Chapter 6
Uncertainty Characterization and Fusion of Information from Unreliable Sources

Lance Kaplan and Murat Şensoy

Abstract Intelligent systems collect information from various sources to support their decision-making. However, misleading information may lead to wrong decisions with significant losses. Therefore, it is crucial to develop mechanisms that will make such systems immune to misleading information. This chapter presents a framework to exploit reports from possibly unreliable sources to generate fused information, i.e., an estimate of the ground truth, and characterize the uncertainty of that estimate as a facet of the quality of the information. First, the basic mechanisms to estimate the reliability of the sources and appropriately fuse the information are reviewed when using personal observations of the decision-maker and known types of source behaviors. Then, we propose new mechanisms for the decision-maker to establish fused information and its quality when it does not have personal observations and knowledge about source behaviors.

Keywords Subjective logic · Unreliable sources · Fusion of information · Quality of information · Uncertainty · Beliefs

6.1 Introduction

Decision-making requires the weighing of risk and benefits in light of uncertain information. While doing so, it is important to estimate the state of the world at sufficient certainty. For a specific decision-making task, this may boil down to estimating the values or a distribution of values for a number of state variables.

L. Kaplan (✉)
RDRL-SES-A, US Army Research Laboratory, Adelphi, MD, USA
e-mail: lkaplan@ieee.org; lance.m.kaplan.civ@mail.mil

M. Şensoy
Computer Science, Özyeğin University, Istanbul, Turkey
e-mail: murat.sensoy@ozyegin.edu.tr

É. Bossé, G. L. Rogova (eds.), *Information Quality in Information Fusion and Decision Making*, Information Fusion and Data Science,
https://doi.org/10.1007/978-3-030-03643-0_6

Let us consider an intelligent agent that needs to solicit help from a person in a specific organization. Instead of asking a random person in the organization, the agent should pick a person with a high probability of accepting and fulfilling the help request. Hence, for a given person, e.g., Joe, in the organization, the agent can compute a probability distribution over possible outcomes of the request. That is, in response to the help request, Joe may help, do nothing, or undermine. These three outcomes are possible values of the state variable about Joe's behavior in response to the help request.

In this work, we adopt subjective logic [10], where opinions describe state variables. A state variable can take values from a domain. For instance, a state variable about Joe's response when help is requested can take three values: help, do nothing, or undermine. Each of these values may lead to a binary proposition, such as "Joe helps when requested," which can be either *true* or *false*. Instantiations of these propositions are observed and used to create opinions about Joe's helpfulness.

The decision-maker may have past history of the number of times Joe helped, did nothing, and undermined the organization's effort to form an opinion about Joe. From this history, the decision-maker can understand and account for the probability of Joe's behavior for the upcoming mission. The more instances of Joe's past behavior, the more certain the decision-maker is about these probabilities. In many cases, the uncertainty about Joe is too high to make a decision, and if time permits, the decision-maker should seek out more information about Joe.

In understanding Joe's tendencies, the decision-maker may have limited experience with Joe and will need to seek reports from other sources about Joe. These sources may or may not provide truthful reports about their experiences with Joe. As a result, the fusion of these reports can lead to wrong probabilities describing Joe's tendencies when help is requested. Furthermore, the decision-maker can become overconfident about these probabilities and make a poor decision.

To overcome these difficulties, the decision-maker needs to develop a trust behavior profile for its reporting sources to estimate how *trustful* and useful their reports are. Then, the decision-maker needs to properly fuse the reports in light of these profiles. It is desirable for the fused opinion about Joe to consistently represent an estimate of the ground truth probabilities of Joe's tendencies and the uncertainty about these probabilities. In this work, the fused opinion is represented as an effective number of observations for which each value of a state variable, e.g., Joe's action to help or not, is instantiated. The effective total number of observations represents the accuracy as a facet of the quality of the information, and it should relate to how close "on average" the estimated probabilities are to the ground truth values.

The development of the trust behavior profiles for the sources is updated as the decision-maker incorporates source reports about instantiations of different state variables. When the decision-maker has its own (limited) observations to form an initial opinion about the values of the state variable, it can leverage the consistency of its own opinion with a particular source's report to update the source's trust profile. In essence, the decision-maker is also its own ego-source. This chapter will review our recent research in trust estimation and fusion when the ego-source is available [15, 26, 27].

In many cases, the decision-maker will not be able to make observations about the values of state variables. This chapter will look at extensions of our previous work for this circumstance. Specifically, the conditions when trust estimation and fusion lead to and do not lead to information with a consistent quality of information characterization will be exposed.

This chapter is organized as follows. Section 6.2 reviews related work, and Sect. 6.3 provides the mathematical foundation to represent subjective opinions to represent distributions for the values of state variables. The trust estimation and fusion problem and corresponding models are presented in Sect. 6.4. Section 6.5 reviews recent solutions and demonstrates their effectiveness when an ego-source is available, and Sect. 6.6 extends the solutions for cases when the ego-source is unavailable. This section also demonstrates the effectiveness of the newly extended solutions. Finally, a discussion of results with concluding remarks is provided in Sect. 6.7.

6.2 Related Work

Fusing uncertain information from unreliable sources has drawn significant attention from the literature. It still stands as an important research problem with wide range of applications in many different domains [10]. There are a number of mathematical frameworks for modeling uncertainty and fusing uncertain information. One prominent example of such frameworks is the evidential theory proposed by Dempster and Shafer [29], where belief masses are assigned to possible outcomes of a proposition, i.e., subsets of a frame of discernment. There have been other approaches inspired from the work of Dempster and Shafer. Jøsang proposed subjective logic (SL), which is a probabilistic logic that explicitly takes uncertainty and belief ownership into account. It is used to model and reason with situations that involve uncertainty and incomplete knowledge. A subjective opinion represents assignment of belief masses to possible values of a state variable, and various logical/analytical operators are used to define a calculus over subjective opinions [7–10, 12, 13]. Each subjective opinion can be represented as a Dirichlet distribution over the values of a state variable, and operators defined over these opinions are performed over the underlying Dirichlet distributions. The statistical underpinning of SL makes it flexible and versatile for many domains and applications. For instance, Liu et al. used SL to compute reputation models of mobile ad hoc networks [18]. Oren et al. proposed to use SL to enhance argumentation frameworks with evidential reasoning [21]. Han et al. used SL for forensic reasoning over the surveillance metadata [5]. Sensoy et al. used it for determination of conflicts in and fusion of information from unreliable sources [25]. In this chapter, we also use Dirichlet distributions to represent and combine subjective opinions from unreliable sources.

Fusion of information from unreliable has been studied in the literature with different scenarios and assumptions. As a result of the rise of Internet, e-commerce has been embraced by users. However, users do not only buy and sell on the Internet but also share their opinions, ratings, and experiences, e.g., through review sites. Therefore, initial work on the fusion of uncertain opinions focuses on propositions about the service quality of online vendors or service providers. A proposition is simply a state variable which can take only two values: *true* or *false*. However, these opinions are collected from unreliable sources, which may aim to mislead the decision-makers, e.g., online buyers. In these works, subjective opinions are usually represented as pairs of positive and negative number of interactions (or experiences) with the service providers (or vendor). Jøsang and Ismail proposed beta reputation systems (BRS) [11], where opinions about a proposition x such as "Bob provides good services" is modeled using beta distributions. Let p_x represent the probability that the proposition is *true*. A beta distribution is used to model the likelihood of each p_x value. Initially, before having any experience with Bob, the beta distribution is represented by parameters $\langle 1, 1 \rangle$, which corresponds to the uniform distribution. This means that p_x can be anything between 0 and 1 with equal probability. However, after having r good and s bad experiences with Bob, the beta distribution parameters are updated as $\langle r + 1, s + 1 \rangle$ using Bayesian update. In BRS, opinions about Bob are collected from a number of information sources, and these opinions are fused using Bayesian update, i.e., evidence aggregation. However, some malicious sources may disseminate misleading opinions.

Whitby et al. extended BRS to filter out misleading opinions provided by the malicious sources. This approach filters out those opinions that do not comply with the significant majority by using an *iterated filtering approach* [37]. Hence, this approach assumes that the majority of sources honestly share their opinions, i.e., liars are in the minority. The extended BRS does not assume that the decision-maker can use its observations to estimate the reliability of information sources. Because BRS is a simple trust-based fusion approach, it has been used in many domains, such as wireless sensor networks [6]. Bui et al. have proposed to use it to estimate trust in sensor readings in body area sensor networks [1]. Ganeriwal et al. proposed a reputation framework for high integrity sensor networks based on the BRS [4].

