Uncertainty-Based Opinion Inference on Network Data Using Graph Convolutional Neural Networks

Xujiang Zhao, Feng Chen
Computer Science Department
University at Albany – SUNY, Albany, NY, USA
xzhao8@albany.edu, fchen5@albany.edu

Jin-Hee Cho *
Department of Computer Science
Virginia Tech, Falls Church, VA, USA
jicho@vt.edu

***Abstract*—In recent years, belief models, such as subjective logic (SL) and collective subjective logic (CSL), have been developed to model an opinion consisting of belief, disbelief, and uncertainty. However, these belief models are designed based on either predefined operators (e.g., discounting and consensus operators) or distribution assumptions (e.g., Markov random fields or MRFs) that are incapable of capturing the heterogeneity of the uncertainty information in large-scale network data. In this paper, we propose a general framework to model and infer heterogeneous uncertainty information in network data based on the state-of-the-art graph convolutional neural networks (GCN). This work is the first that employs a GCN to model the heterogeneous probability density function (PDF) of node-level variables. And then we project this PDF function into a subspace of PDF functions defined based on node-level opinions via knowledge distillation, which provides an effective prediction of the unknown opinion of some nodes based on the observed opinions of the other nodes. Through the extensive simulation experiments, we show that our proposed approach performs better than SL and CSL in predicting unknown opinions when using two road traffic datasets for the validation of the tested algorithms.**

I. Introduction

Uncertainty quantification or management has been realized as a critical area to study for effective decision making process. Uncertainty derives from many different causes, for example, such as information missing during communications, incomplete, corrupted, vague, or conflicting information, and/or biases from humans' limited cognitive capabilities. All these causes generates uncertainty which also significantly affects the effectiveness of decision making.

Data mining researchers also have realized uncertain data as critical factors to minimize and studied in many different domains, including trust in social networks [10], opinion diffusion [3], and graph summarization [13], and so forth.

Subjective Logic (SL), one of well-known belief model derived from Dempster-Shafer Theory (DST) [14], is proposed to deal with the dimension of uncertainty more explicitly. SL defines a binomial opinion with three dimensions, belief, disbelief, and uncertainty. SL also offers a variety of operators that can combine multiple opinions under diverse settings. In this sense, SL allows us to derive a certain structural relation between multiple opinions. However, SL is limited in fusing two opinions, rather than fusing multiple opinions concurrently. Hence, given a large scale network data, it is not scalable. This is proven in our previous work [6] in which we proposed Collective Subjective Logic (CSL) developed based on the combination of Probabilistic Soft Logic (PSL) and Markov Random Fields (MRFs). However, the assumption of distribution based on MRFs limits its capability to deal with heterogeneous network data that may be lossy, noisy, incomplete, or missing, and also in a large-scale.

* This work is done when Jin-Hee Cho was with US Army Research Laboratory.

In this paper, we develop a general framework to address the limitations of SL and CSL as stated above based on the state-of-the-art graph convolutional networks (GCN). This is the first work that employs a GCN to model the heterogeneous probability density function (PDF) of node-level variables and then project this PDF function into a subspace of PDF functions defined based on node-level opinions. This can lead to effective prediction of the unknown opinions of some nodes where the opinions of the other nodes are known.

This work has the following **key contributions**:

1) The proposed GCN-based framework is the first deep learning framework that is capable of predicting the opinions of multiple nodes in a network collectively based on GCN modeling.
2) The proposed GCN-based method achieves both efficiency and effectiveness by leveraging the GCN to model heterogeneous dependencies between the variables in a network and *knowledge distillation* to transfer the heterogeneous dependencies into the prediction of opinions.
3) We validate the performance of our proposed approach through the extensive simulation experiments based on two road traffic datasets. We compare the performance of our proposed approach with that of the existing counterparts with respect to prediction accuracy of unknown opinions and algorithmic complexity.

II. Preliminaries

A. Binomial Opinion in Subjective Logic

In SL, a binomial opinion is defined in terms of belief, disbelief, and uncertainty towards a given proposition x. For simplicity, we omit x in the following notations. To formally put, an opinion w is represented by:

$$w = (b, d, u, a) \tag{1}$$

978-1-5386-7185-6/18/$31.00 ©2018 IEEE

where b is belief (e.g., true), d is disbelief (e.g., false), and u is uncertainty (i.e., ignorance or lack of evidence). a represents a base rate, a prior knowledge upon no commitment (e.g., neither true nor false), where $b+d+u=1$ for $(b,d,u,a) \in [0,1]^4$. We denote an opinion by w, which can be *projected* onto a single probability distribution by removing the uncertainty mass.

A binomial opinion follows a Beta PDF (probability density function), denoted by $\text{Beta}(p|\alpha,\beta)$, where α represents the amount of positive evidence and β is the amount of negative evidence [11].

In SL, uncertainty u decreases as more evidence, α and β, is received over time. An opinion w can be obtained based on α and β as $w=(\alpha,\beta)$. This can be translated to $w=(b,d,u,a)$ using the mapping rule in SL.

