# Trust Estimation of Sources Over Correlated Propositions

Lance Kaplan[1] and Murat Şensoy[2]
[1]US Army Research Lab, Adelphi, MD, US
[2]Department of Computer Science, Ozyegin University, Istanbul, Turkey

***Abstract*—This work analyzes the impact of correlated propositions when estimating the reporting behavior of information sources. These behavior estimates are critical for fusion, and traditional methods assume the propositions are statistically independent. A new source behavior estimation methods is presented that accounts for statistical dependencies between the training propositions. Simulations seem to indicate that the potential performance gains for accounting for the correlations is small relative to the increased computational complexity. One may conclude that the traditional independence assumption in source behavior estimation methods is reasonable even in cases where it is actually violated.**

## I. Introduction

The fusion of information from multiple possibly human sources requires understanding of how these sources introduce distortions in their reports. This is similar to a classical sensor fusion filter that essentially produces a weighted sum of sensor data where the weights are inversely related to the variance of noise each sensor exhibits. Usually, one can calibrate the sensors to understand these noise variances and as long as the senor is not compromised or malfunctioning, the sensor will behave as expected. It is desirable to collect information from human sources that are known to report accurately, but the number of such authenticated sources is small, and in many cases such sources are unavailable. It is necessary to collect information from unreliable sources that introduce errors due to internal group biases, incompetence, or malicious intentions [1].

This work focuses on the problem of determining the reporting behaviors of possibly human sources to enable fusion. It assumes that the semantics of the reports are already grounded so that reports from multiple sources can be associated to the same proposition. The actual problem of machine interpretation of a report and generating computer processable propositional elements is an open research problem being tackled by the natural language processing community and is beyond the scope of this paper. Here, a full report by a source breaks down into elemental reports about a given proposition such as the degree to which 'John distributes good medicine.' Imagine that you represent a non-Government organization that just started operating in John's province to provide disaster relief for a historic flood event that is still emerging. You know nothing about John and are relying on various sources to review their own or second hand experience about how often the medicine that John distributes is good or bad. Clearly, some history of past reports generated by these sources is necessary to properly fuse their current reports.

Generally, the source reporting history considers an aggregate of propositional reports from the various sources. For instance, in our example, another proposition is that the Believer's Church will support your NGO. The problem is that your NGO's primary mission is to represent a minority that is considered to be sinners by the Believer's Church To understand the degree to which the church will be supportive, you need to collect reports from multiple sources report on how well this church helped similar organization in the past. The collection of such propositional reports from the various sources provides a history to evaluate the quality of these sources.

Recently, research from the reputation systems, trust, and crowdsourcing literature has developed methods to characterize the behavior or information sources [1], [2]. These efforts can be divided in cases where the reports can eventually be authenticated and when they cannot. In the first case, it could be established after the fact whether or not a source reported accurately. The beta reputation system builds up a beta distribution to represent uncertain probabilistic belief that the source will report correctly [3]. Then, subjective logic provides the framework to discount new reports from this source and to fuse it with discounted reports from other sources [4]. Other works consider that the fusion agent may have direct evidence to form its own opinion about various proposition. The agents uses the consistency between its own opinions and the reports to form an uncertain probabilistic belief of the source's reporting behavior [2], [5]. This can be viewed as a calibration process similar to a manager initially assigning a new employee low risk tasks with quantifiable performance metrics to establish confidence before assigning more complex critical tasks. At some point, the fusion agent will need to fuse reports from sources without having the luxury of forming its own opinion from direct evidence.

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001 and W911NF-16-2-0173. 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 the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

The second thread of research considers the case when the fusion agent cannot formulate a gold standard to evaluate the source reports. The factfinding literature that begins with [6] assumes that most sources are reporting correctly. It uses the fact that claims are trustworthy if corroborated by reliable sources and a source is reliable if it reports trustworthy claims. In [7], a simple generative source model is proposed so that the source reliability and claim trustworthiness are jointly determined by an expectation-maximization (EM) approach. The model has been extended to consider dependencies either between the sources, the claims or both [1]. In most cases, the factfinding methods work only on crisp reports, e.g., either John distributes good medicine or not, rather than reports representing uncertain probabilistic beliefs. e.g., the belief about the probability that John will distribute good or bad medicine for a given instance.

To the best of our knowledge, all of the works that consider sources reporting uncertain probabilistic beliefs assume the various propositional elements are statistically independent of each other. In general, this assumption does not hold. For instance, consider that it is known that John is a member of the Believer's Church. Therefore, he is influenced by the church's ideology, and the proposition of John distributing good medicine to your NGO is linked to the proposition that the Believer's Church will support your NGO as well. When determining the behavior of information sources, it may be important to consider these correlations.

This paper investigates the performance improvements and computational costs for considering the correlations between various propositions. It builds upon our prior work of source trust estimation that assumes the propositions are statistically independent [8]. The general framework for source trust estimation is first established, and then the paper considers the special case that the propositions are actual the same, i.e., completely correlated. This is done for two reasons. First, the framework simplifies greatly for this case. Second, complete correlation represents an extreme "worst case" scenario for the independence assumption. We hypothesize that the performance gap between methods that consider correlations and that do not is greatest in this case. It is possible that a source can be tasked to report on multiple propositions that are logically equivalent but framed different. The effects for framing bias is well documented [9]. Therefore, it is possible for a source to report propositions differently not realizing that they are actually equivalent. Nevertheless, practical application of the specific methods developed in this work are out scope for this paper. This work is simply the first step to understand whether or not going beyond the independence assumption is necessary.

This paper is organized as follows. Section II reviews subjective logic that underpins the mathematical framework, and Section III reviewers our past effort to develop source behavior opinions. Then, Section IV develops the trust estimation framework over correlated propositions. Experimental evaluation determines the performance tradeoffs between the new and old frameworks in Section V. Finally, Section VI provides concluding remarks.

