Stability of Weighted Majority Voting under Estimated Weights

Shaojie Bai
Zhejiang University
Hangzhou, China
white.shaojie@gmail.com

Dongxia Wang*
Zhejiang University &
ZJU-Hangzhou Global Scientific and Technological Innovation Center
Hangzhou, China
dxwang@zju.edu.cn

Tim Muller
University of Nottingham
Nottingham, United Kingdom
tim.muller@nottingham.ac.uk

Peng Cheng*
Zhejiang University
Hangzhou, China
saodiseng@gmail.com

Jiming Chen
Zhejiang University
Hangzhou, China
cjm@zju.edu.cn

ABSTRACT

Weighted Majority Voting (WMV) is a well-known decision making rule. The weights of sources are determined by the probabilities that sources provide accurate information (*trustworthiness*). However, in reality, the trustworthiness is usually not a known quantity to the decision maker – they have to rely on an estimate called *trust*. An algorithm that computes trust is called *unbiased* when it has the property that it does not systematically overestimate or underestimate the trustworthiness. To formally analyze the uncertainty to the decision process brought by such unbiased trust values, we introduce and analyze two important properties of WMV: *Stability of Correctness* and *Stability of Optimality*. Stability of Correctness measures the difference between the decision accuracy that the decision maker believes he can achieve and the accuracy he actually achieves. We prove Stability of Correctness absolutely holds for WMV – the difference is 0. Stability of Optimality measures the difference between the actual accuracy of decisions made using trust values, and those made using trustworthiness values. We find a relatively tight upper bound on the Stability of Optimality, meaning that, although using (unbiased) trust values is suboptimal compared to using the true trustworthiness values, the difference is small. Meanwhile, a counter-intuitive observation is that while distributions of trustworthiness influence the Stability of Optimality, the number of sources barely influences it. We also provide an overview of how sensitive decision accuracy is to the changes in trust and trustworthiness.

KEYWORDS

Weighted Majority Voting, Trust, Stability of Decision Making

ACM Reference Format:
Shaojie Bai, Dongxia Wang*, Tim Muller, Peng Cheng*, and Jiming Chen. 2024. Stability of Weighted Majority Voting under Estimated Weights. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 9 pages.

* Corresponding author.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1 INTRODUCTION

Crowd wisdom has been playing a fundamental role in helping make better decisions in many scenarios, e.g., hiring workers for labeling tasks in crowdsourcing [25], aggregating classifiers for prediction in ensemble learning [21], asking for opinions of reliability in online rating systems [7, 15], etc. Decisions are derived based on aggregating the information or feedback from a collection of sources, the quality of which can be variable. It can be inaccurate due to lack of expertise, mistakes or malice, e.g., low-quality labeling for machine learning, fake ratings introduced by sellers to promote their reputation, etc.

Among the aggregation mechanisms, Weighted Majority Voting has long been a popular one. Basically, each source supports an option and is assigned a weight. WMV chooses the feedback option that is supported by sources with the maximal total weight. WMV has been seeing its use in a variety of domains ranging from voting [34], crowdsourcing [9], classification [24] to trust systems [46] and even distributed systems [40]. In different contexts, the weight of a source can mean differently. For instance, in determining a collective choice that is widely acceptable to individuals with diverse preference [38], WMV is used for preference aggregation and the weight means the importance of an individual. We are more interested in the scenarios where there is a notion of *correctness* (or accuracy) of decisions, and the weight of a source depends on how trustworthy it is in providing the feedback that corresponds to the correct decision, denoted as *trustworthiness*, which is usually modelled as a probability value. The examples include the aggregation of the crowd-sourced labels [23], the crowd-sensed navigation data [20], or the outputs of the multiple classifiers [27], etc.

While WMV is proven to be optimal when source *trustworthiness* is given [34][1], in practice, decision makers have to resort to an estimation or a belief (denoted as *trust*) that may not equal to the actual values of trustworthiness[2]. Deviation in estimation may

[1] To provide feedback independently is also required for the optimality of WMV [33, 34].
[2] Note that we use "trust" and "trustworthiness" to differentiate between what a decision maker trusts or estimates as the probability of a source's suggesting correctly (regardless of whether he believes in the value or is aware that its just an estimate), and the actual probability value (Refer to their definition difference [41]).

decrease decision accuracy. There exists plenty of effort to improve the estimation of source trustworthiness by learning from historical data (e.g., direct observations or indirect evidences), with a principle that more data increases the confidence in the estimation [3, 43]. Several approaches even treat the belief about source trustworthiness as its actual values [29, 30]. However, no algorithm always produces perfect trust values. It is worth studying how the quality of the trust values impacts the decision quality of WMV. Is WMV able to maintain a tolerant level of decision incorrectness with the inaccuracy in the estimation bounded, meaning having certain levels of stability w.r.t the inaccurate estimation?

