Research Statement

Sumit Chopra
Courant Institute of Mathematical Sciences
New York University
sumit@cs.nyu.edu

I am interested in Machine Learning and Pattern Recognition. My research focuses on understanding the rich contextual information associated with most data sets in a variety of real world domains and using it to infer complex patterns. To this end, I have worked on designing efficient inference and learning algorithms for models capable of handling the uncertainties and interdependencies among samples associated with large scale data sets. The class of models that I have explored is highly diverse, including Energy Based Models, Graphical Models, Deep Learning Architectures, and Relational Graphical Models. Furthermore, I have applied these models to real world problems in a variety of domains, such as in economics for predicting the real estate prices, in computer vision for object recognition, face verification, distance metric learning and data visualization, and in robotics for autonomous navigation.

The importance of contextual information has been well acknowledged in the fields of human and machine vision. For instance, while detecting a car in an image, it is beneficial to use the knowledge that the car is usually parked on the road and sky is always above the ground. Context is equally essential for reasoning in a variety of other domains. Consider one such example where our approach improves substantially over the existing ones: real estate price prediction. The price of a house clearly depends on its individual features, such as the number of bedrooms – the intrinsic price. However, the price also depends on other factors, such as the quality or the desirability of the neighboring houses (the spatial context), and inflation (the temporal context). We capture this spatial context by learning a smooth latent "desirability" manifold over the neighborhood. The value at any point on the manifold can be interpreted as the desirability of that location. When it is combined with the intrinsic price, we get the actual price estimate of the house. The details of this application and the related general framework are discussed below. More generally, in such domains the notion of context is abstracted by the complex interdependencies among the samples. These interactions between samples often exhibit an underlying logical structure. I aim to capture this idea of context in different domains, to enable robust reasoning about a diverse set of complex real world phenomenon.

Another dimension of my research involves answering questions which go beyond the simple "yes/no" questions (classification), or go beyond predicting the value of a function using only the individual features of an input sample (regression). For example, in addition to knowing whether a car is present in an image, it would be interesting to infer other objects and their respective positions. Similarly, instead of just predicting the price of a house, understanding how the economic properties of the entire neighborhood will change over time will be far more useful.

Factor Graphs for Relational Regression

Relational dependencies among samples is an example of structure that one commonly encounters in real world data. Prediction not only depends on a sample's individual features, but also on the features of other related samples. Furthermore, the samples may be related via hidden (latent) features, which are not directly measurable but are only implicit in the data. Intuitively, we would like to use the information associated with one sample to help us reach conclusions about other related samples. Though this area has seen a lot of recent activity [4], most efforts have been directed toward solving relational classification problems. We have developed a general framework that can account for continuous variables and, hence, can be applied even to regression in a relational setting [3].

The idea is to use a single factor graph to model the entire collection of samples. The relationship between samples is captured by a set of factors; a factor measures the compatibility between its input variables by assigning a scalar energy (score). Learning involves adjusting the parameters of the factors so as to minimize the sum of energies of all the factors. The large size and complicated topology of the graph makes the learning and inference tasks challenging. We propose a novel way of designing the factors which leads to an efficient inference and learning algorithm. It reduces the learning process to a simple back-propagation of errors. The framework is novel in a number of ways. Using a non-exponential family of functions in the parameter space allows us to efficiently capture highly complex dependencies among samples. The framework is also capable of capturing latent dependencies that are implicit in the data. The importance of this feature cannot be emphasized enough because a number of data sets exhibit such a characteristic. Predicting the prices of real estate properties is one such important problem.

Real Estate Price Prediction

The price of a house, in addition to being dependent on its individual features, is also influenced by the quality of its neighborhood. Some neighborhood features are directly measurable, such as the median household income. However, most features are hidden and are implicitly reflected in the prices of other houses. These features must be estimated during training.

We applied the framework of Relational Regression using Factor Graphs to solve this problem [2]. In order to capture the types of price dependencies above, each house is assigned two factors. The goal is to simultaneously learn the house-specific dependence on the price and a hidden "desirability" manifold that spans the entire geographic area. The manifold implicitly captures the effect of neighboring houses on the current house. Any point on this manifold can be viewed as a measure of how desirable that particular location is (a function of the price of neighboring houses). The model was tested on real dataset of transactions in the year 2004 for Los Angeles County. The prediction accuracy was superior to all other models traditionally used for this problem. Furthermore, it was used to answer more engaging questions, such as whether a particular house in an area is over priced or under priced (a typical buyer's dilemma). Or, whether adding another room to a house would increase its value appreciably or not (a seller's dilemma).