## II. Subjective Logic

Subjective logic extends probabilistic logic by incorporating uncertainty. We refer the reader to [4] for a formal treatment of subjective logic. This section provides a brief discussion to provide the context for establishing the trust estimation framework. Subjective logic deals with uncertain probabilistic beliefs in propositions. Here, a proposition is as statement about the degree to which a particular attribute of an object can manifest itself with a given set of values. More precisely, the attribute of the object can take on one of $K$ mutually exclusive values; there is an underlying probability for each value to manifest for a given instantiation; and each full observation of the attribute value represents an independent instantiation. A subjective opinion represents incomplete knowledge about these manifestation probabilities as a result of observations (partial or full). Mathematically, the opinion $\omega_X$ about proposition $X$ is composed of non-negative terms $b_{X,k}$ for $k \in \mathbb{X}$ and $u_X$ that sum up to one, i.e.,

$$u_X + \sum_{k=1}^{K} b_{X,k} = 1, \tag{1}$$

representing the belief in the attribute values and uncertainty, respectively. Note that $\mathbb{X} = \{1, \ldots, K\}$ is the frame of discernment (or possible attribute values) for proposition $X$. The subjective opinion also includes non-negative base rates $a_{X,k}$ for $k \in \mathbb{X}$ that sum to one and the prior weight $W$. The base rate and prior weight represent prior knowledge of the manifestation probabilities in the absence of any evidence, i.e., uncertainty $u_X = 1$.

The subjective opinion maps to a Dirichlet distribution for the manifestation probabilities

$$f_\beta(\mathbf{p}_X; \boldsymbol{\alpha}_X) = \frac{1}{B(\alpha_{X,1}, \ldots, \alpha_{X,K})} \prod_{k=1}^{K} p_{X,k}^{\alpha_{X,k}-1}, \tag{2}$$

where $B(\cdot)$ is the beta function

$$B(\alpha_{X,1}, \ldots, \alpha_{X,K}) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_{X,k})}{\Gamma\left(\sum_{k=1}^{K} \alpha_{X,k}\right)}, \tag{3}$$

$\Gamma(\cdot)$ is the gamma function, and the Dirichlet parameters are determined from the opinion by

$$\boldsymbol{\alpha}_X = \frac{W\mathbf{b}_X}{u_X} + W\mathbf{a}_X. \tag{4}$$

Note that for the beliefs to be non-negative, the Dirichlet parameters are lower bounded by $W\mathbf{a}_K$. In this work, we set $W = K$ and $a_k = 1/K$ for $k \in \mathbb{X}$ so that the absence of evidence translates to a uniform distribution over the simplex of possible values of $\mathbf{p}_X$.

The projected probability for a subjective opinion

$$\hat{\mathbf{p}}_X = \mathbf{b}_X + u_X\mathbf{a}_X \tag{5}$$

represents an estimate of the manifestation probabilities for the attributes. It is equivalent to the expected value of $\mathbf{p}_X$ for the opinion's corresponding Dirichlet distribution. Likewise, the uncertainty is inversely related to the Dirichlet strength $S_X = \sum_{k=1}^{K} \alpha_{X,k}$, i.e.,

$$u_X = \frac{W}{S_X}, \tag{6}$$

and represents the spread of the Dirichlet distribution.

Subjective logic provides a mechanism to reason over opinions for multiple propositions, which are statements about different attributes for a single object, the same attribute for different objects or more generally different attributes for different objects. The underlying manifestation probabilities for the different propositions are in general not independent. The operations in subjective logic are designed to infer one propositions from others by taking into account the underlying second-order distributions of the underlying manifestation probabilities [4]. These operations are generally designed to follow the Dirichlet interpretation in the mean, but the uncertainty usually follows an uncertainty maximization principle to limit belief commitments. In our work, we have develop alternative operators that determine the uncertainty to best match the mean and variance of the inferred distribution [10].

Is is well known that the Dirichlet distribution is conjugate prior of the multinomial distribution. This means that belief and uncertainty in the subjective opinion can be interpreted as the number of direct observations for each attribute value, i.e.,

$$\mathbf{r}_x = \frac{W\mathbf{b}_X}{u_X} \tag{7}$$

from a set of independent trials. This paper refers to $\mathbf{r}_X$ as the evidence behind the subjective opinion. It is not necessarily the case that the subjective opinion originated from direct evidence. In fact, the operations in subjective logic provides the mechanism to infer the effective evidence of one opinion from others [4]. Using our moment matching approach, we have derived methods to update the opinion (or effective evidence) in light of partial observations where only the likelihood of attribute values can be observed for the manifestations; rather than the actual values [10]. This updating method underpins the source trust estimation techniques presented in the two following sections.

To consider inferences over multiple statistically dependent propositions, it becomes necessary to form opinions over the cross product of all the attribute values. In general, the underlying frame of discernment does not scale. However, by taking advantage of Markovian structure, it is possible to represent the manifestation probabilities as a Bayesian network. However, the root and conditional probabilities are uncertain based on limited evidence. The concept of the subjective Bayesian network was recently introduced where the probabilities are characterized by subjective opinions (or by Dirichlet distributions) [11]. The subjective Bayesian network for $T$ propositions represents the distribution of the joint manifestation probabilities for the attribute values of all propositions. The distribution is an aggregate $\mathcal{A}$ of the subjective opinions for conditional probabilities[1]

$$f(\mathbf{p}_{1,\ldots,T}; \mathcal{A}) = \prod_{t=1}^{T} \prod_{x \in \mathbb{X}|\text{pa}_t|} f_\beta(\mathbf{p}_{t|\text{pa}_t=x}; \boldsymbol{\alpha}_{t|\text{pa}_t=x}). \tag{8}$$

Efficient methods to infer subjective opinions for individual propositions (e.g., knowledge of $\mathbf{p}_t$ associated to $X_t$) is an active area or research. Recently, we generalized belief propagation to infer opinions over binary propositions in a singly-connected graph [12]. The SBN can represent the complete statistical knowledge binding multiple propositions by a fusion agent and underpins the correlated trust estimation framework in Section IV.