To avoid the need to rely on a majority of sources to be honest, some existing work assumes an ego-agent, i.e., a decision-maker may observe evidence about the ground truth using its own sensors. Hence, an ego-agent can evaluate the information sources by comparing its own observations against those reported by these sources. TRAVOS [31] is one such information fusion framework, which is similar to BRS in terms of representation and fusion of subjective opinions. However, TRAVOS keeps a history of opinions from information sources about propositions, such as the aforementioned proposition about Bob's services. To measure the trustworthiness of a source, the decision-maker compares the source and ego opinions over multiple propositions to determine a beta distribution to describe the trust in each source.

Bayesian modeling has also been used to address fusion of subjective information from malicious sources. Regan et al. proposed BLADE [23] for reputation modeling of sources and fusion of their ratings in e-marketplaces. This model learns parameters of a Bayesian network to fuse subjective and possibly deceptive information from unreliable sources with varying behavior. Most of the existing Bayesian approaches require at least some of the information sources to consistently share honest opinions. These approaches build models for the trustworthiness of information sources and exploit them while fusing their opinions. However, there are approaches, where fused opinions do not directly rely on the opinions from sources. For example, Teacy et al. proposed HABIT, which uses hierarchical Bayesian modeling for fusion of opinions from unreliable sources [32]. It does not directly estimate trustworthiness of information sources. Instead, it uses opinions from sources to measure similarity of the current proposition to past propositions. Then, using the computed similarities as weights, the fused opinion for the current proposition is computed as the weighted average of the decision-maker's opinions about the past propositions. This approach is robust to malicious behaviors, but it requires decision-maker to have accurate opinions for the past propositions similar to the current proposition.

Fact finders address the scenarios where the decision-maker cannot directly observe evidence about the ground truth and the sources only provide absolute claims. They try to identify truth among many conflicting claims without any prior knowledge or observation about the trustworthiness of the information sources. Unlike the previously mentioned approaches, fact-finding approaches assume that the truth is crisp and certain. That is, for a given state variable, only one of k mutually exclusive values can be true. Instantiation of the state variable with each of these values is called a claim. TruthFinder [39] defines trustworthiness of sources as a function of the confidences of the their claims and, conversely, defines the claim confidence as a function of the trustworthiness of the sources espousing them. Then, it iterates by calculating the confidence from the trustworthiness and vice versa. Pasternack and Roth [22] generalizes fact-finding by incorporating source-claim weights in the iterative equations to represent a degree of uncertainty in the observations of claims or in the belief of the sources in the claims.

Recently, Wang et al. formalized fact-finding as a maximum likelihood problem where the expectation-maximization (EM) algorithm [19] is used to estimate the reliabilities of the claims and the users at the same time iteratively [35]. This approach enables the formulation of Cramer-Rao bounds to establish the quality of the estimated reliabilities in terms of the structure of the source-claim network [34]. Furthermore, this approach has been applied for social sensing by estimating the reliability of information from the crowd for sensing situations and events. Specifically, the data from micro-blogging sites such as Twitter[1] has been used to detect social and environmental events earlier than traditional means [36]. The EM approach is further extended in [33] to incorporate the confidence of the users by incorporating weights into the iterative equations similar to [22].

In this work, we aim to exploit behaviors of unreliable sources while fusing their uncertain and possibly misleading opinions. In the literature, different types of information source behaviors are defined and studied [3, 17, 24, 40]. Yu and

[1] http://www.twitter.com

Singh defined four major types of source behaviors over binomial subjective opinions: *honest*, *complementary*, *exaggerated positive*, and *exaggerated negative* [40]. Sources with *honest* behavior share their genuine opinion; on the other hand, the sources adopting non-honest behaviors transform their opinions before sharing. Sources with *complementary* behavior share the opposite of their genuine opinions, i.e., flipping its true opinion. A source with *exaggerated positive* behavior shares an opinion that is more optimistic than its genuine opinion. Similarly, a source with *exaggerated negative* behavior shares an opinion that is more pessimistic than its genuine opinion. The deception models of Yu and Singh received significant amount of attention from the literature. These models are also used in different domains and disciplines. Fung and Boutaba used these deception models for collaborative intrusion detection in networks [3]. In this setting, peers send feedback about the risk levels of a security alert to others.

The honest, complementary, and exaggeration behaviors require information source to know the truth about the state variable in question. However, an information source may still deceive the information requester without knowing the actual truth. In the *Encyclopedia of Deception* [17], fabrication is defined as another type of deception. In the case of fabrication, someone submits statements as truth, without knowing for certain whether or not it actually is true. Therefore, if a source makes up and shares an opinion without actually having any evidence about the proposition in question, then it would be fabricating. This kind of behavior is similar to randomly generating and sharing an opinion when requested.

6.3 Mathematical Preliminaries

A state variable is a random variable that takes on one value from a mutually exclusive set $\mathbb{K}$ at each instantiation. There is a ground truth probability for each possible value to materialize. Given the observations that n^k instantiations of the variable are of value k for all $k \in \mathbb{K}$ are the result of sampling a multinomial distribution, the posterior knowledge about the distribution of the generating probability is the Dirichlet distribution:

$$f_\beta(\mathbf{p}|\mathbf{n}) = \frac{1}{B(\mathbf{n}+1)} \prod_{k\in\mathbb{K}} \left(p^k\right)^{n^k}, \tag{6.1}$$

where

$$B(\mathbf{n}+1) = \frac{\prod_{k\in\mathbb{K}} \Gamma(n^k+1)}{\Gamma\left(\sum_{k\in\mathbb{K}}(n^k+1)\right)} \tag{6.2}$$

is the beta function and $\Gamma(\cdot)$ is the gamma function [16]. Throughout this chapter, the boldfaced variables are $|\mathbb{K}|$ dimensional vectors where their elements are nonbold with a superscript representing the corresponding value in $\mathbb{K}$. Note that in (6.1), the probabilities are constrained to sum to one, i.e., $\sum_{k\in\mathbb{K}} p^k = 1$.

Subjective logic [10] connects the evidence $\mathbf{n}$ to belief mass assignments as used in belief theories for reasoning under uncertainty such as Dempster-Shafer theory [29] and more recently the transferable belief model [30]. Specifically, the connection between the evidence $\mathbf{n}$ and the beliefs $(\mathbf{b}, u)$ is given by the following invertible mapping:

$$b^k = \frac{n^k}{W + \sum_{k\in\mathbb{K}} n^k} \forall k \in \mathbb{K} \quad \text{and} \quad u = \frac{W}{W + \sum_{k\in\mathbb{K}} n^k}, \tag{6.3}$$

where the b^ks are the beliefs for each value of the state variable and u is the remaining uncertainty. The beliefs and uncertainty are constrained to be nonnegative and sum to one. In (6.3), W is the prior weight, In this chapter, we set $W = |\mathbb{K}|$ and consider the uniformative uniform prior. The connection between beliefs and the Dirichlet distribution helps to define many of the operators in subjective logic, which distinguishes it from the prior belief theories by connecting it to second-order Bayesian reasoning.

It is well know that the expected value for the probabilities of the Dirichlet distribution is given by

$$m^k = \frac{n_k + 1}{\sum_{k'\in\mathbb{K}}(n^{k'} + 1)}, \tag{6.4}$$

and the variance is

$$\sigma^{2k} = \frac{m^k(1 - m^k)}{1 + \sum_{k'\in\mathbb{K}}(n^{k'} + 1)} \tag{6.5}$$

for $k \in \mathbb{K}$. In the context that an opinion about a state variable is given by $\mathbf{n}$, the mean given by (6.4) represents the information about, i.e., estimation of, the ground truth probabilities. Likewise, the variance given by (6.5) represents the derived quality of information. The smaller the variance, the higher the quality of the information. Note that the quality of information is proportional to the sum of evidences, i.e., $\sum_{k\in\mathbb{K}} n^k$. The derived quality of information is meaningful if it corresponds to the actual variance through (6.5). This will be discussed by examples throughout this chapter.

Subjective logic provided the inspiration for the fusion and trust characterization operators described in this chapter. The operators described here approximate Bayesian reasoning using the following framework. The input opinions about the state variables and source behaviors translate to Dirichlet distributions to describe the uncertainty about the corresponding appearance probabilities of the various values of these variables. Bayesian reasoning determines the exact output distribution for the appearance probabilities for fusion or discounting, and then this exact distribution is approximated by a Dirichlet distribution such that the mean values match exactly and the variances match in the least squares sense. In other words, moment matching determines the Dirichlet approximation, and the corresponding Dirichlet parameters lead to the fused or discounted opinion.