In this paper, we propose a formal analysis of the stability properties of WMV. Firstly, we study how sensitive the decision accuracy is to the changes in source trust and trustworthiness, with both the arguments taking fixed values. We find that unsurprisingly, decision accuracy decreases with the increasing deviation from trust to trustworthiness, and a sufficiently small deviation barely influences the accuracy. Besides, compared with overestimating, underestimating trustworthiness is usually less harmful to the decision accuracy. Secondly, we study the influence of trust and trustworthiness in a statistical way. Considering that a decision maker may sometimes overestimate source trustworthiness while sometimes underestimate it, the expectation remains correct – unbiased estimation[3]. We define two types of stability based on such unbiased estimation: *Stability of Correctness* and *Stability of Optimality*. *Stability of Correctness* reasons whether the decision accuracy a decision maker believes he achieves (i.e., the accuracy he computes with trust) equals what he actually achieves (i.e., the accuracy computed with trustworthiness). We prove that whatever distribution source trustworthiness follows, as long as the estimation is unbiased, a decision maker gets the accuracy as if the trustworthiness is known – absolute stability. This means that the shape and variance of trustworthiness are irrelevant to the Stability of Correctness.

Stability of Optimality reasons whether the decisions made based on unbiased trust values are as good as those made based on trustworthiness. Considering trustworthiness is usually unknown, Stability of Optimality measures the *gap* between the practical situation where the decision maker decides with trust, and where (magically) he has access to the actual trustworthiness. We prove that *Stability of Optimality* does not hold for WMV, but the degradation in decision accuracy caused by the incorrect but (averagely) unbiased trust is relatively tightly bounded. That is, decision accuracy with unbiased trust will not be too far off the theoretically determined value. Moreover, unlike *Stability of Correctness*, the distribution of trustworthiness influences the upper bound of that accuracy gap, and also determines how well the accuracy can be in the ideal situation, namely where trustworthiness is given. Last but not least, while it may usually be perceived that more sources improve accuracy, we observe counterintuitively that source number influences little on the accuracy gap.

The rest of this paper is organized as follows. In Section 2 the related work is presented. In Section 3 we introduce a formal framework to study WMV decision rule. In Section 4 we present how trust and trustworthiness influence the decision accuracy of WMV. In Section 5 we analyze the two types of stability. The numerical analysis is also performed where needed to demonstrate theories.

[3]Generally, the estimation error always exists, but it is relatively small and can be zero on average with sufficient data [3, 13]

2 RELATED WORK

The Weighted Majority Voting rule has been studied in several domains, e.g., decision theory, voting theory, management science, and receiving various applications. We focus on the scenarios where the weight of a source or "voter" is determined by how trustworthy it is in suggesting the correct decision. Some approaches utilizing WMV assume source trustworthiness is given [4, 34], although in practice it is usually unknown. Plenty of work focuses on modeling and learning source trustworthiness from observation and interaction history [15, 42, 47]. Some researchers model trust as a probability value. To get trust, they either rely on frequency estimation by counting the times of making the right decisions [3], or solving an optimization problem based on their models by minimizing the decision error rate [36] or maximizing the likelihood [10, 26, 31]. Moreover, model-checking-based methods are also applied in quantifying the probability of trust on individual agents, representing the agent's own beliefs [2, 11, 12, 39]. Besides, trust also can be modeled as a random variable. Bayesian models have also been proposed and applied to this problem by [19, 35, 37], combining the prior knowledge and the observations to infer the trustworthiness. Expectation Maximization-based methods are also proposed to estimate source trustworthiness and the correct decision at the same time, via iterative updating [9, 48].