## III. Review of Trust Estimation

This section reviews and builds upon slightly our past work to develop source trust estimates, i.e., subjective opinions about the reporting behavior of each source [8]. The fusion agent must understand how a source may (or may not) distort its reports in light of what it should have observed. In order to form its trust opinion of the sources, the fusion agent uses the past reporting history in cases where it has its own direct opinion of the propositions. This is a calibration process similar to how a manager forms an opinion of a new employee through a series of non-critical assignments. In critical situations, the fusion agent may have to incorporate the subjective opinions of various information sources without having any direct evidence of the particular proposition.

Our work considers a set of canonical behaviors for the sources. For a particular proposition, the source uses its evidence to form its subjective opinion. The source can choose to send this opinion unaltered to the fusion agent. On the other hand, the source many decide to manipulate its opinion. In this paper, we consider binary ($K = 2$) propositions where the sources can decide to either flip its opinions (i.e., $b_1 \leftrightarrow b_2$) or transmit a completely random opinions. Note that the flipped opinion does carry information to the fusion agent in as much the fusion agent can recognize that the opinion has been flipped. On the other hands, a random opinion carries no information, and the fusion agent should ignore the report. For any report, the fusion agent does not know whether the source reported its opinion correctly, flipped it, or acted randomly. By forming a source behavior opinion of each source, the fusion agent can better fuse the reports of multiple behaviors as shown in [8].

Consider that a source reports an opinion $\omega_t^s$ about the $t$-th proposition, and the fusion agent has its own opinion about the proposition $\omega_t^f$. Alternatively, one can represent these opinions in evidence space where $r_{t,1}^s$ and $r_{t,2}^s$ represents the amount of positive and negative evidence, respectively, reported by the source as computed by (7). Similarly, $r_{t,1}^f$ and $r_{t,2}^f$ represents the evidence observed by the fusion agent. Given $p_t$ as the probability for the positive attribute to manifest, the likelihood

[1] In this paper, $f_\beta(\cdot)$ represent a Dirichlet distribution, and $f(\cdot)$ represents an arbitrary distribution.

of the source observing its evidence in the case of truthful ($b_t = 1$) or flipping ($b_t = 2$) reporting is

$$p(\omega_t^s|p_t, b_t) = p_t^{g_{b_t,1}(\mathbf{r}_t^s)}(1-p_t)^{g_{b_t,2}(\mathbf{r}_t^s)}, \quad (9)$$

where $g_{b,1}(\mathbf{r})$ and $g_{b,2}(\mathbf{r})$ are the inverse transformations that converts reported evidence back into the sources observed behavior such that $g_{1,1}(\mathbf{r}) = r_1$ and $g_{1,2}(\mathbf{r}) = r_2$ is for accurate and $g_{1,1}(\mathbf{r}) = r_2$ and $g_{1,2}(\mathbf{r}) = r_1$ is for flipping reporting. The value $p_t$ is only known by the fusion agent within a beta distribution[2], given by its own direct evidence. The expected likelihood of the source correctly reporting given the fusion agent's knowledge of $p_t$ is [3]

$$P(\omega_t^s|b_t, \omega_t^f) = \int p_t^{g_{b_t,1}(\mathbf{r}_t^s)}(1-p_t)^{g_{b_t,2}(\mathbf{r}_t^s)} f_\beta(p_t|\omega_t^f) dp_t. \quad (10)$$

Noting that

$$f(p_t|\omega_t^f) = \frac{p_t^{r_{t,1}^f}(1-p_t)^{r_{t,2}^f}}{B(r_{t,1}^f+1, r_{t,2}^f+1)}, \quad (11)$$

leads to

$$P(\omega_t^s|b_t, \omega_t^f) = \frac{B(g_{b_t,1}(\mathbf{r}_t^s) + r_{t,1}^f + 1, g_{b_t,2}(\mathbf{r}_t^s) + r_{t,2}^f + 1)}{B(r_{t,1}^f+1, r_{t,2}^f+1)}. \quad (12)$$

Now when the source reports randomly ($b_t = 3$), there is no knowledge from the fusion agent about the underlying probability $p_t$ that generated the source's opinion. Therefore knowledge of $p_t$ is vacuous and its distribution in light of the fusion agents opinion is treated as a non-informative uniform distribution, i.e., the beta distribution with no evidence, i.e., $\mathbf{r}_t = \mathbf{0}$. In that case,

$$p(\omega_t^s|b_t = 3, \omega_t^f) = B(r_{t,1}^s+1, r_{t,2}^s+1). \quad (13)$$

While this paper considers a $K = 3$ behavior opinion of the sources, it is straightforward to consider other behaviors. In fact, discovery methods exists to learn the transformations [8].

The likelihoods represent partial observations for the three reporting behaviors. Given that the current subjective opinion of the fusion agent about the source behavior $\omega_{t-1}^{fs}$ due to the first $t-1$ propositions, the actual posterior distribution for behavior manifestation probabilities $\boldsymbol{\rho}_t$ is

$$\begin{aligned} f(\boldsymbol{\rho}|\omega_t^{fs}) &\propto \sum_{b_t=1}^{K} p(\omega_t^s|\omega_t^f, b_t) p(b_t|\boldsymbol{\rho}) f_\beta(\boldsymbol{\rho}|\omega_{t-1}^{fs}), && (14) \\ &= \left(\sum_{b_t=1}^{K} \ell_t^{(b_t)} \rho_{b_t}\right) f_\beta(\boldsymbol{\rho}|\omega_{t-1}^{fs}). && (15) \end{aligned}$$

Note that for simplification of the notation, $\ell_t^{(b_t)} = p(\omega_t^s|b_t, \omega_t^f)$ in the sequel. In [10], the posterior distribution $f(\boldsymbol{\rho}|\omega_t^{fs})$ is approximated by a Dirichlet distribution using the method of moments. Then, $f_\beta(\boldsymbol{\rho}|\omega_t^{fs})$ matches the true posterior in the means and matches the variances in the mean squared sense. The exact closed form update equations for the opinion update is given in [10]. Initially, $\omega_0^{fs}$ is the vacuous opinion ($u_0^{fs} = 1$) and the updating is performed sequentially in order of the propositions. This paper refers to this behavior opinion extraction method as *Seq indTrustEst*.