Automatically Learning a Distance Metric

Knowledge of a distance metric over the data which is capable of capturing the complex dependencies exhibited by it, can be seen as a first step in designing algorithms for more advanced tasks such as information retrieval, contextual classification, and data visualization. Often, however, it is very difficult to intuitively choose an appropriate metric capable of encoding all possible relationships in the data. Samples may be similar in highly complex ways: two images of the same car may be very different if the pictures were taken under different lighting conditions and in different positions. A better strategy would be to learn the metric automatically by looking at the data.

We propose a framework for this task that does not presuppose the existence of a computable metric in the input space, such as the Euclidean distance. Given a set of input samples and a neighborhood graph, the idea is to learn a parametric function that maps the high dimensional input points to a low dimensional output manifold, so that similar points in the input space are mapped close to one another in the output space, and dissimilar points are mapped far away from one another. With no restriction on the architecture of the function, one can capture even complex nonlinear similarities between samples. This method was applied to a number of problems in computer vision, most notably face verification; given a pair of facial images, identify whether or not they belong to the same person [5, 1]. The learned function is capable of correctly labeling the image pairs even if the subjects have very different facial expressions, or even if the faces are partially occluded by artifacts, like sunglasses or a face scarf.

Energy Based Models

My interest in exploiting the contextual information associated with data sets led me towards working with Energy Based Models (EBMs). EBMs can be seen as a natural platform where one can easily design efficient inference and learning algorithms for large scale data sets. Furthermore, the absence of a normalization requirement gives a lot of freedom in designing powerful architectures capable of capturing complex dependencies inherent in the data. I have worked on a number of facets of the EBM framework, both from the theoretical and application perspective.

We have characterized loss functions whose minimization would result in successful and efficient training of the model. This is motivated by the fact that the absence of normalization renders some loss functions useless for the purpose of training in EBMs [6]. Another important issue is the design of a learning algorithm for a system involving latent variables. The EM algorithm in probabilistic models maximizes the log likelihood of the training data. We are working on designing a *deterministic EM* algorithm, which will guarantee the minimization of general loss functions in EBMs. In line with my research goals, I have also worked with advanced connectionist architectures like Deep Belief Networks, Autoencoders, and Convolutional Networks. My primary focus has been on designing unsupervised algorithms for training deep architectures to generate sparse representations of the data. We have applied the algorithm to a number of vision related tasks, such as handwritten digit recognition (for which we hold the current record) and image denoising [7, 8, 9].

Research Agenda

I am currently working towards further exploring the class of latent variable models. This involves extending the Relational Factor Graph framework to capture the hidden temporal dependencies in real estate data, and to problems in social network data, such as slashdot.org. The huge size and elaborate link structure of such data sets is particularly challenging. Their ubiquity, along with the open ended nature of the associated problems, drives me towards exploring them and inferring rich patterns.

My long term research agenda is to bring us closer to the capability of better explaining the intricacies lying behind real world phenomena. In this quest I plan to continue focusing on the design and application of models that are capable of both, compactly representing the relationships between samples and capturing the complex invariance associated with individual samples in data sets belonging to a wide spectrum of domains.

References

[1] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proc. of Computer Vision and Pattern Recognition Conference*. IEEE Press, 2005.

[2] Sumit Chopra, Trivikraman Thampy, John Leahy, Andrew Caplin, and Yann LeCun. Discovering the hidden structure of house prices with a non-parametric manifold model. *13th International Conference on Knowledge Discovery and Data Mining (KDD)*, August 2007.

[3] Sumit Chopra, Trivikraman Thampy, John Leahy, Andrew Caplin, and Yann LeCun. Factor graphs for relational regression. *Technical Report: TR2007-906*, January 2007.

[4] Lise Getoor and Ben Taskar, editors. *Introduction to Statistical Relational Learning.* The MIT Press, Cambridge Massachusetts, 2007.

[5] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *Proc. Computer Vision and Pattern Recognition Conference (CVPR'06)*. IEEE Press, 2006.

[6] Yann LeCun, Sumit Chopra, Raia Hadsell, Ranzato Marc'Aurelio, and Fu-Jie Huang. A tutorial on energy-based learning. In G. Bakir, T. Hofman, B. Schölkopf, A. Smola, and B. Taskar, editors, *Predicting Structured Data*. MIT Press, 2006.

[7] Yann LeCun, Sumit Chopra, Marc'Aurelio Ranzato, and Fu-Jie Huang. Energy based models in document recognition and computer vision. *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, 2007.

[8] Marc'Aurelio Ranzato, Y-Lan Boureau, Sumit Chopra, and Yann LeCun. A unified energy-based framework for unsupervised learning. In *Proc. Conference on AI and Statistics (AI-Stats)*, 2007.

[9] Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra, and Yann LeCun. Efficient learning of sparse representations with an energy-based model. In J. Platt et al., editor, *Advances in Neural Information Processing Systems (NIPS 2006)*. MIT Press, 2006.