SL offers an operator, $\otimes$, to discount trust when an entity does not have any direct experience with another entity. That is, transitive trust based on structural relations is used to derive trust between two entities who have not interacted before. Trust from i to j, denoted by $w^i_j=(b^i_j,d^i_j,u^i_j,a^i_j)$, and trust from j to k, $w^j_k=(b^j_k,d^j_k,u^j_k,a^j_k)$, are used to derive trust from i to k, $w^i_k := (b^i_k,d^i_k,u^i_k,a^i_k)=w^i_j\otimes w^j_k$. It is obtained by:

$$\begin{gathered} b^i_k = b^i_j \otimes b^j_k = b^i_j b^j_k,\ d^i_k = d^i_j \otimes d^j_k = b^i_j d^j_k \\ u^i_k = u^i_j \otimes u^j_k = d^i_j + u^i_j + b^i_j u^j_k,\ a^i_k = a^i_j \otimes a^j_k = a^j_k. \end{gathered} \tag{2}$$

SL also provides a consensus operator, $\oplus$, to find a consensus between two opinions [11] where two entities observe a same entity. An opinion after i exchanges opinions with k is given by $w^i_k \oplus w^j_k$, where:

$$\begin{gathered} b^i_k \oplus b^j_k = \frac{b^i_k u^j_k + b^j_k u^i_k}{\zeta},\ d^i_k \oplus d^j_k = \frac{d^i_k u^j_k + d^j_k u^i_k}{\zeta} \\ u^i_k \oplus u^j_k = \frac{u^i_k u^j_k}{\zeta},\ a^i_k \oplus a^j_k = a^i_k. \end{gathered} \tag{3}$$

where $\zeta = u^i_j + u^j_k - u^i_j u^j_k > 0$. When $\zeta = 0$, $w^i_k \oplus w^j_k$ is defined by:

$$\begin{gathered} b^i_k \oplus b^j_k = \frac{\psi b^i_k + b^j_k}{\psi+1},\ d^i_k \oplus d^j_k = \frac{\psi d^i_k + d^j_k}{\psi+1} \\ u^i_k \oplus u^j_k = 0,\ a^i_k \oplus a^j_k = a_k. \end{gathered} \tag{4}$$

where $\psi = \lim(u^i_k/u^j_k)$. These discounting, $\otimes$, and consensus, $\oplus$, operators [11] are used to derive trust measures based on the trust opinions of relationships. Due to space constraint, we don't show an example of using these operators. Interested readers can be referred to [6].

In this work, we aim to derive a set of unknown opinions $\mathbf{x}=\{x_1,\cdots,x_n\}$ when a set of observed opinions $\mathbf{y}=\{y_1,\cdots,y_m\}$ is given where both opinions are represented by a binomial opinion with four dimensions, as described in Eq. 1 (i.e., w_{x_i} for $i=1\cdots n$ and w_{y_j} for $j=1\cdots m$).

B. Graph Convolutional Networks (GCN)

Denote a graph as $\mathbb{G}=(\mathbb{V},\mathbb{E},A)$, where $\mathbb{V}=\{1,\cdots,n\}$ refers to the set of nodes and $\mathbb{E}\subseteq\mathbb{V}\times\mathbb{V}$ refers to the set of edges. Let $A\in\{0,1\}^{n\times n}$ be the adjacency matrix, where $A_{i,j}=1$ if $(i,j)\in\mathbb{E}$ and, otherwise, $A_{i,j}=0$. The (unnormalized) graph Laplacian matrix is an $n\times n$ symmetric positive-semidefinite matrix $L=D-A$, where D is the degree matrix and $D_{i,i}$ refers to the degree of node i and $D_{i,i}=0$ for $i\neq j$.

The graph Laplacian has an eigen decomposition $L=\Phi\Lambda\Phi^T$, where $\Phi=(\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_n)$ are the orthonormal eigenvectors and $\Lambda=diag(\lambda_1,\cdots,\lambda_n)$ is the diagonal matrix of corresponding eigen values. The eigenvalues serve as the role of frequencies in classical harmonic analysis and the eigenvectors are interpreted as Fourier atoms. Given a signal $\mathbf{r}\in\mathbb{R}^n$ (or a vector of feature values) on the nodes of graph $\mathbb{G}$, where r_i refers to a feature value at node i, its graph Fourier transform is given by $\hat{\mathbf{r}}=\Phi^T\mathbf{r}$. Given two signals $\mathbf{r}$ and $\mathbf{b}$ on the graph, we can define their spectral convolution as the element-wise product of their Fourier transformations,

$$\mathbf{r}\star\mathbf{b}=\Phi^T(\Phi^T\mathbf{r})\circ(\Phi^T\mathbf{b})=\Phi diag(\hat{r}_1,\cdots,\hat{r}_n)\hat{\mathbf{b}}, \tag{5}$$

which is a property of the well-known Convolutional Theorem in the Euclidean case.

As a graph is irregular, it is difficult to directly define a convolution on the nodes. Instead, Bruna et al. [5] used the spectral definition of convolution (Eq. (5)) to generalize Convolutional Neural Networks (CNNs) on graphs, which has a spectral convolutional layer of the form as:

$$g_\theta\star\mathbf{r}=\Phi g_\theta\Phi^T\mathbf{r}. \tag{6}$$

The filter g_θ can be defined as a function of the eigenvalues of L, i.e., $g(\Lambda)$. Evaluation of Eq. (6) is computationally expensive because multiplication with the eigenvector matrix Φ is $O(n^2)$, in addition to the high computational cost in computing the eigendecomposition of L in the first place. To address this problem, Hammond et al. [7] suggest that $g_\theta(\Lambda)$ can be well-approximated by a truncated expansion in terms of Chebyshev polynomials $T_k(r)$ up to K-th order:

$$g_\theta(\Lambda)\approx\sum_{k=1}^{K}\theta_k T_k(\tilde{\Lambda}), \tag{7}$$

with a rescaled $\tilde{\Lambda}=\frac{2}{\lambda_{max}}\Lambda-I_n$. λ_{max} refers to the largest eigenvalue of L. $\theta\in\mathbb{R}^K$ is now a vector of Chebyshev coefficients. The Chebyshev polynomial can be recursively defined as $T_k(r)=2xT_{k-1}(r)-T_{k-2}(r)$, with $T_0(r)=1$ and $T_1(r)=r$. Applying the approximation based on Chebyshev polynomials, a convolution of a signal x with a filter g_θ now has the approximated form:

$$g_\theta\star\mathbf{r}\approx\sum_{k=1}^{K}\theta_k T_k(\tilde{L})\mathbf{r}. \tag{8}$$