Such learned trust is sometimes treated as an estimation of the source trustworthiness with the deviation considered [14, 44], while sometimes treated equivalently as trustworthiness, namely as the probability of a source suggesting correctly and is further used to evaluate decision accuracy [18, 24, 28]. However, trust essentially represents the belief of a decision maker about the source quality, which may deviate from the actual probability. And he may not gain the claimed decision accuracy based on trust.

Besides the efforts in modeling and learning trustworthiness, there exists work that theoretically analyzes how trustworthiness and trust would influence decision accuracy, which is most relevant to ours. Given trustworthiness, the decision accuracy of WMV is analyzed in [4] without considering the learning process of trustworthiness. On the other hand, some other work takes the learning process into consideration. To measure the estimation quality, the decision accuracy bounds for learning algorithms have been proposed through PAC techniques in [16, 22, 45]. Considering trust is derived from finite samples, some researchers then study precise characterizations of the relationship between the decision accuracy and the sample size in [14]. More recently, tighter bounds for decision accuracy under arbitrary estimation are provided, ignoring particular assumptions for trustworthiness [27]. Unfortunately, none of them have analyzed the relationship between the estimate error and the decision accuracy of WMV in a quantitative way.

3 PRELIMINARIES

In this section, we outline a formal framework to support our study of the stability of Weighted Majority Voting decision rule. Note that the capital letters represent random variables, and the lower cases represent non-random variables. The bold letters represent

a vector of multiple variables, and the non-bold letters represent single variables.

Consider a decision-making scenario, a decision maker is faced with multiple possible decisions $O = \{o_1, \ldots, o_K\}$ and only one of them is correct. The random variable O determines which of the options is actually correct, e.g., $O{=}o_1$ if o_1 is the correct decision. The decision maker receives feedback from a set of sources, $\mathcal{S} = \{s_1, .., s_n\}$. The random variable F_i, with f_i denoting an outcome, represents the feedback of source s_i. The feedback may or may not correspond to the correct decision. For WMV, we assume[4] a one-to-one correspondence between the feedback that suggests the correct decision and the correct decision itself, and denote $F_i = O$ iff f_i suggests correctly, $F_i \in O$. Feedback of all the sources is represented by random variable $\boldsymbol{F} : \boldsymbol{F} = (F_1, \ldots, F_n)$, with $\boldsymbol{f}{:}\boldsymbol{f}{=}(f_1, \ldots, f_n)$ denoting an outcome, and its sample space is defined as $\mathcal{F}{:}\boldsymbol{f}{\in}\mathcal{F}$. A decision mechanism is a function: $\mathcal{D} : \mathcal{F}{\rightarrow}O$. The quantity that the decision maker wants to maximize is the probability of making the correct decisions (which we shorthand as *decision accuracy* or *decision correctness* throughout the paper): $\mathbb{P}(\mathcal{D}(\boldsymbol{F}) = O)$.

we define Υ_i as a $\{-1, 1\}$-indicator random variable of whether source s_i suggests the correct decision and v_i as one of its outcome: $\Upsilon_i{=}1$ if $F_i{=}O$ and $\Upsilon_i{=}-1$ if $F_i{\neq}O$. For the indicator variables of all the sources i.e., $\boldsymbol{\Upsilon} : \boldsymbol{\Upsilon} = (\Upsilon_1, \ldots, \Upsilon_n)$, one of its samples is an *indicator vector* i.e., $\boldsymbol{v}{=}(v_1, \ldots, v_n), \boldsymbol{v} \in \mathcal{T}$. $\mathcal{T}$ denote the sample space of $\boldsymbol{\Upsilon}$. Let $-\boldsymbol{v}{=}(-v_1, \ldots, -v_n)$ denote the *opposite* indicator vector of $\boldsymbol{v}$ where source indicators are flipped. The set of all the possible feedback under $\boldsymbol{v}$ is denoted as $\mathcal{F}_{\boldsymbol{v}}$.

We use the following running example in this section to demonstrate the relevant concepts.

