the model have low tree-width.\\

A natural question to ask is if there are other, differ- ent restrictions that we can place on models to ensure tractable exact or approximate inference. Indeed, a celebrated result is the ability of the “graph cuts” algo- rithm to exactly find the maximum a posteriori (MAP) assignment in any pairwise graphical model with bi- nary variables, where the internal potential structure is restricted to be submodular. Along similar lines, polynomial-time algorithms can exactly compute the partition function in an Ising model if the underlying graph is planar (Fisher, 1961).\\

Extensions of these results have been a topic of much recent interest, particularly for the case of MAP in- ference. Gould (2011) shows how to do exact MAP inference in models with certain higher order terms via graph cut-like algorithms, and Ramalingham et al. (2008) give results for multilabel submodular models. Tarlow et al. (2010) provide efficient algorithms for a number of other high-order potentials.\\

1realatedwork-gould

-------------------

Models containing higher order factors are able to encode complex dependencies between groups of vari- ables, and can encourage solutions which match the statistics of the ground truth solution (Potetz, 2007; Roth \& Black, 2005; Woodford et al., 2009). However, the high computational cost of performing MAP infer- ence in such models has inhibited their use (Lan et al., 2006). Instead, there has been a widespread adoption of the simpler and less powerful pairwise-CRF models which allow efficient inference (Szeliski et al., 2008).\\

While learning the parameters of models with insuffi- cient expressivity, researchers can penalize certain mis- representations of the solution space using a ‘loss func- tion’ which specifies the deviations from ground truth that the learning algorithm should avoid (Tsochan- taridis et al., 2005; Taskar et al., 2003). Most previ- ous works on these topics have used simple choices of the loss function, such as the Hamming loss or squared loss, which lead to tractable learning algo- rithms (Szummer et al., 2008). However, in real world applications, researchers might prefer more general loss functions which penalize deviations in some higher order statistics.