By stacking multiple convolutional layers of the form of Eq. (8) in which each layer is followed by a point-wise non-linearity filter, we can therefore design a multi-layer convolutional neural network model based on graph convolutions.

For example, a two-layer GCN model [12] for the task of node classification on a network with a symmetric adjacency matrix A (binary or weighted) can be formulated as:

$$p = g(\mathbf{r}, A) = \text{softmax}\left(g_{W^{(1)}} \star \text{ReLU}(g_{W^{(0)}} \star \mathbf{r})\right), \quad (9)$$

where the output matrix $p \in [0,1]^{n\times 2}$ provides the predicted probabilities of the binary classes of the n nodes. Here, $W^{(0)}, W^{(1)}$ are weight matrixs.

III. Problem Formulation

In this section, we describe an example to motivate a problem to solve in this work. We also show how to formulate a given uncertainty-based inference problem.

A. Example Scenario

In this work, we aim to infer unknown opinions based on a set of known opinions in terms of the applications in traffic congestion prediction in a road network. Given a network, defined as $\mathbb{G} = (\mathbb{V}, \mathbb{E} = \mathbb{E}_1 \cup \mathbb{E}_2, f)$, where $\mathbb{V} = \{1, 2, \cdots, l\}$ is the set of vertices (i.e., intersections in the road network), $\mathbb{E} \subseteq \mathbb{V}\times\mathbb{V}$ is the set of edges (i.e., road links), and the function $f : \mathbb{E} \rightarrow \{0, 1\}$ refers to a boolean variable $f(e)$ for edge e, in which state 0 indicates 'non-congested' while state 1 refers to 'congested' at time t. Suppose there is a subset of edges $\mathbb{E}_1 = \{e_1, \cdots, e_M\} \subseteq \mathbb{E}$ with traffic sensors (e.g., cameras) installed at these edges where the edges without traffic sensors are represented by $\mathbb{E}_2 = \{o_1, \cdots, o_N\}$, and $\mathbb{E} = \mathbb{E}_1 \cup \mathbb{E}_2$.

Suppose that we are given the current observations of the congestion variables on $\mathbb{Y}$, $\mathbf{y} = [f(e_1), \cdots, f(e_M)] \in \{0, 1\}^M$, and the beliefs (i.e., states) towards these variables estimated based on their observations at the T historical time stamps. A belief over the states of a congestion variable y_i can be represented as a subjective opinion, defined as $w_{y_i} = (b_{y_i}, d_{y_i}, u_{y_i}, a_{y_i})$ in Eq. (1), or equivalently as a Beta distribution with the evidence parameters Beta$(p_{y_i}; \alpha_{y_i}, \beta_{y_i})$ [11]. For an edge $e_i \in \mathbb{E}_1$ that has the T most recent observations $\{f(e_i^1), \cdots, f(e_i^T)\}$, where r or s is the number of 0's (i.e., no traffic) or 1's (i.e., traffic) in these observations, respectively. Beta$(p_{y_i}; \alpha_{y_i}, \beta_{y_i})$ can be estimated by:

$$\alpha_{y_i} = r + a_{y_i}W, \;\; \beta_{y_i} = s + (1 - a_{y_i})W. \quad (10)$$

where r and s are the amounts of positive and negative evidence, W is a predefined non-informative prior weight (i.e., the amount of uncertain evidence) and a_{y_i} is a predefined probability of prior general background knowledge on proposition y_i used to interpret uncertain evidence W.

Given these information, we aim to predict the beliefs about the states of the congestion variables at the edges in $\mathbb{X}$ without sensors (i.e., intersections without any camera), denoted as $\boldsymbol{\omega}_\mathbf{x} = [\boldsymbol{\omega}_{x_1}, \cdots, \boldsymbol{\omega}_{x_N}]$, where x_i refers to the state variable of a link $o_i \in \mathbb{X}$ and $\boldsymbol{\omega}_{x_i}$ refers to evidence parameters of a Beta distribution Beta$(p_{x_i}; \alpha_{x_i}, \beta_{x_i})$ that represents the belief about the state of x_i. As the edges in $\mathbb{X}$ do not have any observations, their beliefs cannot be directly inferred. Therefore, we can infer the unknown beliefs towards the edges $\mathbf{x}$ based on the structural relations between the known beliefs on other edges.

TABLE I
Key notations and their meanings.