Example 1. *There are three sources $\mathcal{S}{=}\{s_1, s_2, s_3\}$. If $O = \{A, B\}$, $O{=}B$ and the indicator vector $\boldsymbol{v} = (1, 1, -1)$, then $\boldsymbol{f} = (B, B, A)$ and $\mathcal{F}_{\boldsymbol{v}} = \{(B, B, A), (A, A, B)\}$.*

When decision "correctness" is a concern, Weighted Majority Voting usually considers how probable each source suggests the correct decision. For source i, let $\mathbb{P}(F_i{=}O) = p_i$ and $\boldsymbol{p} = (p_1, \ldots, p_n)$. Hence $\mathbb{P}(\Upsilon_i = 1) = p_i$. We refer to p_i as the *trustworthiness* of source s_i. In practice, the trustworthiness of a source is usually unknown to a decision maker. And an estimation is used, denoted as $\hat{p}_i$, with $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots \hat{p}_n)$. We call the value $\hat{p}_i$ *trust*, which represents the subjective estimation or belief of the decision maker regarding how probable a source suggests correctly. There exist multiple ways to compute $\hat{p}_i$, e.g., counting the frequency of making correct decisions, or Bayesian learning methods based on prior interaction data. In the literature, the trustworthiness of a source can have different meanings, e.g. honesty of an agent in a rating system [33], competency of a voter [8], reliability of a worker in crowdsourcing [9], correctness of a sensor in crowdsensing [32], etc. Whatever the meanings, p_i represents an intrinsic quality or the fact that how probable the source reports correctly, while $\hat{p}_i$ represents how the decision maker thinks of or estimates that probability [41]. We assume sources independently provide feedback, hence $\mathbb{P}(\boldsymbol{\Upsilon} = \boldsymbol{v}) = \prod_{i:v_i=1} p_i \cdot \prod_{i:v_i=-1}(1{-}p_i)$.

In Example 1, suppose $\boldsymbol{p} = (0.6, 0.6, 0.7)$, the estimation of $\boldsymbol{p}$ by a decision maker may be inaccurate: $\hat{\boldsymbol{p}} = (0.6, 0.7, 0.8)$.

[4]For model general decision-making scenarios, the options of feedback and that of decisions may not necessarily equal and may take a many-to-one mapping.

Below, we introduce the Weighted Majority Voting (WMV) decision scheme. It can be treated as an extension of the more commonly known *Majority Voting* decision scheme. The difference is that Majority Voting treats sources without distinguishing, while WMV assigns sources different weights. The weight of a source is usually determined by how trustworthy its feedback is. Formally:

Definition 1 (Weighted Majority Voting $\mathcal{D}_W$). *Given a set of n sources $\mathcal{S}$, their trustworthiness $\boldsymbol{p}$ and independent feedback $\boldsymbol{f}$, $\mathcal{D}_W$ makes decisions via the function [33, 34]:*

$$\mathcal{D}_W(\boldsymbol{f}) = argmax_{o \in O}\left(\sum_{i:f_i=o} w_i\right) \tag{1}$$

where $f_i \in O$, $w_i = \log(p_i/1{-}p_i)$ with $p_i \geq 0.5$.

To give an instance, consider Example 1, suppose $\boldsymbol{p} = (0.6, 0.6, 0.9)$, $O = (A, B)$, $O = B$ and $\boldsymbol{v} = (1, 1, -1)$, then $\boldsymbol{f} = (B, B, A)$. $w_1 \approx 0.18, w_2 \approx 0.18, w_3 \approx 0.60$. since $w_1 + w_2 < w_3$, $\mathcal{D}_W(\boldsymbol{f}) = A$.

Here $p_i \geq 0.5$ and the log weight function are well-known for classical WMV in the literature [17, 34], where trust and trustworthiness are not distinguished. The assumption $p_i \geq 0.5$ means that sources with $p_i < 0.5$ are ignored. For a source with $p_i < 0.5$, a decision maker may assign negative weight to its feedback. Or he can just simply reverse the vote of the source (e.g., replacing the reported option A with C). But if either the operation is realized by the malicious sources, they can push the decision to a wrong one by reporting correctly, purposely reducing the chance of correct option being selected. Therefore, it is in the interest of the decision maker to ignore such sources.