It is more accurate to impose the Dirichlet approximation after all the training proposition reports have been processed. In this case, it is easy to see that the actual posterior is

$$\begin{aligned} f(\boldsymbol{\rho}|\omega_T^{fs}) &\propto \prod_{t=1}^{T}\left(\sum_{b_t=1}^{K} \ell_t^{(b_t)} \rho_{b_t}\right) f_\beta(\boldsymbol{\rho}|\omega_0^{fs}) && (16) \\ &= \sum_{b_1=1}^{K} \cdots \sum_{b_T=1}^{K} \left(\prod_{t=1}^{T} \ell_t^{(b_t)}\right)\left(\prod_{t=1}^{T} \rho_{b_t}\right) f_\beta(\boldsymbol{\rho}|\omega_0^{fs}). && (17) \end{aligned}$$

The posterior is a mixture of Dirichlets and the means and variances for $\boldsymbol{\rho}$ can be computed in closed form. Again, the method of moments is used to approximate the posterior by a single Dirichlet distribution to form the subjective opinion. A recursive programming method is presented in [10]. The problem is that is that the computational complexity grows as $O(K^t)$ and [10] provides empirical evidence that the behavior profile error is not significantly lower than Seq indTrustEst.

As a mixture of Dirichlets, many of the terms in (17) repeat the same Dirichlet components. Therefore, the posterior can be represent with much fewer terms as

$$f(\boldsymbol{\rho}|\omega_T^{fs}) = \left(\sum_{\mathbf{n}} \ell_T(\mathbf{n}) \rho_1^{n_1} \cdots \rho_K^{n_K}\right) f_\beta(\boldsymbol{\rho}|\omega_0^{fs}) \quad (18)$$

where $\mathbf{n}$ is a vector of non-negative integers such that $\sum_{k=1}^{K} n_k = T$ and $\ell_T(\mathbf{n})$ is the sum of likelihoods for the hypotheses such that the $k$-th behavior is selected $n_k$ times, i.e.,

$$\ell_T(\mathbf{n}) = \sum_{\mathbf{b}\in\mathcal{N}} \prod_{t=1}^{T} \ell_t^{(b_t)}, \quad (19)$$

where $\mathcal{N} = \left\{b_1, \ldots, b_T | \sum_{t=1}^{T} I_k(b_t) = n_k \text{for } k = 1, \ldots, K\right\}$ and $I_k$ is the $k$-th indicator function

$$I_k(x) = \begin{cases} 1 & \text{if } x = k, \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

It is easy to see that the posterior (18) can be built up sequentially one proposition at a time because the composite likelihood $\ell_t(\mathbf{n})$ can be composed from $\ell_{t-1}(\mathbf{n})$ and $\ell_t^{(b_t)}$ for $b_t = 1, \ldots, K$ because of the statistical independence of the propositions. As a result, the complexity of composing (18) grows as $O(t^{K-1})$ with respect to the number of propositions. Determining the Dirichlet distribution of (18) is the same as in [10] for (17), except sorting through significantly less terms. This paper incorporates this more efficient method to update the source behavior opinion over a group of $T$ propositions and refers to the method as *indTrustEst*.

[2] The beta distribution is the special case of a Dirichlet distribution when $K = 2$.

[3] The terminology $f_\beta(\cdot|\omega_X)$ is used interchangeably with $f_\beta(\cdot|\mathbf{r}_X + \mathbf{1})$ to represent the distribution for the manifestation probabilities given by the subjective opinion about proposition $X$.

## IV. Trust Estimation over Correlated Propositions

In the previous section, the aggregation of propositions assume that for a given instantiation of proposition attribute values, the manifest probability for the values is the product of the manifest probabilities for each individual proposition. In general, the probability that one attribute manifests as a particular value does depend on the other attribute values that have manifested. This section extends our previous work to account for the fusion agent's knowledge of these dependencies. Probabilistic graphical models provides tools to exploit inherent structure in the joint probability, and one can view this joint probability as a Bayesian network. As discussed in Section II the fusion agents knowledge about this network forms a subjective Bayesian network where the distribution of the joint probability is represented by (8). For ease of presentation, the discussion focuses on the reporting of binary propositions.

The source still reports their subjective opinions about each proposition separately. In doing so, it can exhibit one of the $K$ behaviors for each report. In light of the fusion agent's subjective knowledge of the dependency structure, the likelihood of a particular sequence of reporting behaviors for the set of propositions (i.e., a behavior path) is determined. To this end, the overall likelihood is the expected likelihood of the individual propositional evidences in light of the distribution of the joint manifestation probabilities given by the subjective Bayesian network of the fusion agent $f(\mathbf{p}_{1,\ldots,T}|\Omega^f)$. Specifically,

$$p(\omega^s_{1:T}|\mathbf{b}_{1:T},\Omega^f) = \int \prod_{t=1}^{T} p_t^{g_{b_t,1}(\mathbf{r}^s_t)}(1-p_t)^{g_{b_t,2}(\mathbf{r}^s_t)} h_{b_t}(p_t,q_t,\nu_t) \cdot f(\mathbf{q}_{1,\ldots,T}|\Omega^f)\prod_{t=1}^{T} f_\beta(\nu_t|\mathbf{1})d\mathbf{p}d\mathbf{q}d\boldsymbol{\nu}. \tag{21}$$

Like the independent processing case, the $g(\cdot)$ functions transforms the reported evidence back into the true evidence in light of the behavior hypothesis. The term

$$h_{b_t}(p_t,q_t,\nu_t) = \begin{cases} \delta(p_t - \sum_{k_1,\ldots,k_T} q_{k_1,\ldots,t,\ldots,k_T}) & \text{for } b_t \neq K \\ \delta(p_t-\nu_t) & \text{for } b_t = K \end{cases} \tag{22}$$

associates the manifestation probability of the reported propositions either to the marginal manifestation probability from the fusion agents's subjective Bayesian network or to vacuous knowledge about the probability, i.e., the uniform distribution $f_\beta(\nu_t|\mathbf{1})$ for the random behavior hypothesis. Note that (21) represents the full likelihood for the source to exhibit the sequence of behaviors $b_1,\ldots,b_T$. In general (21) is difficult to calculate.