$w_x = (b_x, d_x, u_x, a_x)$	A binomial subjective opinion of a binary random variable x as defined in Eq. (1)
p_x	A truth probability of a binary random variable x
$\omega_x = (\alpha_x, \beta_x)$	Evidence parameters of Beta$(p_x\|\alpha_x, \beta_x)$ that corresponds to a subjective opinion w_x.
$\mathbf{y} = [y_1, \cdots, y_M]$, $\mathbf{p_y} = [p_{y_1}, \cdots, p_{y_M}]$, $\boldsymbol{\omega}_\mathbf{y} = [\omega_{y_1}, \cdots, \omega_{y_M}]$	$\mathbf{y}$ is a vector of M input binary random variables whose subjective opinions are known. $\mathbf{p_y}$ and $\boldsymbol{\omega}_\mathbf{y}$ are the corresponding vectors of truth probabilities and subjective opinions of $\mathbf{y}$, respectively.
$\mathbf{x} = [x_1, \cdots, x_N]$, $\mathbf{p_x} = [p_{x_1}, \cdots, p_{x_N}]$, $\boldsymbol{\omega}_\mathbf{x} = [\omega_{x_1}, \cdots, \omega_{x_M}]$	$\mathbf{x}$ is a vector of N output binary random variables, whose subjective opinions are to be predicted. $\mathbf{p_x}$ and $\boldsymbol{\omega}_\mathbf{x}$ are the corresponding vectors of truth probabilities and subjective opinions of $\mathbf{x}$, respectively.

B. Problem Formulation

We formulate the problem of uncertainty-based inference by:

Problem 1 (Uncertainty-based opinion inference in network data)**:** *Let us define the following notations:*

- *Let $\mathbb{G} = (\mathbb{V}, \mathbb{E}, f)$ be the input network as defined above.*
- *Let $\mathbf{y} = (y_1, \cdots, y_M)$ be a vector of input boolean variables over $\mathbb{E}_1$, whose opinions are denoted as $\omega_\mathbf{y} = (\omega_{y_1}, \cdots, \omega_{y_M})$, implying that the PDF (probability density function) of the truth probability of the variable y_i, p_{y_i}, is Beta$(p_{y_i}; \omega_{y_i})$; Let y_i, p_{y_i} follows a Bernoulli distribution, Bern$(y_i; p_{y_i})$. Let $\{\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(T)}\}$ be the observations of $\mathbf{y}$ in the T most recent observations.*
- *Let $\mathbf{x} = (x_1, \cdots, x_N)$ be a vector target boolean variables over $\mathbb{E}_2$, whose opinions are denoted as $\omega_\mathbf{x} = (\omega_{x_1}, \cdots, \omega_{x_N})$, implying that the PDF (probability density function) of the truth probability of the variable x_i, p_{x_i}, is Beta$(p_{x_i}; \omega_{x_i})$; Let x_i, p_{x_i} follows a Bernoulli distribution, Bern$(x_i; p_{x_i})$.*
- *Let $\mathbf{p_x} = (p_{x_1}, \cdots, p_{x_N})$, $\mathbf{p_y} = (p_{y_1}, \cdots, p_{y_M})$.*

Given

- *$\mathcal{G} = (\mathbb{V}, \mathbb{E} = \mathbb{Y} \cup \mathbb{X}, f)$, the input network;*
- *$\{\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(T)}\}$, the observations of vector of input Boolean variables and $\boldsymbol{\omega}_\mathbf{y} = (\omega_{y_1}, \cdots, \omega_{y_M})$, the subjective opinions on $\mathbf{y}$.*

Predict $\boldsymbol{\omega}_\mathbf{x}$, *the opinion on the vector of target Boolean variables* $\mathbf{x}$.

IV. Our proposed approach

In this section, we present a novel deep learning framework that combines a probabilistic model of opinions (i.e., SL) with a GCN model to estimate the unknown opinions $\boldsymbol{\omega}_\mathbf{x}$ while taking into account the heterogeneous dependencies among the variables $\mathbf{x}$ and $\mathbf{y}$. We first introduce the two separate models and then discuss how these two models can be effectively bridged for the task of opinion inference using the framework of knowledge distillation. For notational simplicity, we consider all the binary variables in $\mathbf{x}$ and $\mathbf{y}$ as dummy variables, i.e., $x_i, y_j \in [0, 1]^2$.

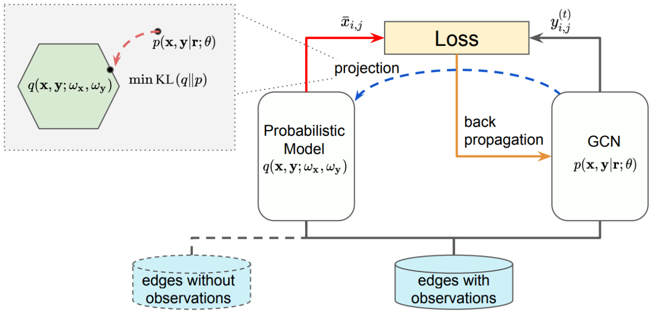


Fig. 1. Framework overview. At each iteration ℓ, the unknown opinions $\boldsymbol{\omega}_\mathbf{x}^{(\ell+1)}$ are estimated by projecting the conditional PDF of the GCN model, $p(\mathbf{x},\mathbf{y}|\mathbf{r};\theta^{(\ell)})$, to a subspace $\mathcal{Q}$ of PDF functions $q(\mathbf{x},\mathbf{y};\omega_\mathbf{x})$ defined by the opinions $\omega_\mathbf{x}$. After that, the parameters $\theta^{(\ell+1)}$ of the GCN model are updated to minimize the prediction loss of the true labels $\mathbf{y}$ and the feedback soft labels $\bar{\mathbf{x}}$ obtained from the estimated opinions $\boldsymbol{\omega}_\mathbf{x}^{(\ell+1)}$.