It has been shown in the literature that the decision accuracy of WMV is determined by the indicator vectors where it always decides correctly (an example is where all sources report correctly). For such indicator vectors, whether a decision is correct is not influenced by the feedback of the sources that suggest incorrectly. To give an opposite example, consider Example 1. Suppose $O = \{A, B, C\}$, $\boldsymbol{p} = (0.70, 0.65, 0.65)$ and $\boldsymbol{v} = (1, -1, -1)$ (only s_1 reports correctly). We get $(w_1, w_2, w_3) \approx (0.37, 0.27, 0.27)$. Both the feedback $\boldsymbol{f} = (A, B, C)$ and $\boldsymbol{f'} = (A, C, C)$ are possible under $\boldsymbol{v}$ (both belong to $\mathcal{F}_{\boldsymbol{v}}$. However, WMV decides correctly by choosing A under $\boldsymbol{f}$ and decides incorrectly by choosing C under $\boldsymbol{f'}$. Whether WMV decides correctly is influenced by what incorrect feedback is. Given the same $\boldsymbol{p}$, for $\boldsymbol{v'} = (1, 1, -1)$, it can be seen that whatever s_3 reports, WMV can always decides correctly by trusting s_1, s_2.

Let $\mathbb{D}_W(\boldsymbol{p})$ denote the set of all the indicator vectors where WMV always decides correctly when using $\boldsymbol{p}$, namely $\mathbb{D}_W(\boldsymbol{p}) = \{\boldsymbol{v} | \mathcal{D}_W(\boldsymbol{f}) = O, \boldsymbol{f} \in \mathcal{F}_{\boldsymbol{v}}\}$. It has been proven that $\mathbb{D}_W(\boldsymbol{p}) = \{\boldsymbol{v} | \mathbb{P}(\boldsymbol{v}) \geq \mathbb{P}(-\boldsymbol{v})\}$ and the decision accuracy of $\mathcal{D}_W$ is (Refer to [4, 33, 34]):

$$\begin{aligned}\mathbb{P}(\mathcal{D}_W(\boldsymbol{F}) = O) &= \sum_{\boldsymbol{v}:\boldsymbol{v} \in \mathbb{D}_W(\boldsymbol{p})} \mathbb{P}(\boldsymbol{v}) \\ &= \sum_{\boldsymbol{v}:\mathbb{P}(\boldsymbol{v}) \geq \mathbb{P}(-\boldsymbol{v})} \left(\prod_{i:v_i=1} p_i \cdot \prod_{i:v_i=-1}(1{-}p_i)\right)\end{aligned} \tag{2}$$

Equation 2 indicates that the accuracy of WMV is determined by the probabilities of indicator vectors, which depend on the trustworthiness values of the sources.

In Example 1, if $\boldsymbol{p} = (0.6, 0.6, 0.9)$, then $\mathbb{D}_W(\boldsymbol{p}) = \{(1,1,1), (-1,1,1), (1,-1,1), (-1,-1,1)\}$ and the decision accuracy is 0.9 (i.e., a.l.a source s_3 reports correctly).

WMV has been proved to be optimal when trustworthiness $\boldsymbol{p}$ and the log weight function are used for decision making and the sources are independent in providing feedback [34].

In practice, when trustworthiness is unknown, the weight assigned to each source depends on the trust, that is, $w_i = \log(\hat{p}_i/1-\hat{p}_i)$. Besides, the decision maker computes the probabilities of indicator vectors with trust values, which we use the subscript $\mathbb{P}_{\hat{\boldsymbol{p}}}$ to distinguish from their actual probabilities: $\mathbb{P}_{\hat{\boldsymbol{p}}}(\Upsilon = \boldsymbol{v}) = \prod_{i:v_i=1} \hat{p}_i \cdot \prod_{i:v_i=-1}(1-\hat{p}_i)$. With $\mathbb{P}(\boldsymbol{v})$ replaced by $\mathbb{P}_{\hat{\boldsymbol{p}}}(\boldsymbol{v})$, the decisions would always be correct for those indicator vectors which the decision maker thinks are more probable than their opposite, namely $\mathbb{D}_W(\hat{\boldsymbol{p}}) = \{\boldsymbol{v} | \mathbb{P}_{\hat{\boldsymbol{p}}}(\boldsymbol{v}) \geq \mathbb{P}_{\hat{\boldsymbol{p}}}(-\boldsymbol{v})\}$. As a result, $\mathbb{D}_W(\hat{\boldsymbol{p}})$ and $\mathbb{D}_W(\boldsymbol{p})$ may be different. In Example 1, if $\boldsymbol{p} = (0.6, 0.6, 0.9)$ and $\hat{\boldsymbol{p}} = (0.8, 0.6, 0.8)$, then $(-1,-1,1) \in \mathbb{D}_W(\boldsymbol{p})$ while its opposite indicator vector $(1,1,-1) \in \mathbb{D}_W(\hat{\boldsymbol{p}})$. This may result in different decision accuracy. We introduce $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$ to distinguish:

$$\mathbb{P}(\mathcal{D}_W(\boldsymbol{F}) = O) \triangleq \omega(\hat{\boldsymbol{p}}, \boldsymbol{p}) = \sum_{\boldsymbol{v} \in \mathbb{D}_W(\hat{\boldsymbol{p}})} \mathbb{P}(\boldsymbol{v}) \tag{3}$$