We consider a special case that the fusion agents knows that the propositions are all semantically the same. This is the case of complete correlation where the subjective Bayesian network simplifies to

$$f(p_1,\ldots,p_T|\Omega^f) = \prod_{t=2}^{T}\delta(p_t-p_1)f_\beta(p_1|\omega^f_1) \tag{23}$$

so that $p_T = p_{t-1} = \cdots = p_1$. Insertion of (23) into (21) yields

$$p(\omega^s_{1:T}|\mathbf{b}_{1:T},\Omega^f) = \frac{B(e_1(\mathbf{b}_{1:T}) + r^f_{1,1}+1, e_2(\mathbf{b}_{1:T}) + r^f_{1,2}+1)}{B(r^f_{1,1}+1, r^f_{1,2}+1)} \cdot \prod_{t:b_t=K} B(r^s_{t,1}+1, r^s_{t,2}+1) \tag{24}$$

where the accumulated evidence for the attribute values $\kappa \in \mathbb{X}_t$ is

$$e_\kappa(\mathbf{b}_{1:T}) = \sum_{t:b_t\neq K} g_{b_t,\kappa}(\mathbf{r}^s_t). \tag{25}$$

Now the likelihood represents one particular behavior path, and the posterior for the source behavior manifestation probability $\boldsymbol{\rho}$ aggregates all possible paths

$$f(\boldsymbol{\rho}|\omega^{fs}_T) = \left(\sum_{b_1=0}^{K}\cdots\sum_{b_T=0}^{K}\ell_T^{(\mathbf{b}_{1:T})}\rho_{b_1}\cdots\rho_{b_T}\right) f_\beta(\mathbf{p}|\omega^{fs}_0), \tag{26}$$

where $\ell_T^{(\mathbf{b}_{1:T})} = p(\omega^s_{1:T}|\mathbf{b}_{1:T},\Omega^f)$ for ease of presentation. Again, the posterior is a mixture of Dirichlet distributions. For $K$ behaviors, the posterior includes the mixture of $K^T$ Dirichlet distributions, and thus the computational complexity grow exponentially with the number of propositions $T$. Like the independent proposition case, the means and variances for the mixture distribution are easily computed, and the method of moments is employed to approximate the posterior to a Dirichlet distribution to form a subjective opinion for the source behavior profile. We refer to this methods as *corTrustEst*.

In this paper, $K = 3$ canonical behaviors are considered. While there are only $(T+2)(T+1)/2$ unique Dirichlet distributions in the mixture, one must enumerate the likelihood for each possible path $\mathbf{b}_{1:T}$ separately, because the likelihoods for various values of $b_t$ are dependent on the values for $\mathbf{b}_{1:t-1}$. This is in contrast to independent propositions in the previous section. Still, one can sequentially build up the mixture and remove terms with very low likelihoods to reduce the computational complexity.

Algorithm 1 shows the steps of *seq corTrustEST*, the sequential method to build up the Dirichlet mixture and determine the effective source behavior opinion. Steps 3-8 updates the accumulated evidences and likelihoods by polyfurcating the current set of behavior paths to consider the $K$ possible behaviors for reporting the $t$-th proposition. This forms the Dirichlet mixture in step 9. In step 10, the terms whose likelihood values $\ell$ divided by the maximum likelihood is less than a threshold $\lambda$ are pruned. In this paper, $\lambda$ is set to a value of $1e-4$. After processing all $T$ source opinions, the means and variances are computed in steps 13-16. A good Dirichlet strength that "best" matches the variances and does not lead to negative belief is determined in steps 17-20. The final step determines the updated source behavior opinion. When the threshold $\lambda = 0$, none of behavior paths are pruned, and Algorithm 1 becomes equivalent to full corTrustEst.

**Algorithm 1** **Seq corTrustEst**

**Input:**$\omega_0^{fs}$,$\omega_1^{f}$ and $\omega_t^s$ for $t = 1, \ldots, T$. **Output:**$\omega_T^{fs}$.

Set $\mathbf{b}_{1,1:0} = []$, $N = 1$, $e_\kappa(\mathbf{b}_{1:0}) = 0$ for $\kappa \in \mathbb{X}$.<br>
**for all** $t = 1, \ldots, T$, **do**<br>
  **for all** $i = 1, \ldots, N$ **do**<br>
    $\mathbf{b}_{i+N(k-1),1:t} = [\mathbf{b}_{i,1:t-1} \quad k]$ for $k = 1, \ldots, K$.<br>
    $e_\kappa(\mathbf{b}_{i+N(k-1),1:t}) = e_\kappa(\mathbf{b}_{i,1:t-1}) + g_{k,\kappa}(\mathbf{r}_t^s)$ for $\kappa = 1, 2$.

$$\ell(\mathbf{b}_{i+N(k-1),1:t}) = \ell(\mathbf{b}_{i,1:t})\cdot \cdot \frac{B(e_1(\mathbf{b}_{i,1:t}) + r_{1,1}^f + 1, e_2(\mathbf{b}_{i,1:t}) + r_{1,2}^f + 1)}{B(e_1(\mathbf{b}_{i,1:t-1}) + r_{1,1}^f + 1, e_2(\mathbf{b}_{i,1:t-1}) + r_{1,2}^f + 1))},$$