A. A Probabilistic Model of Opinions ($q(\mathbf{x},\mathbf{y};\omega_\mathbf{x},\omega_\mathbf{y})$).

Based on the properties of Beta opinions as discussed in Section II-A, we have the following Bayesian distributions: $y_i \sim \mathrm{Bern}(y_i;p_{y_i})$; $p_{y_i} \sim \mathrm{Beta}(p_{y_i};\omega_{y_i})$. The probability density function (PDF) of y_i based on its opinion ω_{y_i} can be calculated as $q(y_i;\omega_{y_i}) = \int \mathrm{Beta}(p_{y_i};\omega_{y_i})\mathrm{Bern}(y_i;p_{y_i})dp_{y_i} = \mathrm{Bern}(y_i;\frac{\alpha_{y_i}}{\alpha_{y_i}+\beta_{y_i}})$, where $\omega_{y_i} = (\alpha_{y_i},\beta_{y_i})$. The PDF function $q(x_i;\omega_{x_i})$ can be calculated in a similar manner. The joint PDF function of $\mathbf{x}$ and $\mathbf{y}$ based on the opinions $\boldsymbol{\omega}_\mathbf{x}$ and $\boldsymbol{\omega}_\mathbf{y}$ can be then calculated as:

$$q(\mathbf{x},\mathbf{y};\omega_\mathbf{x},\omega_\mathbf{y}) = \prod_{i=1}^{N} q(x_i;\omega_{x_i}) \prod_{j=1}^{M} q(y_j;\omega_{y_j}) \tag{11}$$

B. A GCN Model of $\mathbf{x}$ and $\mathbf{y}$ ($p(\mathbf{x},\mathbf{y}|\mathbf{r};\theta)$)

To employ a GCN model here, we first design a pseudo feature vector $\mathbf{r} \in \mathbb{R}^{N+M}$, in which $r_i = 0$ if $i \leq N$; $r_j = -1$ for $j > N$ if $y_{j-N,0} = 1$; and otherwise, $r_j = 1$. A GCN model defines a conditional PDF $p(\mathbf{x},\mathbf{y}|\mathbf{r};\theta)$ by using a softmax output layer that produces a $((N+M)\times 2)$-dimensional soft prediction matrix $p_{\mathbf{x},\mathbf{y}} \in [0,1]^{(N+M)\times 2}$ as defined below,

$$p_{\mathbf{x},\mathbf{y}} = g(\mathbf{r};A,\theta) : \mathbb{R}^{M+N} \rightarrow [0,1]^{(M+N)\times 2}, \tag{12}$$

where $g(\mathbf{r};A,\theta)$ is defined in Eq. (9) and $\theta = \{W^{(0)},W^{(1)}\}$ refers to the parameters of the GCN model. The conditional PDF function $p(\mathbf{x},\mathbf{y}|\mathbf{r};\theta)$ has the form as:

$$p(\mathbf{x},\mathbf{y}|\mathbf{r};\theta) = \prod_{i=1}^{N} p(x_i|r_i;\theta) \prod_{j=1}^{M} p(y_i|r_{N+j};\theta) \tag{13}$$

where $p(x_i|r_i;\theta) = \prod_{k=1}^{2}[g_{i,k}(r,A;\theta)]^{x_{i,k}}$ and $p(y_i|r_{N+j};\theta) = \prod_{k=1}^{2}[g_{i+N,k}(r,A;\theta)]^{y_{i,k}}$.

C. Bridging the Two Models for Prediction of $\omega_\mathbf{x}$ based on Knowledge Distillation.

The goal is to predict the unknown opinions $\boldsymbol{\omega}_\mathbf{x}$ based on the observations and opinions of the input variables $\mathbf{y}$, including $\{\mathbf{y}^{(1)},\cdots,\mathbf{y}^{(T)}\}$ and $\boldsymbol{\omega}_\mathbf{y}$. **Employing the opinion-based PDF model $q(\mathbf{x},\mathbf{y};\omega_\mathbf{x},\omega_\mathbf{y})$ alone is insufficient as it fails to model the heterogeneous dependencies between $\mathbf{x}$ and $\mathbf{y}$ based on their relations in the network. Similarly, the GCN model $p(\mathbf{x},\mathbf{y}|\mathbf{r};\theta)$ can not be directly applied to predict $\boldsymbol{\omega}_\mathbf{x}$ as it is incapable of modeling the opinions directly, although it has well modeled the heterogeneous dependency information among $\mathbf{x}$ and $\mathbf{y}$.** We develop a novel iterative knowledge distillation algorithm that transfers the heterogeneous dependency information between $\mathbf{x}$ and $\mathbf{y}$ in the GCN model into the estimation of the unknown opinions $\boldsymbol{\omega}_\mathbf{x}$, while at the same time, we use the estimated $\boldsymbol{\omega}_\mathbf{x}$ in each iteration to help improve the estimate of the parameters θ for the GCN model. The main steps of our proposed algorithm are summarized in Algorithm 1 and its key idea is illustrated in Fig. 1. In particular, our approach is composed of the following two main components:

Algorithm 1: Prediction of Unknown Opinions

Input: $\{\mathbf{y}^{(1)},\cdots,\mathbf{y}^{(T)}\}$ and $\boldsymbol{\omega}_\mathbf{y}$
Output: $\boldsymbol{\omega}_\mathbf{x}$