The first parameter of the function $\omega()$ represents the value used for decision making, and the second parameter represents the value used to compute the probability of deciding correctly. For $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$, decisions are made using trust values $\hat{\boldsymbol{p}}$, while the decision accuracy that the decision maker actually obtains still depends on source trustworthiness, which challenges the optimality of WMV.

Generally, both the parameters of $\omega()$ can be either trust or trustworthiness, and we assume that the parameter (either trust or trustworthiness) used for decision-making is at least 0.5. Trust values are, by definition, known to the decision-maker. Therefore, it's reasonable to apply the assumption for trust, meaning ignoring sources with trust below 0.5. For trustworthiness, we assume it is at least 0.5 only when it is used to decide (e.g., in Section 4.1), and otherwise, its value ranges from $(0, 1)$ (e.g., in Section 4.2, 4.3 and 5).

Depending on what we equip the parameters with, trust or trustworthiness, we will obtain different meanings for decision accuracy as follows. The quantity $\omega(\boldsymbol{p}, \boldsymbol{p})$ denotes the "ideal" decision accuracy, where the decision maker knows and uses the trustworthiness values to decide and compute. The quantity $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$ denotes the "practical" decision accuracy, where the decision maker decides with the trust values $\hat{\boldsymbol{p}}$, but the accuracy he actually achieves depends on trustworthiness. The quantity $\omega(\hat{\boldsymbol{p}}, \hat{\boldsymbol{p}})$ denotes the "perceived" decision accuracy that the decision maker thinks he can obtain (decides and computes accuracy with trust), while the actual accuracy may not equal $\omega(\hat{\boldsymbol{p}}, \hat{\boldsymbol{p}})$.

4 PARAMETER SENSITIVITY

In this section, we analyze how changes in the values of trust and trustworthiness influence the decision accuracy or the correctness of WMV. There are several ways: 1) how the decision accuracy changes when the trustworthiness and trust change simultaneously; 2) how the decision accuracy changes with trustworthiness when trust remains constant; 3) how the decision accuracy changes with trust when trustworthiness remains constant. If the changes show relatively little effect on the correctness, then we can say that WMV is not very sensitive to the parameters. Sensitivity relates to stability, the analysis in this section provides several important insights for the analysis in the next section.

We will also take numerical analysis based on the setting in the following running Example 2 to further illustrate the theories.

Example 2. *There are four sources $\mathcal{S}$=$\{s_1, s_2, s_3, s_4\}$ and their trustworthiness values are $\boldsymbol{p} = (0.8, 0.75, 0.7, 0.6)$ respectively.*

4.1 Direct Sensitivity Analysis

Here, we analyze the case where the parameters used for making decisions and that for computing accuracy are equal. There are two different rationales for doing this, but the mathematics is identical for both. First, consider that the decision maker is given the actual trustworthiness values to make decisions. Second, consider analyzing the sensitivity of the beliefs of the decision maker. Assume the decision maker only knows trust values, and uses them to compute their belief about how probable a decision is correct.

For simplicity, we use trustworthiness everywhere, but the analysis remains unchanged when using trust instead (simply put a hat on all p's and $\boldsymbol{p}$'s). Observe that if trustworthiness of only 1 source varies, then decision accuracy would appear to be a piecewise linear non-decreasing convex function. In Figure 1(a), we depict Example 2, with each plot representing a source trustworthiness variable.

