Abstract

In this thesis, we consider the problem of finding underlying true states of events (e.g., articles, movies, and signals) using biased observations from agents (e.g., sensors and individuals). We discuss the following three cases of the problem: each event is a signal evolving over time, agents are nodes of a social network, and each event is a graph signal.

In the case where each event is a signal evolving over time, we consider the application of blind calibration of a sensor (agent) network where the sensor gains, offsets and the ground truth values of signals (events) are estimated from noisy observations. This is in general a non-identifiable problem, unless restrictive assumptions on the signal subspace or sensor observations are imposed. We show that if each signal observed by the sensors follows a known dynamic model with additive noise, then the sensor gains and offsets are identifiable. We propose a dynamic Bayesian nonparametric model to infer the sensors' gains and offsets. Our model allows different sensor clusters to observe different unknown signals, without knowing the sensor clusters a priori. We develop an offline algorithm using block Gibbs sampling and a linearized forward filtering backward sampling method that estimates the sensor clusters, gains and offsets jointly. Furthermore, for practical implementation, we also propose an online inference algorithm based on particle filtering and local Markov chain Monte Carlo. Simulations using a synthetic dataset, and experiments on two real datasets suggest that our proposed methods perform better than several other blind calibration methods, including a sparse Bayesian learning approach, and methods that first cluster the sensor observations and then estimate the gains and offsets.

In the case where agents are nodes of a social network, we investigate the application of truth discovery based on opinions from multiple agents who may be unreliable or biased. We consider the case where agents' reliabilities or biases are correlated if they belong to the same community, which defines a group of agents

with similar opinions regarding a particular event. We incorporate knowledge of the agents' social network in our truth discovery framework.

When the observation model (i.e., the relationship between agent reliabilities, event truths, and agent opinions) is known, we develop Laplace variational inference methods (VISIT) to estimate agents' reliabilities, communities, and the event states. We also develop a stochastic variational inference method to scale our model to large social networks. Simulations and experiments on real data suggest that when observations are sparse, our proposed methods perform better than several other inference methods, including majority voting, TruthFinder, AccuSim, the Confidence-Aware Truth Discovery method, the Bayesian Classifier Combination (BCC) method, and the Community BCC method.

When the observation model is unknown, we use an autoencoder to learn the observation model. A Bayesian network model is proposed to guide the learning of the autoencoder by modeling the dependence of agent reliabilities corresponding to different data samples. At the same time, it also models the social relationships between agents in the network. The proposed approach is unsupervised and is applicable when ground truth labels of events are unavailable. A variational inference method is used to jointly estimate the hidden variables in the Bayesian network and the parameters in the autoencoder. Simulations and experiments on real data suggest that the proposed method performs better than several other inference methods, including majority voting, the Bayesian Classifier Combination (BCC) method, the Community BCC method, and the VISIT method.

In the case where each event is a graph signal, we develop an graph convolution network (GCN) method. GCN has been extensively studied in the literature and one major direction is based on spectral graph theory and graph signal processing. In this thesis, we study the problem from a completely different perspective, by introducing parallel flow decomposition of graphs. The essential idea is to decompose a graph into families of non-intersecting one dimensional (1D) paths, after which, we apply a 1D CNN along each family of paths. Our method is tested on a news article classification dataset in which each article (event) is a graph with keywords as nodes and similarities between keywords as edges. We demonstrate our method, which we call GFCN, achieves better performance than baseline methods.

List of Author's Awards, Patents, and Publications[1]

Publications

- **J. Yang**, X. Zhong, and W. P. Tay, "A dynamic Bayesian nonparametric model for blind calibration of sensor networks," IEEE Internet Things J., vol. 5, no. 5, pp. 3942-3953, Oct. 2018

- **J. Yang**, J. Wang, and W. P. Tay, "Using social network information in Bayesian truth discovery.", IEEE Transactions on signal and information processing over networks,accepetd, April.2019

- **J. Yang**, W. P. Tay, and X. Zhong, A dynamic Bayesian nonparametric model for blind calibration of sensor networks, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, New Orleans, USA, Mar. 2017

- **J. Yang**, and W. P. Tay, "An Unsupervised Bayesian Neural Network for Truth Discovery.", submitted to NIPS.

- F. Ji, **J. Yang* (Equally Contributed First Author)**, and W. P. Tay, "GraphFlow: A new graph convolutional network based on parallel flows.", submited to NIPS

[1]The superscript * indicates joint first authors