6.4 The Source Estimation and Fusion Problem

In general, a decision-maker collects, over the course of his/her duties, reports from different sources about many different state variables. The decision-maker employs A unique sources and evaluates I variables. The decision-maker may or may not be able to form an initial opinion about each variable. We denote the opinion about the i-th state variable from the a-th source using a subscript as $\mathbf{n}_{i,a}$. When the a-th source does not have any observations about the i-th variable, it should report the vacuous opinion $n^k_{i,a} = 0$ for all $k \in \mathbb{K}$. The decision-maker may or may not be able to form an initial opinion about a state variable and acts as an ego-source. We index the ego-source as $a = 0$ and other sources as positive integers $a > 0$. For ease of illustration in the chapter, all of the I state variables are binary, i.e., their instantiations are propositions where $\mathbb{K} = \{+, -\}$ and $+$ and $-$ represent a positive and negative variable value, respectively. An example of such a proposition is that a particular vendor provides a satisfactory $(+)$ or an unsatisfactory $(-)$ transaction.

The sources do not necessarily correctly report their opinions based upon their individual observations. Many times, some sources intentionally lie and report opinions in direct conflict with other sources. The ultimate problem for the decision-maker is to form a fused opinion that portrays information about ground truth probabilities of the values of state variables consistent with the opinion's apparent quality of information. This fused opinion should represent higher quality of information than can be obtained from any smaller subset of sources.

To enable an effective solution to the fusion problem, we incorporate the beta model from [20]. Specifically, the behavior of a source is a state variable itself where the variable values are particular behaviors describing how the source transforms its truthful opinion into its reported opinion. While a large number of source behaviors may exist, we restrict the discussion in this chapter to the three well-studied behaviors from the literature [3, 17, 24, 40]: (1) good, (2) flipping, and (3) random. In the good behavior case, the a-th source accurately reports the number of positive and negative instantiations of state variables it observed. When the source exhibits flipping behavior, it exchanges the number of positive and negative instantiations. Finally in the random case, a source randomly selects the number of positive and negative instantiations to report independent of the actual numbers it observed. This chapter will examine the robustness of such a beta model by considering that the ground truth source behaviors are one of these three, but the fusion algorithms either account only for two behaviors (good and random) or all three. Clearly, performance drops when there is a model mismatch, and in real applications, one may want to incorporate a richer set of behavior models. In recent work, we developed methods to learn new behavior models using an ego-source [27]. These richer behavior models are beyond the scope of this chapter. Nevertheless, the results in this chapter do provide insights about the impact of model mismatches.

The a-th source's behavior profile is the ground truth probabilities to exhibit each one of the three behaviors adopted while sharing its opinions. The decision-maker builds up an opinion about the behavior profile by determining the effective number of instances t_a^k that the a-th source exhibited behavior $k \in \{g, f, r\}$. After each time the decision-maker collects opinions from the different sources, it cannot directly determine which behavior each source actually followed. The next two sections describe methods to build up opinions about the source behaviors and then use these opinions to fuse the source reports. Due to the lack of direct observations, the behavior opinions t_a^k are not necessarily integers, which also means the fused opinions about the variables also need not take integer values.

To demonstrate the effectiveness of the methods presented in the next two sections, 100 sources reporting 1000 variables are simulated using the three behavior models. A given percentage of the source agents will be predominately good, flippers, and random, respectively. Predominately good sources report their true opinions about a particular variable with a probability of 0.7, and their flipping and random probabilities are 0.15. Similarly, the predominately flipping and random source exhibits their dominating behaviors with probability 0.7 and the other two behaviors with probabilities of 0.15. Predominately good sources can lie, albeit with a much smaller probability. In contract a predominately flipping source can provide a truthful opinion. For the i-th state variable, the number of direct observations $N_{i,a}$ that the a-th source achieves for the variable's values is a random number drawn uniformly between 0 and 100. The underlying ground truth probability for the positive value p_i^+ of each of the state variables is sampled over the uniform distribution between 0 and 1. The a-th source's true opinion about the i-th variable is the result of $N_{i,a}$ draws from a Bernoulli process with probability p_i^+. Each source then determines its behavior for the variable as a random multinomial draw using its behavior probabilities. If this draw selects the good behavior, the source reports it true opinion. If the draw selects the flipping behavior, the source swaps its $n_{i,a}^+$ and $n_{i,a}^-$ values. Otherwise the random behavior means that the sources chooses the integer $n_{i,a}^+$ uniformly between 0 and $N_{i,a}$ and sets $n_{i,a}^- = N_{i,a} - n_{i,a}^+$.

If the decision-maker does not account for the various behaviors of the sources and assumes all the reported opinions for the i-th variable are correct, the fusion process is rather straightforward. The fusion operations make two weaker assumptions: (1) each reported opinion is statistically independent of the others (i.e., the observed evidence of the sources do not overlap), and (2) the prior distribution in light of no observed evidence is uniform (which is an uninformative prior). With these assumptions, it can be shown that the distribution of the fused opinion is a beta distribution given by

$$f_\beta(p|\mathbf{n}_{i,f}) \propto \prod_{a=1} f_\beta(p|\mathbf{n}_{i,a}) \tag{6.6}$$

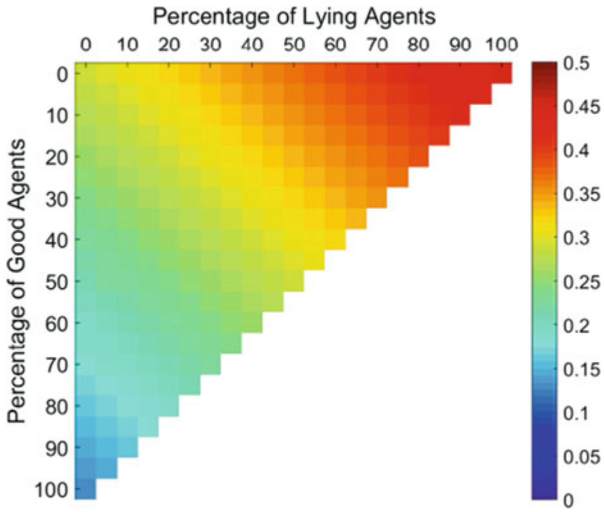

Fig. 6.1 RMSE of consensus fusion for various mixtures of predominately good, flipping, and random sources represented as a heat map

where $n^k_{i,f} = \sum_{a=1}^{A} n^k_{i,a}$ for $k \in \{+, -\}$. In other words, the fused opinion in evidence space is the output of the operator $\texttt{Consensus}(\mathbf{n}_{i,1}, \ldots, \mathbf{n}_{i,A})$ that simply sums the evidence supplied by each source. `Consensus` fusion is one of the more commonly used operators in subjective logic [10].

When the consensus operator is applied over all simulated reports from 100 agents covering 1000 variables, the resulting root mean square error (RMSE) between the expected fused probability (see (6.4)) and the ground truth opinion for various mixtures of sources types is given in Fig. 6.1 as a heat map. When most of the sources are predominately good, the RMSE is fairly low at 0.13. As the those sources are replaced by predominately random sources, the RMSE grows to about 0.3, which is consistent to a complete random guess from a uniform distribution. When most of the sources are predominately flipping, the RMSE grows to above 0.4 as the flipped reports are moving the estimated probabilities far from the ground truth. The predicted RMSE can be calculated as the root mean of the expected variance of the fused opinions given by (6.5). For all cases, the predicted RMSE is similar with a mean value of 0.0070, which is much smaller than the actual RMSE (even with 100% good sources). This is because `consensus` fusion assumes all source reports are good, which is not even true 30% of the time for good sources. Clearly, the behavior of the sources must be accounted for in the fusion process.

6.5 Fusion Using Behavior Estimates via Ego-Sources

When the decision-maker can directly observe instantiations of different state variables, it can act as its own ego-source. Assuming the decision-maker is competent and acting in its own best interest, the ego-source's opinion will always be good. Then, the decision-maker compares its opinions against those of the a-th source over a set of I variables to determine the source behavior profile. The procedure to determine the opinions about source behaviors was first derived in [15] for good and random behaviors (the two-mode model), and it can trivially be generalized for a finite set of known behaviors. Following the derivation in [15], the posterior distribution for the probabilities that the a-th source follows particular behaviors given the set of opinions from the ego-source about propositions and a-th source is

$$f(\mathbf{p}|\mathbf{t}_a) \propto \prod_i \left(\sum_{k \in \mathbb{K}} \text{Prob}(\mathbf{n_{i,0}}|\mathbf{n_{i,a}}, k) p^k \right), \tag{6.7}$$

where the likelihood of the a-th source exhibiting the k-th behavior when reporting its opinion about the i-th variable is

$$\begin{aligned} \text{Prob}(\mathbf{n_{i,0}}|\mathbf{n_{i,a}}, k) &= \int p^{n_{i,0}^{+}} (1-p)^{n_{i,0}^{-}} f_\beta(p|h^k(\mathbf{n_{i,a}}))dp, \\ &= \frac{B(h^k(\mathbf{n}_{i,a}) + \mathbf{n}_{i,0} + 1)}{B(h^k(\mathbf{n}_{i,a}) + 1)}, \end{aligned} \tag{6.8}$$

where $h^g(\mathbf{n}) = \mathbf{n}$, $h^f(\mathbf{n}) = [n^-, n^+]$, and $h^r(\mathbf{n}) = [0, 0]$ represent the accurate information that can be obtained from the source when it is known to employ good, flipping, or random behavior, respectively, for the given report. For the random behavior, the opinion is completely independent of the source's actual observation, and therefore the sources report is vacuous.