    for $k = 1, \ldots K - 1$.<br>
    $\ell(\mathbf{b}_{i+N(K-1),1:t}) = \ell(\mathbf{b}_{i,1:t}) B(r_{t+1}^s + 1, s_{t+1}^s + 1)$.<br>
  **end for**<br>
  $f(\boldsymbol{\rho}|\Omega_t) = \left(\sum_{i=1}^{KN} \ell(\mathbf{b}_{i,1:t}) p_{b_{i,1}} \cdots p_{b_{i,t}}\right) f_\beta(\mathbf{p}|\omega_0^{fs})$.<br>
  Prune all terms such that $\ell(\mathbf{b}_{i,1:t}) < \lambda \max_i \ell(\mathbf{b}_{i,1:t})$, let $N$ be the number of surviving terms, and reindex the surviving terms from $1, \ldots, N$.<br>
**end for**<br>
**for all** $k = 1, \ldots, K$ **do**<br>
  $n_{i,k} = \sum_{t=1}^{T} I_k(b_{i,t})$ for $i = 1, \ldots, N$.<br>
  $d = \sum_{i=1}^{N} \frac{B(n_{i,1}+r_{0,1}^{fs}+1, \ldots, n_{1,K}+r_{0,K}^{fs}+1)}{B(r_{0,1}^{fs}+1, \ldots, r_{0,K}^{fs}+1)}$.<br>
  $m_k = \sum_{i=1}^{N} \frac{B(n_{i,1}+r_{0,1}^{fs}+I_1(k)+1, \ldots, n_{1,K}+r_{0,K}^{fs}+I_K(k)+1)}{d \cdot B(r_{0,1}^{fs}+1, \ldots, r_{0,K}^{fs}+1)}$.<br>
  $v_k = \sum_{i=1}^{N} \frac{B(n_{i,1}+r_{0,1}^{fs}+2I_1(k)+1, \ldots, n_{1,K}+r_{0,K}^{fs}+2I_K(k)+1)}{d \cdot B(r_{0,1}^{fs}+1, \ldots, r_{0,K}^{fs}+1)}$.<br>
  $s_k = \frac{m_k - v_k}{v_k - m_k^2}$.<br>
**end for**<br>
$s = 0.5 \min_k s_k + 0.5 \frac{\sum k{=}1K m_k(1-m_k)(m_k-v_k)}{\sum_{k=1}^{K} m_k(1-m_k)(v_k-m_k^2)}$.<br>
$s = \max(s, \frac{1}{m_1}, \ldots, \frac{1}{m_K})$.<br>
$b_{T,k}^{fs} = m_k - \frac{1}{s}$ for $k = 1, \ldots, K$ and $u_T^{fs} = \frac{K}{S}$.

## V. Experimental Results

This section analyzes the potential advantages of accommodating the statistical dependencies between propositions used by a fusion agent to determine its subjective opinion about the source behavior profile. As mentioned earlier, we compare previous opinion formation methods that assume the propositions are independent to formation methods that fully account for the statistical dependency. Specifically, this paper considers the case that the propositions are all fully correlated in that the manifestation probabilities for each attribute value are the same. This represents and extreme case that we hypothesize represents the biggest performance advantage for the new formation methods that account for the dependencies.

The simulations create synthetic opinions from a source reporting opinions about $N_{cor}$ binary propositions to a fusion agent. Because the fusion agent understands the propositions all share the same manifestation probabilities, it forms a single opinion from its own internal observations. To this end, the underlying manifestation probability $p_1$ for $k = 1$ is randomly drawn from a uniform distribution over $[0, 1]$ ($p_2 = 1 - p_1$ for binary propositions). Then the source takes $N_{cor}$ sets of full observation to form $N_{cor}$ different subjective opinions. For each set, the source performs $N_O$ full observations where $N_O$ is drawn uniformly over $[0, 100]$ and $\mathbf{p} = [p_1, 1 - p_1]$ are the manifestation probabilities for the two attribute values. For each of the $N_{cor}$ opinions, the source decides to report it to the fusion agent as is, flip it, or created a different opinion using another randomly generated value for $p_1$ based upon the sources behavior manifestation probabilities $\boldsymbol{\rho}$, whose values are set initially by drawing uniformly over the simplex. The fusion agent also creates its own single subjective opinion similar to the source. It then compares its opinion to the $N_{cor}$ reported opinions to form its subjective source behavior opinion. The baseline approach is for the fusion agent to repeat its own opinion $N_{cor}$ times and run either seq indTrustEst or full indTustEst as described in Section III. The fusion agent also executes corTrustEst and pruned cor TrustEest to from a source behavior opinion.

The process is repeated over 100 rounds to refine the source behavior opinions further. Each round represents another set of completely correlated $N_{cor}$ reports where the propositions across rounds are statistically independent. For each round the manifestation probabilities for the set of propositions are redrawn, i.e., each round represents completely different propositions. The 100 rounds represent the calibration of a single source. Finally, the entire process is repeated 1000 times where new behavior probabilities $\boldsymbol{\rho}$ are drawn for the source. This creates instances of subjective opinion formation about 1000 different sources.

The four opinion formation methods are evaluated by two performance criteria. The first is root mean squared (RMS) error between the projected probabilities of the opinion (see (5)) and the ground truth $\boldsymbol{\rho}$. The second characterizes the uncertainty of the opinion. Each opinion can be converted into a beta distribution for the values $\rho_1$, $\rho_2$ and $\rho_3$, and confidence bounds can be obtained to capture a desired $\gamma$ percent of the distribution. The actual ratio $\tilde{\gamma}(\gamma)$ of ground truth that falls within these bounds is determined. If the uncertainty (i.e., distribution), is correctly characterized then $\tilde{\gamma}(\gamma) = \gamma$. The deviation from this ideal curve illustrates how well (or not) the uncertainty in the opinion represents the true uncertainty. More details about the generating the curves can be found in Appendix 3 of [12]. In this paper, we define the mean DEsired Confidence BOund Divergence (DeCBoD) as

$$\text{mean DecBod} = \int_0^1 |\tilde{\gamma}(\gamma) - \gamma| d\gamma \approx \frac{1}{N} \sum_{i=1}^{N} |\tilde{\gamma}(\gamma_i) - \gamma_i|. \quad (27)$$

Similarly, a max DecBOD can be defined. In the approximation used in this paper, the $\tilde{\gamma}(\cdot)$ curve is calculated by sampling the desired significance $\gamma$ values uniformly over 100 points.