- $\ell = 1$;
- Estimate the initial $\theta^{(\ell)}$ by solving Problem (18) with $\pi = 1$ using the back propagation algorithm;
- $\pi = 0.5$;
- **repeat**
  - Calculate $\boldsymbol{\omega}_\mathbf{x}^{(\ell+1)}$ via Eq. (17);
  - Estimate $\theta^{(\ell+1)}$ by solving Problem (18) using the back propagation algorithm;
  - $\ell = \ell + 1$;
- **until** *convergence*
- **return** $\boldsymbol{\omega}_\mathbf{x}^{l}$

(1) Transferring of the dependency information from the GCN model to the probabilistic model of opinions for predicting opinions $\boldsymbol{\omega}_\mathbf{x}$ based on the estimated parameters $\theta^{(\ell)}$ at iteration ℓ. In this component, we assume that an estimation of the parameters θ of the GCN model at iteration ℓ is given as $\theta^{(\ell)}$. Our proposed approach is to estimate an approximate PDF function $q(\mathbf{x},\mathbf{y})$ in the parametric space $\mathcal{Q} = \{q(\mathbf{x},\mathbf{y}) \mid q(\mathbf{x},\mathbf{y}) = q(\mathbf{x},\mathbf{y};\omega_\mathbf{x},\omega_\mathbf{y}), \forall\omega_\mathbf{x}\}$ that is as close as possible to the conditional PDF function defined by the GCN model. We consider the Kullback–Leibler divergence as the distance metric between PDF functions and the resulting optimization problem can then be formulated as follows:

$$\min_{\boldsymbol{\omega}_\mathbf{x}} \mathrm{KL}\Big(\prod_{t=1}^{T} q(\mathbf{x}^{(t)},\mathbf{y}^{(t)};\boldsymbol{\omega}_\mathbf{x},\boldsymbol{\omega}_\mathbf{y})\| \prod_{t=1}^{T} p(\mathbf{x}^{(t)},\mathbf{y}^{(t)}|\mathbf{r}^{(t)};\theta^{(\ell)})\Big), \tag{14}$$

which can be interpreted as the a projection of the PDF function $q(\mathbf{x}^{(t)},\mathbf{y}^{(t)};\boldsymbol{\omega}_\mathbf{x},\boldsymbol{\omega}_\mathbf{y})$ to the space of PDF functions $\mathcal{Q}$ as parameterized by the unknown opinions $\boldsymbol{\omega}_\mathbf{x}$. Further, Problem (14) can be shown to be equivalent to

$$\min_{\boldsymbol{\omega}_\mathbf{x}} \sum_{t=1}^{T}\sum_{i=1}^{N} \mathbb{E}[\log q(x_i^{(t)};\omega_{x_i})] + \sum_{t=1}^{T}\sum_{j=1}^{M} \mathbb{E}[\log q(y_j^{(t)};\omega_{y_j})] - \sum_{t=1}^{T}\sum_{i=1}^{N} \mathbb{E}[\log p(x_i^{(t)}|r_i^{(t)};\theta^{(\ell)})] - \sum_{t=1}^{T}\sum_{j=1}^{M} \mathbb{E}[\log p(y_j^{(t)}|r_{N+j}^{(t)};\theta^{(\ell)})],$$

where $\mathbb{E}[f(\mathbf{x}, \mathbf{y})] = \sum_{\mathbf{x},\mathbf{y}} f(\mathbf{x}, \mathbf{y}) q(\mathbf{x}, \mathbf{y}; \boldsymbol{\omega}_{\mathbf{x}}, \boldsymbol{\omega}_{\mathbf{y}})$. As the above objective function is an additive function of the opinions $\{\omega_{x_1}, \cdots, \omega_{x_N}\}$, we conclude that Problem (14) can be decomposed N independent subgraph problem as follows:

$$\min_{\omega_{x_i}} \sum_{t=1}^{T} \mathbb{E}\left([\log q(x_i^{(t)}; \omega_{x_i})] - \mathbb{E}[\log p(x_i^{(t)} | r_i^{(t)}; \theta^{(\ell)})] \right), \quad (15)$$

where $\omega_{x_i} = (\alpha_{x_i}, \beta_{x_i})$ refers to the parameters of the opinion of the variable x_i corresponding to the node o_i in $\mathbb{E}_2$. Problem (15) can be shown to be equivalent to

$$\min_{\bar{p}_{x_i}} \sum_{t=1}^{T} \bar{p}_{x_i} \log \frac{\bar{p}_{x_i}}{g(r^{(t)}, A; \theta^{(\ell)})} + (1 - \bar{p}_{x_i}) \log \frac{1 - \bar{p}_{x_i}}{1 - g(r^{(t)}, A; \theta^{(\ell)})},$$

where $\bar{p}_{x_i} = \frac{\alpha_{x_i}}{\alpha_{x_i} + \beta_{x_i}}$. By calculating the gradient of the above objective function and conditioning the gradient to 0, we obtain the the analytical solution for $\bar{p}_{x_i}$:

$$\bar{p}_{x_i}^{(\ell)} = \frac{\sqrt[T]{\prod_{t=1}^{T} \frac{g(r^{(t)}, A; \theta^{(\ell)})}{1 - g(r^{(t)}, A; \theta^{(\ell)})}}}{\sqrt[T]{\prod_{t=1}^{T} \frac{g(r^{(t)}, A; \theta^{(\ell)})}{1 - g(r^{(t)}, A; \theta^{(\ell)})}} + 1} \quad (16)$$