The source behavior characterization method approximates (6.7) by finding the Dirichlet distribution that matches the means of (6.7) and matches the variances as closely as possible (in the least squares sense). Closed form expressions for the means and variances of (6.7) are available because the distribution is a mixture of Dirichlets. However, the number of modes grows exponentially with respect to the number of variables I. In [15], a method is presented that updates a source behavior opinion by sequentially performing moment matching over one state variable at a time. It is shown in [15] that this sequential updating method is almost as accurate as the much more computationally complex method that incorporates all propositions at once. We refer to the operator $\mathbf{t}_a = \texttt{SourceBehavior}(\mathbf{n}_{1,0}, \mathbf{n}_{1,a}, \ldots, \mathbf{n}_{I,0}, \mathbf{n}_{I,a})$ as the sequential method that approximates the Dirichlet distribution for the source behavior probabilities using the parameters $\mathbf{t}_a$ as effective evidences of the source behaviors. In this chapter, while the simulated sources are randomly picking one of three behaviors for each propositional report, the source characterization method is either employing the two-mode model, i.e., $\mathbb{K} = \{g, r\}$ or the three-mode model, i.e., $\mathbb{K} = \{g, f, r\}$. This allows understanding of the performance loss when the assumed model does not fully characterize the data.

In essence, the `SourceBehavior` operator calibrates the a-th source relative to the ego-source. Kaplan et al. [15] explain how the behavior opinions are updated based upon the consistency between the opinions of the ego-source and those of the a-th source. When both opinions represent similar probabilities and are supported by large evidence (i.e., small uncertainty), then the likelihood for the good behavior in (6.8) becomes very large relative to the other likelihoods. This actually means that the evidence for the good behavior is incremented by a number near 1 and the evidence for the other behaviors is slightly decremented. Similarly, if the probabilities are consistent with flipping, the evidence for flipping behavior is incremented by a number near 1. Otherwise when the probabilities are inconsistent to either good or flipping behaviors, the evidence for random behavior is incremented by a number near to 1. The increment of the behavior evidence update decreases as the uncertainty associated to either source opinion increases. When the ego-source's opinion becomes vacuous, i.e., $n^+_{i,0} = n^-_{i,0} = 0$, the likelihoods for each of the behaviors become equal, and the update does not change any of the source behavior opinions. In other words, when the uncertainty of the reported propositions are low, the behavior update is comparable to directly observing which behavior the a-th source used in reporting the given proposition. As the uncertainty of either reported opinion grows, the increments to the source behavior evidence go to zero. The strength of the update depends on how much direct evidence the ego-source is able to observe.

Given the characterization of the behavior of the sources, the subjective logic method discounts each source's behavior followed by `consensus` fusion [14]. The discount operation in subjective logic, which originates from Dempster-Shafer theory [29], only considers the belief that the source provides a good report. Specifically, the function $\texttt{Discount}(\mathbf{t}, \mathbf{n}) = \frac{t^g+1}{\sum_{k\in\mathbb{K}}(t^k+1)}\mathbf{n}$ discounts the opinion based upon the expected probability of a good report. The reports of all sources for the i-th variable are discounted by `Discount` using their respective behavior profiles $\mathbf{t}_a$, and the outputs are passed through the `consensus` operator. In effect, the discount operator acts as a "soft" censor for sources.

Figure 6.2a shows the RMSE result s of subjective logic discounting and fusion. The error is clearly reduced as compared to consensus fusion. Specifically, the RMSE performance relies mostly on the percentage of predominately good sources. When the percentage of predominately good sources is 10%, the RMSE is about 0.2 (much lower than consensus alone), and this value decreases to 0.13 when all the sources are predominately good (comparable to consensus alone). Like consensus, the uncertainty associated to the fused opinion still greatly underpredicts the actual RMSE, where the predicted RMSE averages around 0.053 for the various mixtures of sources. The predicted RMSE is higher than that of simple consensus fusion

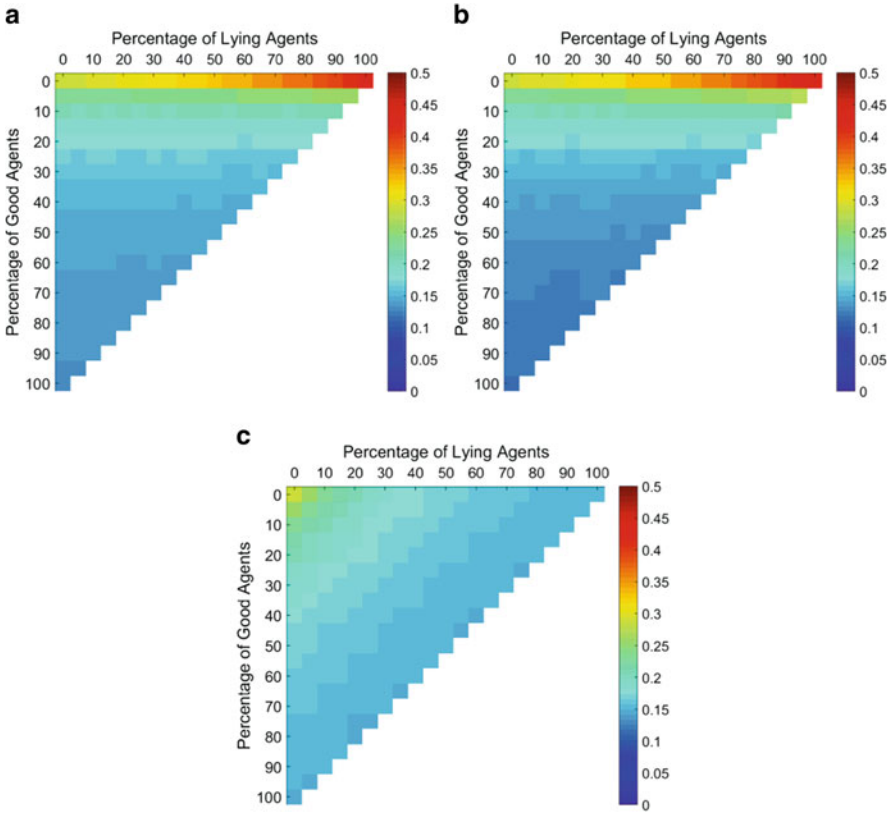

Fig. 6.2 RMSE of discounted fusion for various mixtures of predominately good, flipping, and random sources represented as a heat map: (**a**) subjective logic discounting followed by consensus fusion, (**b**) two-mode behavior discounting followed by consensus fusion, and (**c**) three-mode behavior discounting followed by consensus fusion

due to the discounting. It is greatest when the percentage of predominately flipping sources is 100% (0.13) and is smallest when the percentage of predominately good sources is 100% (0.026).

While the discount operator is intuitively appealing, it is ad hoc. By the two-mode beta model, the distribution for the probability of the i-th variable due to the a-th source's report is

$$f(p|\mathbf{t}_a, \mathbf{n}_{i,a}) \propto \frac{t_a^g + 1}{t_a^g + t_a^r + 2} f_\beta(p|[n_{i,a}^+, n_{i,a}^-]) + \frac{t_a^r + 1}{t_a^g + t_a^r + 2} f_\beta(p|[0,0]). \tag{6.9}$$

The $\texttt{Discount}_2(\mathbf{t}_a, \mathbf{n}_{i,a})$ determines the discounted report as the evidence parameters of the beta distribution whose means and variances match the distribution in (6.9). This form of discounting was used in the TRAVOS trust and reputation model [31], and Eqs. (11)–(15) in [31] implement the $\texttt{Discount}_2$ operator.

Figure 6.2b shows the RMSE results when employing the $\texttt{Discount}_2$ operator followed by consensus. The results look very similar to SL discounting results. If one squints and look at the actual numbers, one will actually see a slight improvement using the two-mode discounting. Again, the predicted RMSEs are much lower than the actual errors. The predicted values are actually slightly larger than SL discounting, but not by much.