The DeCBOD values provide statistics to easily assess how well the $\tilde{\gamma}(\cdot)$ curve fits the ideal. Figure 1 and Table I illustrates

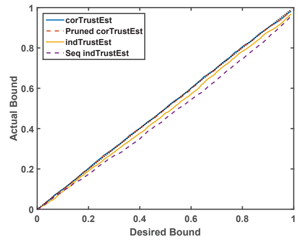
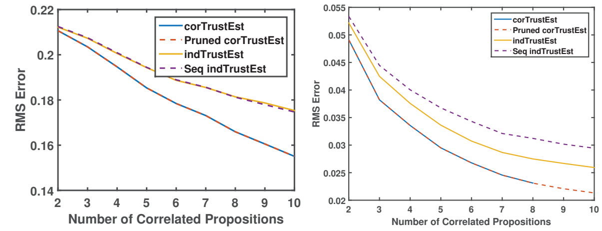



Fig. 1. Actual versus desired confidence bounds for various methods to establish source behavior opinions after processing one round of 10 correlated propositions. These curves are used to establish the DeCBoD statistic, which are provided in Table I.

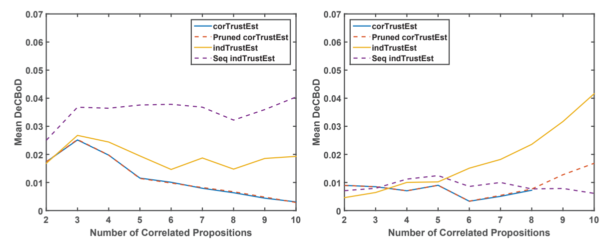
TABLE I
DECBOD VALUES OF SOURCE BEHAVIOR OPINIONS AFTER UPDATE OF 10 CORRELATED PROPOSITIONS.

| Statistic | corTrustEst | Pruned corTrustEst | indTrustEst | Seq indTrustEst |
|---|---|---|---|---|
| Mean | 0.0031 | 0.0029 | 0.0193 | 0.0404 |
| Max | 0.0083 | 0.0100 | 0.0303 | 0.0617 |




(a) (b)

Fig. 2. Accuracy of the source behavior opinions versus the number of correlated propositions: (a) RMS error after the first round and (b) RMS erro after the last round.




(a) (b)

Fig. 3. Quality of the uncertainty characterization of the source behavior opinions versus the number of correlated propositions: (a) Mean DeCBoD after the first round, and (b) mean DecBoD after the last round. Note that the curves for max DecBoD are similar.

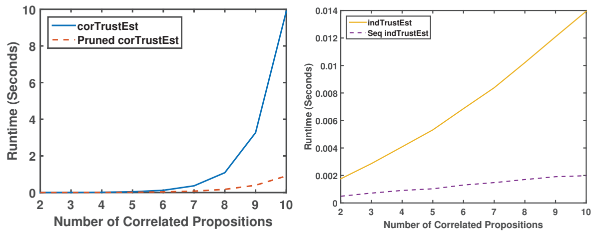
this assessment. The figure plots the $\tilde{\gamma}(\cdot)$ generated by the four behavior formation methods when $N_{cor} = 10$ after one round of processing, and the table provides the corresponding DeCBoD values. The correlated processing methods resulted in curves that lie on the ideal linear curve, i.e, $\tilde{\gamma}(x) = x$ . The indTrustEst method is not far off from ideal, and the seq indTrustEst method is clearly further from ideal. One can see that the DeCBoD numbers in the tables indicate these curve fits to the ideal. Overall, the figure and table demonstrates that the behavior opinion formation methods that consider the correlation between reported propositions better captures uncertainty. For the independent methods, it is better to build up the Dirichlet mixture rather than performing the single Dirichlet approximation after each sequential proposition update.

Figure 2 compares the accuracy of the projected probabilities of the behavior opinions after the first and last round of updates for various values of the number of correlated propositions $N_{cor}$ processed per a round. As expected the error decreases as $N_{cor}$ grows. After the first round, the faster pruned and sequential update methods perform indistinguishably worse than their full mixture processing counterparts. The gap between the correlated and independent methods grows with $N_{cor}$. Still, even for $N_{cor} = 10$, the error using the correlated methods is only 11% smaller than that of the independent methods. After the last round, the seq indTrustEst performs slightly worse than the full indTrustEst method. Furthermore, the gap between the independent and correlated method grows more. Now for $N_{cor} = 10$, the error reduction of the correlated methods is about 27% and 18% with respect to seq and regular indTrustEst, respectively.

Figure 3 illustrates the quality of the uncertainty characterization in the behavior opinions after the first and last rounds for various values of $N_{cor}$. After the first round, the uncertainty characterization is best for the correlated methods, and there is little distinction between the pruned and full methods. The indTrustEst method provides better uncertainty characterization than seq intTrustEst after the first round. For some reason the seqTrustEst methods exhibits better uncertainty characterization (and the best for $N_{cor} = 10$) after the last round. However, this comes at a cost of higher RMS error.

Figures 4(a) and 4(b) provides the runtimes of the two correlated and two independent methods, respectively. The exponential growth of complexity as $N_{cor}$ increases is clearly evident for the correlated methods. The advantages for the pruning is also evident as pruned corTrustEst is about 11 times faster than corTruestEst for $N_{cor} = 10$. Nevertheless, indTrustEst and seq indTrustEst are over 65 and 450 times faster, respectively, than pruned corTrustEst for $N_{cor} = 10$.

The whole process of forming the behavior opinions for sources is to enable fusion of reports from multiple sources; especially in cases when the fusion agent has no direct evidence of the specific proposition. The big question is whether or not the more accurate behavior estimation that considers the correlation between training propositions has any significant impact on fusion, especially in light of the high computational cost. The traditional method is to use the behavior opinions to discount the reported opinion and then incorporate consensus fusion [4]. Here we use our method from [8] were the fused




(a) (b)

Fig. 4. Average runtime to determine source behavior opinions versus the number of correlated propositions to process: (a) The methods that consider correlated propositions, and (b) the methods that assume independent propositions.