As the parameters α_{x_i} and β_{x_i} satisfy the property that $\alpha_{x_i} + \beta_{x_i} = T$, we obtain the analytical form of the updated opinions $\omega_{x_i}^{(\ell+1)} = (\alpha_{x_i}^{(\ell+1)}, \beta_{x_i}^{(\ell+1)})$ as follows:

$$\alpha_{x_i}^{(\ell+1)} = \bar{p}_{x_i}^{(\ell)} \cdot T; \;\; \beta_{x_i}^{(\ell+1)} = T - \alpha_{x_i}^{(\ell+1)}, \quad \forall i = 1, \cdots, N. \; (17)$$

(2) Estimation of θ for the GCN model based on feedback from the predicted opinions $\boldsymbol{\omega}_{\mathbf{x}}^{(\ell)}$ at iteration ℓ. Here, we assume that the predicted opinions $\boldsymbol{\omega}_{\mathbf{x}}^{(\ell)}$ are given, which provides a soft prediction for each variable x_i: $\bar{x}_{x_i} = \frac{\alpha_{x_i}^{(\ell)}}{\alpha_{x_i}^{(\ell)} + \beta_{x_i}^{(\ell)}}$. The soft predictions transfer back the knowledge of opinions $\boldsymbol{\omega}_{\mathbf{x}}^{(\ell)}$ to the training process of the GCN model. The resulting objective for the estimation of θ can be defined as:

$$\theta^{(\ell+1)} = \arg\min_{\theta} \sum_{t=1}^{T} \sum_{i=1}^{M} \sum_{j=1}^{2} \pi y_{i,j}^{(t)} \cdot \log g_{i+N,j}(r^{(t)}, A; \theta) + \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{2} (1 - \pi) \bar{x}_{i,j} \log g_{i,j}(r^{(t)}, A; \theta), \quad (18)$$

where the objective function is defined based on the cross entropy loss for classification and π is an imitation parameter that calibrates the relative importance of the two components in the objective function. Problem (18) can then be solved using the general back propagation procedure for deep learning models. We note that a similar imitation procedure has been shown effective in regularizations of deep learning models, such as an imitation procedure that transfers the knowlege of logic rules to a deep neural network for modeling of sequences [9] and another one for model compression where the procedure is called *distillation* [8].

V. Results and Analysis

A. Experimental Settings

1) Road traffic datasets: We collected traffic data from June 1, 2013 to March 31, 2014 across two cities from INRIX [1], Washington D.C. and Philadelphia (PA), as summarized in Table II. The raw INRIX dataset provides traffic speed and reference speed information for each road link per hour interval. A reference speed is defined as the "uncongested free flow speed" for each road segment [2]. It is calculated based upon the 85-th percentile of the measured speed for all time periods over a few years, where the reference speed serves as a threshold separating two traffic states, *congested* vs. *uncongested*. The road traffic dataset for each of the two cities has 43 weeks in total. An hour is represented by a specific combination of hours of a day ($h \in \{6, 9, 12, ..., 21\}$), days of a week ($d \in \{1, 2, 3, 4, 5\}$), and weeks ($w \in \{1, 2, ..., 43\}$): (h, d, w).

TABLE II
Description of the three real-world datasets

Dataset name	# nodes	# edges	# weeks	# snapshots (hours) in total
Washington, D.C.	1,383	1,878	43	3440
Philadelphia	603	708	43	3440

Estimation of opinions of training and testing edges in each dataset. For each road traffic dataset, the opinion of a specific (training or testing) link s at an hour (h, d, w) is estimated based on the observations of the same hour in previous T weeks $\{x_{s,h,d,w}, x_{s,h,d,w-1}, ..., x_{s,h,d,w-T+1}\}$ as the evidence, where $x_{s,h,d,w}$ refers to the congestion observation (0 or 1) of the link s at hour (h, d, w) and T refers to a predefined time window size.

2) Parameter settings: The main parameters for all the datasets include T (time window size) and TR (test ratio or the percentage of edges that are tested). The values of T are set as $T \in \{2, 3, 6, 8, 11\}$. The uncertainty mass values are set to $\{50\%, 40\%, 25\%, 20\%, 15\%\}$. The values of TR are set to $\{10\%, 20\%, 30\%, 40\%\}$.

3) Performance metrics: The uncertainty mass, u_s, for each training or testing edge is a known and constant value, u, after the window size T is predefined, without the actual observations of this link. For this reason, the empirical analysis based on the road traffic datasets focuses the comparison between the proposed method and comparable methods based on the two metrics: *Expected Belief MSE* (EB-MSE) and running time complexity (second.). As Table I shows, EB-MSE is defined as:

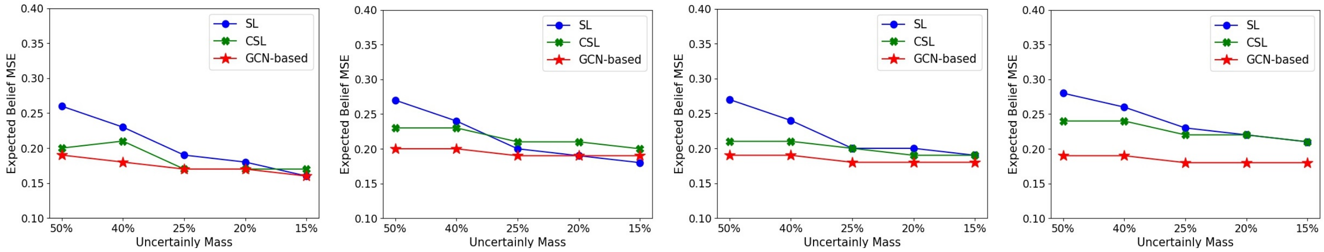
$$\text{EB-MSE}(\boldsymbol{\omega}_{\mathbf{x}}) = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{a_{x_i}}{a_{x_i} + b_{x_i}} - \frac{a_{x_i}^{\star}}{a_{x_i}^{\star} + b_{x_i}^{\star}} \right| \quad (19)$$

where $\omega_{x_i} = (a_{x_i}, b_{x_i}, u)$ and $\omega_{x_i}^{\star} = (a_{x_i}^{\star}, b_{x_i}^{\star}, u)$ refers to the predicted and true opinions of a target variable x_i, respectively; and $\frac{a_{x_i}}{a_{x_i} + b_{x_i}}$ refers to the expected belief of the opinion ω_{x_i}.