It is now natural to wonder about how the three-mode beta model can fair. In this case, the distribution for the probability of the i-th variable due to the a-th source's report is

$$f(p|\mathbf{t}_a,\mathbf{n}_{i,a}) \propto \frac{(t_a^g+1)f_\beta(p|[n_{i,a}^+,n_{i,a}^-])+(t_a^f+1)f_\beta(p|[n_{i,a}^-,n_{i,a}^+])+(t_a^r+1)}{t_a^g+t_a^f+t_a^r+3}. \tag{6.10}$$

Again, the method of moments can be employed to extract a discounted opinion by approximating (6.10) by a beta distribution. We refer the process as the $\texttt{Discount}_3(\mathbf{t}_a,\mathbf{n}_{i,a})$ operator. The actual operator is a special case of the joint discounting and consensus fusion operator described in [27] which will be discussed soon.

Figure 6.2c provides the RMSE results when employing $\texttt{Discount}_3$ before consensus. The results improve significantly over the two previous discounting operators as the number of predominately flipping sources increases. This is because the discounting operator actually to some extent "unflips" the reports from the flipping sources. Now, the performance of the fusion is primarily a function of the percentage of predominately random sources. With no random sources, the error is about 0.15 and grows to 0.29 when all sources are predominately random. Like the previous discounting operators, the predicted RMSE is much lower than the actual error. The predicted error is as low as 0.028 for no random source and is as high as 0.075 when all sources are predominately random.

The large gap between the predicted and actual fusion results for all the discounting methods indicated that more can be done. The discounted reports as given by the beta mixtures in (6.9) and (6.10) are poorly fitted by a single beta distribution. It is actually better to perform the fusion with the beta mixtures before finding an approximate beta distribution fit. Under the fairly general assumption that the prior on the distribution of values of propositions is uniform, the distribution after fusing the reports from all sources is

$$f(p|\mathbf{T},\mathbf{N}_i) \propto \prod_{a=1}^{A} f(p|\mathbf{t}_a,\mathbf{n}_{i,a}), \tag{6.11}$$

where $f(p|\mathbf{t}_a,\mathbf{n}_{i,a})$ is given by (6.9) or (6.10) for the two-mode or three-mode behavior model, respectively. The operator $\mathbf{n}_{i,f} = \texttt{JointConDis}(\mathbf{t}_1,\mathbf{n}_{i,1},\ldots,\mathbf{t}_A,$ $\mathbf{n}_{i,A})$ determines the fused opinion by selecting the opinion associated to the beta distribution that is determined through moment matching to (6.11). The distribution in (6.11) is a mixture of beta distributions, which leads to analytical expressions for

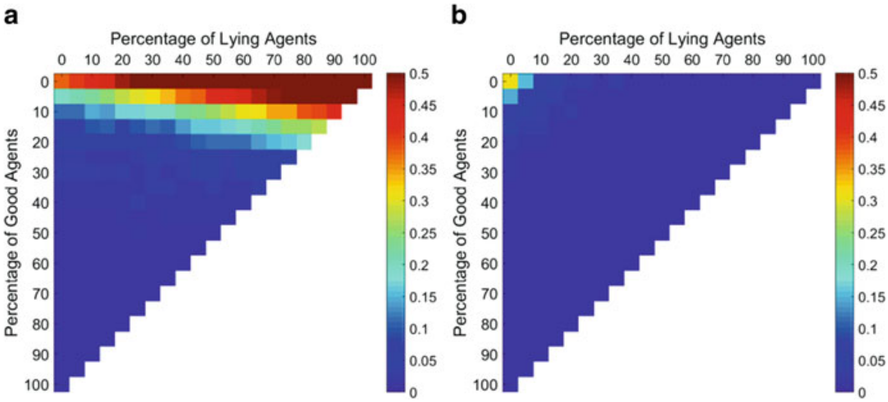

Fig. 6.3 RMSE of joint discounting and fusion for various mixtures of predominately good, flipping, and random sources represented as a heat map: (**a**) two-mode and (**b**) three-mode

the moments. However, the number of mixture components grows exponentially. A practical implementation of the `JointConDis` operator is presented in [27] that clusters similar components into a single component as more sources are integrated.

Figure 6.3a shows the RMSE results for two-mode `JointConDis`. The improvement over the previous discounting methods is obvious when the percentage of predominately good sources is 10% or greater. Otherwise, the reports from the predominately flipping agents can cohere in the fusion process and drown out the good reports. Therefore the error can become worse than the previous discounting methods when the percentage of predominately flipping sources is large. Now the predicted RMSE matches the actual RMSE as long as the percentage of predominately good agents dominates the flipping sources. Table 6.1 provides the actual and predicted RMSE numbers for various percentages of source types. The discrepancy between the actual and predicted values as more flipping sources are included is due to the fact that the two-mode model does not account for the flipping behavior.

Figure 6.3b shows the RMSE results for three-mode `JointConDis`. The error is now very small except for cases when the percentage of predominately random agents is 95% or more. Because the flipping behavior is modeled in the fusion approach, the reports of flippers can be “unflipped” so that predominately flipping sources are providing comparable information as predominately good agents, and the fusion method is able to exploit that information. Because the fusion method is modeling all the behaviors inherent in the synthesized sources, the predicted and actual errors are comparable except when all the sources are predominately random as provided in Table 6.2. It seems that the joint consensus fusion process that models all the source behaviors is able to achieve the lowest possible error and the fused opinion is able to represent the quality of information after fusion. The error is the lowest when none of the sources are predominately random. In such cases, the actual and predicted RMSE is 0.0075, which is slightly higher than the predicted RMSE

Table 6.1 RMSE and (predicted RMSE) of joint fusion and discounting using the two-mode model with an ego-source for various mixtures of predominately good, flipping, and random sources

% of good sources	% of flipping sources					
	0	20	40	60	80	100
0	0.3791 (0.1342)	0.4789 (0.1212)	0.5458 (0.0899)	0.5680 (0.0663)	0.5787 (0.0344)	0.5802 (0.0269)
20	0.0424 (0.0274)	0.0590 (0.0261)	0.0860 (0.0391)	0.1126 (0.0412)	0.1757 (0.0535)	— —
40	0.0139 (0.0130)	0.0189 (0.0135)	0.0188 (0.0140)	0.0169 (0.0143)	— —	— —
60	0.0112 (0.0105)	0.0118 (0.0105)	0.0119 (0.0106)	— —	— —	— —
80	0.0100 (0.0091)	0.0102 (0.0091)	— —	— —	— —	— —
100	0.0091 (0.0082)	— —	— —	— —	— —	— —

Table 6.2 RMSE and (predicted RMSE) of joint fusion and discounting using the three-mode model with an ego-source for various mixtures of predominately good, flipping, and random sources

% of good sources	% of flipping sources					
	0	20	40	60	80	100
0	0.3124 (0.2004)	0.0284 (0.0246)	0.0113 (0.0112)	0.0095 (0.0094)	0.0083 (0.0083)	0.0076 (0.0075)
20	0.0281 (0.0252)	0.0117 (0.0113)	0.0092 (0.0094)	0.0085 (0.0083)	0.0075 (0.0075)	— —
40	0.0110 (0.0119)	0.0093 (0.0094)	0.0081 (0.0083)	0.0075 (0.0075)	— —	— —
60	0.0096 (0.0094)	0.0085 (0.0083)	0.0077 (0.0075)	— —	— —	— —
80	0.0081 (0.0083)	0.0075 (0.0075)	— —	— —	— —	— —
100	0.0075 (0.0075)	— —	— —	— —	— —	— —

of standard consensus fusion as discussed in the previous section. This is because consensus fusion alone assumes all reports are honest, whereas the predominately honest and flipping sources still provide random reports 15% of the time, which is accounted for in joint consensus and discount fusion in (6.11).

6.6 Fusion Using Behavior Estimates Without Ego-Sources

The three-mode `JointConDis` is probably at the estimation limit when dealing with sources that probabilistically decide to manipulate their reported opinions. The problem is that it requires an ego-source to "calibrate" the source behavior profiles. This section investigates what is possible when an ego-source is unavailable. This occurs when the decision-maker does not have direct access to observe the values over different instantiations of the various variables. This section is inspired by the fact-finding work described in [36, 39].