TABLE II
AVERAGE FUSION RESULTS IN LIGHT OF SOURCE BEHAVIOR OPINIONS OBTAINED BY EITHER CONSIDERING CORRELATION OR ASSUMING INDEPENDENCE OVER SETS OF TEN PROPOSITIONS DURING TRAINING.

| | | pruned corTrustEst | | seq indTrustEst | |
|---|---|---|---|---|---|
| Num Sources | Round | RMS Error | mean DeCBoD | RMS Error | mean DeCBoD |
| 10 | First | 0.2528 | 0.0676 | 0.2547 | 0.0634 |
| 10 | Last | 0.1782 | 0.0855 | 0.1775 | 0.0852 |
| 100 | First | 0.0187 | 0.0182 | 0.0198 | 0.0231 |
| 100 | Last | 0.0095 | 0.0184 | 0.0095 | 0.0164 |

opinion is formed by fitting a Dirichlet distribution to

$$f(\mathbf{p}) = \prod_{s=1}^{S} \left( \sum_{b=1}^{K} \hat{\rho}_{s,b} f_\beta(\mathbf{p}|\mathbf{g}_b(\mathbf{r}^s) + \mathbf{1}) \right) f_\beta(\mathbf{p}|\mathbf{1}). \quad (28)$$

Note that in [8], (28) is approximated by clustering and aggregation of terms to control the number of mixture terms. We grouped the 1000 sources in groups of 10 or 100. Each group then reported 100 opinions to the fusion agent, and the fusion agent combined these opinions using (28) where the projected source behavior probabilities were obtained either from the first or final round of training via the either the seq pruned corTrustEst or seq indTrustEst method when $N_{cor} = 10$. This represents the case with the largest performance gap between the two methods. Table II provides the RMS error and uncertainty characterization of the fused result. The fusion process is clearly more accurate when incorporating more sources. In either event, there is very little advantage to the more accurate correlated-based opinions when training over one round, and when training over all rounds, there appears to be no advantage.

## VI. CONCLUSIONS

This paper investigates the possible performance advantage to the consideration of existing (but ignored) dependencies between propositions used from training to form source behavior opinions. This paper considers a simple extreme case when all the training propositions are completely correlated. The consideration of the correlations does result in more accurate behavior opinions in both estimating the expected behavior probabilities and the uncertainty about these probabilities. However, the computational cost to incorporate the correlation is significant. Furthermore, the impact of the improved opinions on fusion appears to be very slight. Therefore, the improved performance does not appear to justify the higher computational cost.

The known dependencies between propositions enable exploitation of propositions whose opinions are not directly reported by the fusion agent. If this is the case, the known dependencies between propositions can be used to infer opinions of reliable sources about those propositions using deduction [13] to propagate the evidence over the subjective Bayesian network. Hence, we can have indirect evidence from the fusion agent about the ground truth. Then, the inferred opinions can be used to estimate the behaviors of the sources either by treating the collection of direct and inferred opinions as independent or by accounting for dependencies via (21) for each possible behavior path. This paper shows that the benefit of accounting for the dependencies between propositions is quite limited for the extreme case that the propositions are equivalent, and we hypothesize that the benefits are even less for more general scenarios. Nevertheless, there is still a large gain for the fusion agent to infer opinions for propositions for which it does not have direct evidence before employing trust estimation with the independent proposition assumption.

## REFERENCES

[1] D. Wang, T. Abdelzaher, and L. Kaplan, *Social Sensing: Building Reliable Systems on Unreliable Data*. Morgan Kaufmann, 2015.

[2] W. L. Teacy, M. Luck, A. Rogers, and N. R. Jennings, "An efficient and versatile approach to trust and reputation using hierarchical Bayesian modelling," *Artificial Intelligence*, vol. 193, pp. 149 – 185, 2012.

[3] A. Jøsang and R. Ismail, "The beta reputation system," in *Proceedings of the Fifteenth Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy*, June 2002, pp. 48–64.

[4] A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer, 2016.

[5] K. Regan, P. Poupart, and R. Cohen, "Bayesian reputation modeling in e-marketplaces sensitive to subjecthity, deception and change," in *Proc of the 21st Natl Conf on Artificial Intelligence*, 2006, pp. 1206–1212.

[6] X. Yin, J. Han, and P. S. Yu, "Truth discovery with multiple conflicting information providers on the web," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 6, pp. 796–808, 2008.

[7] D. Wang, L. Kaplan, and T. F. Abdelzaher, "Maximum likelihood analysis of conflicting observations in social sensing," *ACM Transactions on Sensor Networks*, vol. 10, no. 2, pp. 30:1–30:27, Jan. 2014.

[8] M. Şensoy, L. Kaplan, G. de Mel, and T. D. Gunes, "Source behavior discovery for fusion of subjective opinions," in *19th International Conference on Information Fusion*, 2016, pp. 138–145.

[9] A. Tversky and D. Kahneman, "The framing of decisions and the psychology of choice," *Science*, vol. 211, no. 4481, pp. 453–458, Jan. 1981.

[10] L. Kaplan, M. Şensoy, S. Chakraborty, and G. de Mel, "Partial observable update for subjective logic and its application for trust estimation," *Information Fusion*, vol. 26, pp. 66–83, 2015.

[11] M. Ivanovska, A. Jøsang, L. Kaplan, and F. Sambo, "Subjective networks: Perspectives and challenges," in *Proc. of the 4th International Workshop on Graph Structures for Knowledge Representation and Reasoning*, Buenos Aires, Argentina, Jul. 2015, pp. 107–124.

[12] L. M.Kaplan and M. Ivanovska, "Efficient belief propagation in second-order bayesian networks for singly-connected graphs," *International Journal of Approximate Reasoning*, vol. 93, pp. 132–152, Feb. 2018.

[13] L. M. Kaplan, M. Şensoy, Y. Tang, S. Chakraborty, C. Bisdikian, and G. de Mel, "Reasoning under uncertainty: Variations of subjective logic deduction," in *Information Fusion (FUSION), 2013 16th International Conference on*. IEEE, 2013, pp. 1910–1917.