4) Comparison methods: We compared our proposed GCN-based method with the comparable counterpart methods, including CSL [6] and SL [11], as discussed in Section II-A.

B. Experimental Results based on Real-World Datasets

Fig. 2 compares the performance of our proposed GCN-based approach with that of the two counterpart methods (i.e., CSL and SL) with respect to EB-MSE based on two road taffic datasets (PA and DC). The results indicate that our GCN-based approach performs the best among all in EB-MSE. The



(a) 10% Test Ratio, PA Dataset (b) 30% Test Ratio, PA Dataset (c) 10% Test Ratio, DC Dataset (d) 30% Test Ratio, DC Dataset

Fig. 2. Comparison of CSL and counterpart methods under test ratio of 10% and 30%: EB-MSE vs. uncertainty mass

GCN-based approach is less sensitive than CSL and SL to u by showing that EB-MSE in SL significantly decreases as u decreases. When u is low (i.e., 20% or 15%), EB-MSE of the three methods are comparable under certain settings while the GCN-based approach still outeprforms with 30% of the test ratio in overall.

Fig. 3 shows the average log snapshot-level running times for different test ratios (i.e., 10% and 30%) on the two road traffic datasts obtained for PA and DC. When the network size increases, the time complexity of SL increases in an exponential order while those of the GCN-based approach and CSL increase in a linear order. Since SL needs to identify all independent paths among the observed opinions [11], it has much higher complexity for large network data than the other two methods. Further, our GCN-based method shows better running time than CSL since the GCN-based method uses GCN and knowledge distillation whose scales are in a linear order.

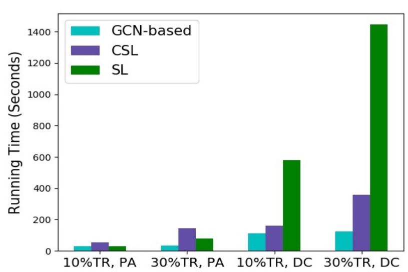


Fig. 3. Running Time vs. $TR = 10\%, 30\%$ on the Real-World Datasets

Summary. Overall, our GCN-based method outperforms all other counterparts. In particular, our GCN-based method shows less sensitivity over a wide range of the uncertainty mass, implying high resilience, compared to CSL and SL. The performance order in EB-MSE follows: GCN-based method $>$ CSL $>$ SL. The higher performance of GCN-based method and CSL over SL is because they both use collective opinion inference procedures. The performance order in running time is also the same as that in EB-MSE.

VI. Conclusion

In this work, we propose a novel deep-learning based approach via GCN and knowledge distillation to derive unknown opinions based on known opinion probabilities and the heterogeneous dependencies between node-level variables in the input network. In our future work, we plan to validate the performance of our GCN-based method based on more real-world datasets (e.g., cyber security datasets) and extend our proposed method to address uncertainty-based online opinion inference problems.

Acknowledgments

This work is partially supported by ARL's Competitive Basic Research Program under Computational and Information Sciences Directorate and by the US Army Research Office under grant number W911NF1720129. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] "Inrix," http://inrix.com/publicsector.asp.
[2] "Reference speed for congestion evaluation," http://www.inrix.com/scorecard/methodology.asp/.
[3] S. Bach, M. Broecheler, L. Getoor, and D. O'leary, "Scaling mpe inference for constrained continuous markov random fields with consensus optimization," in *NIPS*, 2012, pp. 2654–2662.
[4] M. Bröcheler, L. Mihalkova, and L. Getoor, "Probabilistic similarity logic," in *UAI* 2010, pp. 73–82.
[5] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.
[6] F. Chen, C. Wang, and J.-H. Cho, "Collective subjective logic: Scalable uncertainty-based opinion inference," in *IEEE Big Data*, 2017, pp. 7–16.
[7] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
[8] G. E. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *CoRR*, vol. abs/1503.02531, 2015.
[9] Z. Hu, X. Ma, Z. Liu, E. H. Hovy, and E. P. Xing, "Harnessing deep neural networks with logic rules," in *ACL*, 2016.
[10] B. Huang, A. Kimmig, L. Getoor, and J. Golbeck, "Probabilistic soft logic for trust analysis in social networks," in *International Workshop on Statistical Relational AI*, 2012, pp. 1–8.
[11] A. Jøsang, *Subjective Logic*. Springer, 2016.
[12] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *CoRR*, 1609.02907, 2016.
[13] A. Memory, A. Kimmig, S. H. Bach, L. Raschid, and L. Getoor, "Graph summarization in annotated data using probabilistic soft logic," in *Proceedings of the 8th ICUR*. 2012, pp. 75–86.
[14] G. Shafer *et al.*, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.