It is interesting to look at the performance of the two-mode `JointConDis` when the source behavior opinion is vacuous, i.e., $t_a^g = t_a^r = 0$. Figure 6.4a shows the RMSE over the various mixtures of source types. Despite the lack of knowledge about the source behavior, the fusion still works well when the percentage of

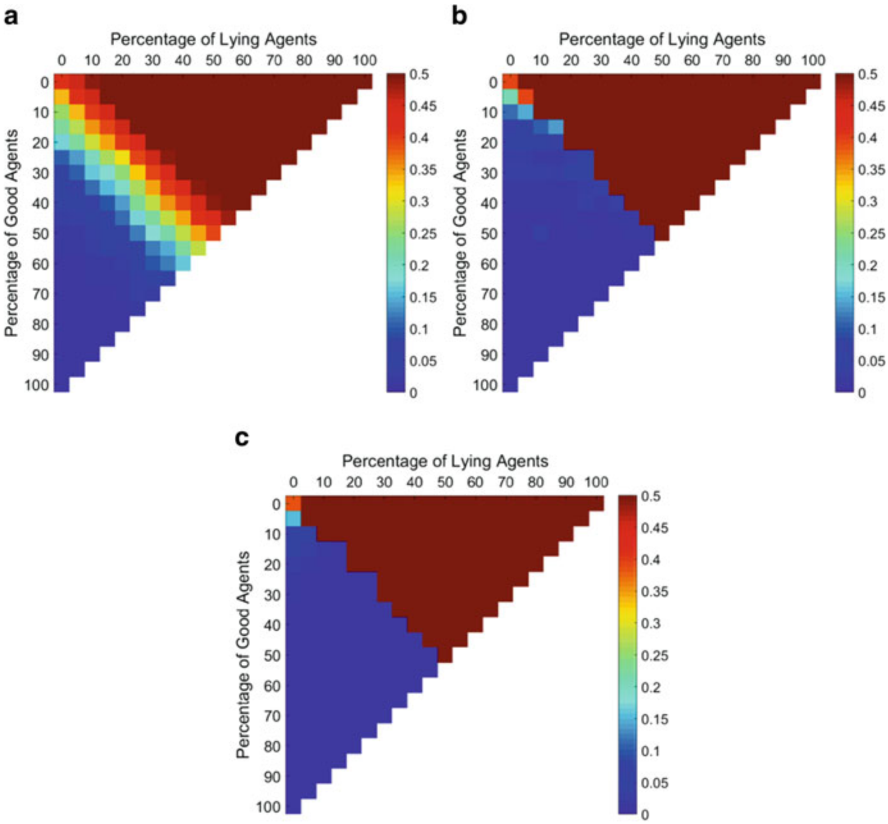

Fig. 6.4 RMSE without an ego-source for various mixtures of predominately good, flipping, and random sources represented as a heat map: (**a**) two-mode joint discounting and fusion using vacuous source behavior profiles, (**b**) two-mode fact-finding, and (**c**) three-mode fact-finding

Table 6.3 RMSE and (predicted RMSE) of joint fusion and discounting using the two-mode model without any ego-source for various mixtures of good, flipping, and random sources

% of good sources	% of flipping sources					
	0	20	40	60	80	100
0	0.4049 (0.0764)	0.5453 (0.0406)	0.5733 (0.0168)	0.5759 (0.0111)	0.5766 (0.0098)	0.5772 (0.0090)
20	0.1702 (0.0479)	0.4009 (0.0757)	0.5462 (0.0480)	0.5736 (0.0172)	0.5763 (0.0100)	— —
40	0.0442 (0.0137)	0.1377 (0.0461)	0.3927 (0.0693)	0.5474 (0.0424)	— —	— —
60	0.0128 (0.0113)	0.0265 (0.0119)	0.1628 (0.0482)	— —	— —	— —
80	0.0104 (0.0098)	0.0132 (0.0100)	— —	— —	— —	— —
100	0.0090 (0.0090)	— —	— —	— —	— —	— —

predominately good sources is much larger than the percentage of predominately flipping sources. In fact, the RMSE is mostly a function of the difference of these two percentages. Table 6.3 provides the actual and predicted RMSE obtained by the fused opinions. When all the sources are predominately good, the match is very close. In this case, the good reported opinions are able to cohere in the `JointConDis` operation against the noncoherent random opinions. The match between the actual and predicted errors slowly deteriorates as the difference between the percentages of predominately good and flipping sources decreases. Once there are more flipping sources, the `JointConDis` is cohering to the flipped opinions, and the actual RMSE becomes large because the estimate is a flipped version of the ground truth. Overall, the performance of two-mode `JointConDis` with the vacuous behavior is not as good as using the ego-source-generated behavior profile, but it does significantly outperform the earlier discounting methods when the predominately good sources outnumber the flipping ones. This indicates that fusion without an ego-source to calibrate the sources is possible, but more can be done as we will now see.

The three-mode `JointConDis` operator using a vacuous source belief profile is ineffective because the distribution given by (6.11) is bimodal due to the modeling of the flipping behavior and the two modes are equiprobable in the absence of prior knowledge of the relative number of sources that are exhibiting good and flipping behaviors for the given variable. Fitting a single beta distribution to this bimodal distribution leads to a poor characterization of the fused opinion, and it is not clear which mode is representative of the ground truth and which mode is representative of the flipped ground truth.

The performance of the two-mode `JointConDis` operator using a vacuous source belief profile appears to provide a surprisingly good representation of the ground truth when the majority of sources are predominately good. The fact-finding methods used for non-probabilistic propositions that have certain values either *true* or *false* [36, 39] provide inspiration to do more. The fact-finding methods alternate between estimating the trustworthiness of the sources given an estimate of the truth of their claims and estimating the truth of the claims given an estimate of the trustworthiness of the sources. In other words, the fused opinion that is the output of the two-mode `JointConDis` operator using the vacuous source behavior profile can serve as an initial surrogate for an ego-source opinion. Then, the `SourceBehavior` operator using a two-mode model provides an updated source behavior profile opinion for each source. Next, updated fused opinions are obtained using the two-mode `JointConDis` operator with the updated source behavior profile opinions, and then the `SourceBehavior` operator updates the source behavior profile opinions using the updated fused opinions. The process repeats until the source behavior profile opinions converge. The details of the two-mode fact-finding method is shown in Operator 1 as $\texttt{FactFind}_2$.

Operator 1 $[\mathbf{n}_{1,f},\ldots,\mathbf{n}_{I,f},\mathbf{t}_1,\ldots,\mathbf{t}_A] = \texttt{FactFind}_2(\mathbf{n}_{1,1},\ldots,\mathbf{n}_{i,a},\ldots,\mathbf{n}_{I,A})$

$t_a^g = t_a^r = 0$ for $a = 1,\ldots,A$
$t_a^{g'} = t_a^{r'} = 1$ for $a = 1,\ldots,A$
while $\sum_{a=1}^{A} \|\mathbf{t}_a^{'} - \mathbf{t}_a\|^2 > \epsilon$ **do**
  $\mathbf{t}_a^{'} = \mathbf{t}_a$ for $a = 1,\ldots,A$
  /* Use 2-mode source behavior model */
  $\mathbf{n}_{i,f} = \texttt{JointConDis}(\mathbf{n}_{i,1},\mathbf{t}_1,\ldots,\mathbf{n}_{i,A},\mathbf{t}_A)$ for $i = 1,\ldots,I$
  $\mathbf{t}_a = \texttt{SourceBehavior}(\mathbf{n}_{1,f},\mathbf{n}_{1,a},\ldots,\mathbf{n}_{I,f},\mathbf{n}_{I,a})$ for $a = 1,\ldots,A$
end while

Figure 6.4b shows the RMSE of the two-mode fact-finding method. As long as the predominately good sources outnumber the predominately flipping agents, the RMSE is very low. Otherwise, the error is large because the flipped version of the ground truth prevails. Table 6.4 compares the actual and predicted RMSE values for various mixtures of source types. As long as the number of predominately good sources significantly outnumbers the other source types, the predicted and actual errors are comparable. The discrepancy between the actual and predicted errors when the predominately good sources are slightly in the majority indicates that more still can be done.

The two-mode fact-finding method does not correct for flipping behavior. A three-mode fact-finding method can do better, but without an initial estimate of the source behaviors, three-mode `JointConDis` suffers from the two-mode problem discussed earlier. Operator 2 describes the three-mode fact-finding operator $\texttt{FactFind}_3$ that alternates between joint discounting and fusion and source behavior characterization using the fused opinions as ego-source surrogates. It is initialized by $\texttt{FactFind}_2$ to determine initial evidence for the good behavior associated to each source. To this end, the second step in Operator 2 is actually transferring the belief in the random behavior into uncertainty. In setting up a three-

Table 6.4 RMSE and (predicted RMSE) of two-mode fact-finding for various mixtures of good, flipping, and random sources

% of good sources	% of flipping sources					
	0	20	40	60	80	100
0	0.3947 (0.1071)	0.5721 (0.0256)	0.5757 (0.0124)	0.5764 (0.0104)	0.5763 (0.0090)	0.5764 (0.0081)
20	0.0434 (0.0271)	0.5689 (0.0312)	0.5746 (0.0159)	0.5763 (0.0104)	0.5762 (0.0091)	— —
40	0.0139 (0.0130)	0.0200 (0.0127)	0.5759 (0.0145)	0.5764 (0.0105)	— —	— —
60	0.0112 (0.0104)	0.0120 (0.0104)	0.0122 (0.0105)	— —	— —	— —
80	0.0100 (0.0090)	0.0104 (0.0091)	— —	— —	— —	— —
100	0.0092 (0.0081)	— —	— —	— —	— —	— —

mode belief, the beliefs in flipping and random behavior are set to zero, and the good behavior belief and uncertainty are transferred from the two-mode behavior opinion. This can be verified by the evidence to belief mapping given by (6.3). The rationale for moving the random behavior belief to uncertainty is because the two-mode fact-finding cannot distinguish between random and flipping behavior and only the belief in the good behavior is valid. Then, the fact-finding methods alternate between performing three-mode `JointConDis` to estimate fused opinions as surrogates for the ego-source opinions and three-mode `SourceBehavior` to update the source behavior profile opinions until convergence.

Operator 2 $[\mathbf{n}_{1,f}, \ldots, \mathbf{n}_{I,f}, \mathbf{t}_1, \ldots, \mathbf{t}_A] = \texttt{FactFind}_3(\mathbf{n}_{1,1}, \ldots, \mathbf{n}_{i,a}, \ldots, \mathbf{n}_{I,A})$

$[\tilde{}, \ldots, \tilde{}, \mathbf{t}_1, \ldots, \mathbf{t}_A] = \texttt{FactFind}_2(\mathbf{n}_{1,1}, \ldots, \mathbf{n}_{i,a}, \ldots, \mathbf{n}_{I,A})$

$t_a^g = \frac{3t_a^g}{2+t_a^r}$ for $a = 1, \ldots, A$

$t_a^f = t_a^r = 0$ for $a = 1, \ldots, A$

$t_a^{g'} = t_a^{f'} = t_a^{r'} = 1$ for $a = 1, \ldots, A$

while $\sum_{a=1}^{A} \|\mathbf{t}_a^{'} - \mathbf{t}_a\|^2 > \epsilon$ **do**

  $\mathbf{t}_a^{'} = \mathbf{t}_a$ for $a = 1, \ldots, A$

  /* Use 3-mode source behavior model */

  $\mathbf{n}_{i,f} = \texttt{JointConDis}(\mathbf{n}_{i,1}, \mathbf{t}_1, \ldots, \mathbf{n}_{i,A}, \mathbf{t}_A)$ for $i = 1, \ldots, I$

  $\mathbf{t}_a = \texttt{SourceBehavior}(\mathbf{n}_{1,f}, \mathbf{n}_{1,a}, \ldots, \mathbf{n}_{I,f}, \mathbf{n}_{I,a})$ for $a = 1, \ldots, A$

end while

Figure 6.4c shows the RMSE of the three-mode fact-finding method. The boundary where the numbers of predominately good and flipping agents are comparable is sharper than that of the results from the two-mode fact-finding method. The actual

Table 6.5 RMSE and (predicted RMSE) of three-mode fact-finding for various mixtures of good, flipping, and random sources

% of good sources	% of flipping sources 0	20	40	60	80	100
0	0.3820 (0.1239)	0.5743 (0.0238)	0.5761 (0.0112)	0.5764 (0.0094)	0.5764 (0.0083)	0.5765 (0.0075)
20	0.0243 (0.0262)	0.5761 (0.0115)	0.5763 (0.0094)	0.5765 (0.0083)	0.5765 (0.0075)	— —
40	0.0108 (0.0114)	0.0093 (0.0094)	0.5764 (0.0083)	0.5765 (0.0075)	— —	— —
60	0.0096 (0.0094)	0.0085 (0.0083)	0.0077 (0.0075)	— —	— —	— —
80	0.0081 (0.0083)	0.0076 (0.0075)	— —	— —	— —	— —
100	0.0075 (0.0075)	— —	— —	— —	— —	— —

and predicted RMSE numbers are provided in Table 6.5. When the predominately good sources outnumber the flippers, the three-mode fact-finding lowers the error as compared to two-mode fact-finding because it does explicitly correct for flipping behaviors. Furthermore, the agreement between the actual and predicted errors is maintained as long as the predominately good sources outnumber the flippers. It seems that the three-mode fact-finding is pushing the limits of what is possible for jointly performing source and fused opinion in the absence of an ego-source. Without the ego-agent, one must make the implicit assumption that lying sources are in the minority. This is true for state variables in general and is also true for traditional fact-finding methods that operate over crisp propositions [36, 39]. When the assumption is violated, the fact-finding method will fail. This seems to be a fundamental barrier when an ego-source is unavailable.

6.7 Discussion and Conclusions

This chapter demonstrates how to perform fusion of subjective opinions from possibly unreliable sources to estimate values of probabilistic state variables. Specifically, a subjective opinion about a state variable summarizes the evidence about the possible probabilities that the state variable takes one of K values. It encodes both the expected probabilities as the information and the amount of evidence that has been collected as the quality of the information. As shown in this chapter, the quality of information represents the spread (or difference) between the actual ground truth probabilities and the expectation information derived from the observations.

When the decision-maker (or its very trusted advisor) has direct observations about many of the state variables or propositions, the decision-maker can use the consistency of his/her observations and the corresponding reported opinions of a particular source to establish a source behavior profile opinion for that particular source. The decision-maker can achieve very effective fusion by accounting for its source behavior opinions of each source in conjunction with the reported opinions from each source. It is demonstrated that it is much more effective to perform the fusion in one shot rather than discounting each source's opinion by its corresponding behavior profile opinion. This is because a discounted opinion is unable to jointly capture the uncertainty of the reported opinion in light of the source's various behaviors.

Simulations of three (good, flipping, and random) source behaviors help to demonstrate the effectiveness of the various fusion methods. When the fusion method models all three behaviors, the fusion leads to a very tight estimate of ground truth that is well characterized by the ground truth. When the fusion method only models good and random behavior, the estimate of the ground truth is not as tight because the method is only able to censor (and not correct for) the flipping behavior, which is usually inconsistent with good behavior. The quality of information is still able to characterize the difference between the estimates and ground truth as long as the unmodeled flipping behavior does not become overly prevalent.

It is possible to perform fusion of a set of subjective opinions when the decision-maker does not have any direct observations to calibrate the behavior profiles of sources. In these cases, fused opinions act as a surrogate for the direct opinions so that inspired by fact-finding methods, one can iterate between fusion and source behavior estimation where the estimates progressively improve as long as the good behaviors occur more frequently than the bad behaviors. The fact-finding principle provides good estimates whose errors are well characterized by the quality of information as long a good behaviors occur more than flipping behaviors.

In general, sources can exhibit more than the three behaviors considered in this chapter. Nevertheless, the two-mode behavior model is a robust behavior model because the random behavior can capture source behaviors intended to move a fused estimate farther from the ground truth. The problem with the two-mode behavior model is that it does not allow fusion method to incorporate "bad" reports by implicitly transforming them into "good" reports. In essence, the two-mode behavior model only enables the fusion to censor (but not correct for) bad behaviors. The three-mode behavior model can correct for flipping behaviors, but it will not be able to correct for other unmodeled behaviors. The insight of the results in this chapter mean that in light of additional source behaviors, the three-mode fusion methods will still achieve good fusion with a meaningful quality of information characterization as long as the decision-maker has direct observations to calibrate the sources. It is just that the fusion performance could be improved by explicitly modeling the behaviors, and methods such as in [27] could be employed to learn new behaviors. Without the direct observation, the fact-finding methods will still be effective as long as the good behavior is the majority behavior exhibited in the collections of reports. Furthermore, the three-mode fact-finding should still beat out two-mode fact-finding because it can use flipped reports as information.

The beta model to characterize source behavior is nice because it captures the idea that sources do not always lie or tell the truth. However, a clever and malicious source would try to be truthful as much as possible to build a good reputation and decide to lie at the moment that makes the decision-maker's organization the most vulnerable. This chapter does not present "the" technique to handle such a case, but this chapter does provide insights in the challenges to successfully deceive or protect from such deception. For instance, the malicious source can only be effective if he/she coordinates his/her lie with other sources and those sources are not drowned out by a larger group of good sources. Likewise, the decision-maker can use other stereotypical or profile information about the source, e.g., see [2, 28], along with risk/benefit analysis to build the sources reputation based upon its past forgone opportunities to cause harm in light of its likely affiliations. In other words, each proposition need not be considered equal in forming the source behavior profile. Furthermore, the fusion methods in this chapter assume independent sources, and as a result, they are vulnerable to coordinating sources. Understanding how the source profile information forms an influence network among sources can lead to better methods. For instance, social EM is a fact-finding method for binary propositions that is resilient to the "echo chamber" effect in social networks [38].

Specific applications will drive the exact source reputation and fusion system that is required. The methods presented in this chapter are generic. While they are not necessarily best for a particular scenario, such as a set of cooperating sources waiting for the exact right time to lie, the methods presented here can serve as the building blocks for customized systems. Overall, there are opportunities to design fusion systems to be resilient to conflicting and malicious sources. However, there are limitations to how resilient the system can be built. The chapter has identified some of these limitation and opportunities.

Acknowledgements Research was sponsored by the US Army Research Laboratory and was accomplished under agreement numbers W911NF-14-1-0199. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory or the US government. The US government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation hereon. Dr. Şensoy thanks the US Army Research Laboratory for its support under grant W911NF-14-1-0199 and The Scientific and Technological Research Council of Turkey (TUBITAK) for its support under grant 113E238.

References

1. V. Bui, R. Verhoeven, J. Lukkien, R. Kocielnik, A trust evaluation framework for sensor readings in body area sensor networks, in *Proceedings of the 8th International Conference on Body Area Networks, BodyNets '13* (ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, 2013), pp. 495–501

2. C. Burnett, T.J. Norman, K. Sycara, Stereotypical trust and bias in dynamic multiagent systems. ACM Trans. Intell. Syst. Technol. **4**(2), 26:1–26:22 (2013)
3. C. Fung, R. Boutaba, *Intrusion Detection Networks: A Key to Collaborative Security* (CRC Press, London, 2013)
4. S. Ganeriwal, L. Balzano, M. Srivastava, Reputation-based framework for high integrity sensor networks. ACM Trans. Sens. Netw. (ToSN) **4**(3), 15 (2008)
5. S. Han, B. Koo, A. Hutter, W. Stechele, Forensic reasoning upon pre-obtained surveillance metadata using uncertain spatio-temporal rules and subjective logic, in *Analysis, Retrieval and Delivery of Multimedia Content*, ed. by N. Adami, A. Cavallaro, R. Leonardi, P. Migliorati (Springer, New York, 2013), pp. 125–147
6. G. Han, J. Jiang, L. Shu, J. Niu, H.-C. Chao, Management and applications of trust in wireless sensor networks: a survey. J. Comput. Syst. Sci. **80**(3), 602–617 (2014)
7. A. Jøsang, A logic for uncertain probabilities. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. **9**(3), 279–311 (2001)
8. A. Jøsang, The consensus operator for combining beliefs. Artif. Intell. J. **142**(1–2), 157–170 (2002)
9. A. Jøsang, Conditional reasoning with subjective logic. J. Multiple-Valued Log. Soft Comput. **15**(1), 5–38 (2009)
10. A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty* (Springer, Cham, 2016)
11. A. Jøsang, R. Ismail, The beta reputation system, in *Proceedings of the Fifteenth Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy*, Bled, June 2002, pp. 48–64
12. A. Jøsang, R. Hayward, S. Pope, Trust network analysis with subjective logic, in *Proceedings of the 29th Australasian Computer Science Conference*, Hobart (2006), pp. 85–94
13. A. Jøsang, J. Diaz, M. Rifqi, Cumulative and averaging fusion of beliefs. Inf. Fusion **11**(2), 192–200 (2010)
14. A. Jøsang, T. Aǆderska, S. Marsh, Trust transitivity and conditional belief reasoning, in *6th IFIP WG 11.11 International Conference, IFIPTM 2012*, Surat, May 2012, pp. 68–83
15. L. Kaplan, M. Şensoy, S. Chakraborty, G. de Mel, Partial observable update for subjective logic and its application for trust estimation. Inf. Fusion **26**, 66–83 (2015)
16. S. Kotz, N. Balakrishnan, N.L. Johnson, *Continuous Multivariate Distributions*, vol. 1 (Wiley, New York, 2000)
17. T.R. Levine, *Encyclopedia of Deception* (SAGE Publications, Los Angeles, 2014)
18. Y. Liu, K. Li, Y. Jin, Y. Zhang, W. Qu, A novel reputation computation model based on subjective logic for mobile ad hoc networks. Futur. Gener. Comput. Syst. **27**(5), 547–554 (2011)
19. T.K. Moon, The expectation-maximization algorithm. IEEE Signal Process. Mag. **13**(6), 47–60 (1996)
20. T. Muller, P. Schweitzer, On beta models with trust chains, in *Proceedings of Trust Management VII: 7th IFIP WG 11.11 International Conference*, Malaga (2013), pp. 49–65
21. N. Oren, T.J. Norman, A. Preece, Subjective logic and arguing with evidence. Artif. Intell. **171**(10), 838–854 (2007)
22. J. Pasternack, D. Roth, Making better informed trust decisions with generalized fact-finding, in *IJCAI* (Spatial Cognition, Bremen, 2011), pp. 2324–2329
23. K. Regan, P. Poupart, R. Cohen, Bayesian reputation modeling in e-marketplaces sensitive to subjecthity, deception and change, in *Proceedings of the 21st National Conference on Artificial Intelligence* (AAAI Press, Menlo Park, 2006), pp. 1206–1212
24. M. Şensoy, P. Yolum, Experimental evaluation of deceptive information filtering in context-aware service selection, in *International Workshop on Trust in Agent Societies* (Springer, Berlin/Heidelberg, 2008), pp. 326–347
25. M. Sensoy, G. de Mel, T. Pham, L. Kaplan, T.J. Norman, TRIBE: trust revision for information based on evidence, in *Proceedings of 16th International Conference on Information Fusion*, Istanbul (2013)

26. M. Şensoy, L. Kaplan, G. Ayci, G. de Mel, FUSE-BEE: fusion of subjective opinions through behavior estimation, in *18th International Conference on Information Fusion*, Washington, DC (2015), pp. 558–565
27. M. Şensoy, L. Kaplan, G. de Mel, T.D. Gunes, Source behavior discovery for fusion of subjective opinions, in *19th International Conference on Information Fusion*, Heidelberg (2016), pp. 138–145
28. M. Şensoy, B. Yilmaz, T.J. Norman, Stage: stereotypical trust assessment through graph extraction. Comput. Intell. **32**(1), 72–101 (2016)
29. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976)
30. P. Smets, The combination of evidence in the transferable belief model. IEEE Trans. Pattern Anal. Mach. Intell. **12**(5), 447–458 (1990)
31. W.T.L. Teacy, J. Patel, N.R. Jennings, M. Luck, TRAVOS: trust and reputation in the context of inaccurate information sources. Auton. Agents Multi-Agent Syst. **12**(2), 183–189 (2006)
32. W.L. Teacy, M. Luck, A. Rogers, N.R. Jennings, An efficient and versatile approach to trust and reputation using hierarchical Bayesian modelling. Artif. Intell. **193**, 149–185 (2012)
33. D. Wang, C. Huang, Confidence-aware truth estimation in social sensing applications, in *12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Seattle, June 2015, pp. 336–344
34. D. Wang, L. Kaplan, T. Abdelzaher, C.C. Aggarwal, On credibility estimation tradeoffs in assured social sensing. IEEE J. Sel. Areas Commun. **31**(6), 1026–1037 (2013)
35. D. Wang, L. Kaplan, T.F. Abdelzaher, Maximum likelihood analysis of conflicting observations in social sensing. ACM Trans. Sens. Netw. (ToSN) **10**(2), 30 (2014)
36. D. Wang, T. Abdelzaher, L. Kaplan, *Social Sensing: Building Reliable Systems on Unreliable Data* (Morgan Kaufmann, Waltham, 2015)
37. A. Whitby, A. Jøsang, J. Indulska, Filtering out unfair ratings in Bayesian reputation systems. ICFAIN J. Manag. Res. **4**(2), 48–64 (2005)
38. S. Yao, S. Hu, S. Li, Y. Zhao, L. Su, L. Kaplan, A. Yener, T. Abdelzaher, On source dependency models for reliable social sensing: algorithms and fundamental error bounds, in *IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, Nara (2016), pp. 467–476
39. X. Yin, J. Han, P.S. Yu, Truth discovery with multiple conflicting information providers on the web. IEEE Trans. Knowl. Data Eng. **20**(6), 796–808 (2008)
40. B. Yu, M.P. Singh, Detecting deception in reputation management, in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems* (ACM, New York, 2003), pp. 73–80