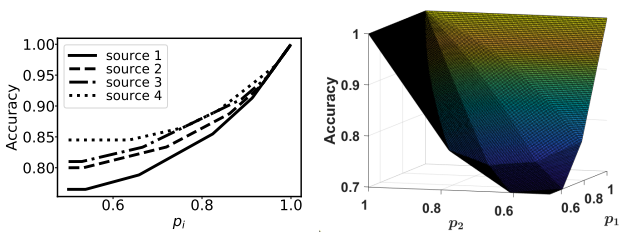
Lemma 1. *Let $f(p_i) = \omega(\boldsymbol{p}, \boldsymbol{p})$, where p_j is constant for $j \neq i$. The function $f(p_i)$ is a piecewise linear non-decreasing convex function.*

Sketch of Proof. The computation for correctness of a decision can be characterized as $p_i \cdot x + (1-p_i) \cdot y$, where $x - y \geq 0$ and this coefficient increases with p_i increasing. Each decision based on corresponding $\mathbb{D}_W(\boldsymbol{p})$ represents a non-decreasing line. □

Generally, if a source is more probable to be trustworthy, the decision will be better and improved faster. Figure 1(a) illustrates Lemma 1. The accuracy of WMV is determined by comparing the $\mathbb{P}(\boldsymbol{v})$ and $\mathbb{P}(-\boldsymbol{v})$ for all the indicator vectors. The relation between $\mathbb{P}(\boldsymbol{v})$ and $\mathbb{P}(-\boldsymbol{v})$ either remains or changes, depending on the value of p_i and how much it changes. Intuitively, this results in the piecewise characteristics of $f(p_i)$. Moreover, we can vary trustworthiness values of multiple (or even all) sources. If we vary k trustworthiness, then we get an k-dimensional piecewise surface. In Figure 1(b), we depict our running example with p_1 and p_2 on the two axes, and p_3 and p_4 remaining constant. The surface appears like a collection of intersecting planes, but in fact, the graph consists of surfaces described by a polynomial, rather than a linear one.

In Lemma 1, we assume that trustworthiness of all the sources remains constant and independent. However, sources can collude to influence decisions, or one can update the trust values of multiple sources at a time. For such situations, we assume the trustworthiness values of multiple sources are consistently equal, meaning they are not independent.

Lemma 2. *In the special case of the identical sources, let $f(p) = \omega(\boldsymbol{p}, \boldsymbol{p})$, where $p_1 = \cdots = p_m = p$, $m \leq n$ and p_j are constant, $j > m$. The function $f(p)$ is a piecewise non-decreasing function, and it is concave (linear or strictly concave) in each segment.*



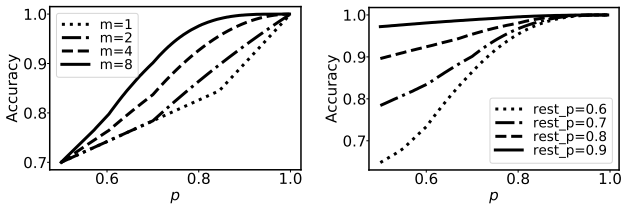
(a) One Source p_i (b) Two sources p_1,p_2

Figure 1: Sensitivity of $\mathcal{D}_W$ to $\boldsymbol{p}$ when $\boldsymbol{p} = \hat{\boldsymbol{p}}$

Sketch of Proof. The computation for correctness can be characterized as a summation: $\omega(\boldsymbol{p}, \boldsymbol{p}) = \sum_i g_i(p)$. For each $g_i(p)$, it meets piecewise non-decreasing property, and concave property in each segment. Thus, $\omega(\boldsymbol{p}, \boldsymbol{p})$ also holds. □

Note that in Lemma 2, if m=n, meaning all the sources are identical, WMV becomes the classical Majority Voting decision rule and $f(p)$ becomes a concave monotonically increasing function [5].

Figure 2(a) shows how the decision accuracy $\omega(\boldsymbol{p}, \boldsymbol{p})$ changes with varying p and m values where there are originally 2 sources with the same $rest\_p = 0.7$, then m identical sources join with their $p \geq 0.5$ and $n = m + 2$. Given m value, $f(p)$ increases piecewisely with p, and specifically in each segment, it is concavely increasing. Given p value, $f(p)$ increases monotonically with m. In Figure 2(b), we fix $n = 10, m = 6$ and vary the trustworthiness of the m identical sources p and the rest sources $rest\_p$. This figure illustrates that even the rest sources are in minority, but the higher $rest\_p$ is, the more insensitive the decision to p is. Besides, it demonstrates that when the trustworthiness of multiple sources updates in a particular way, the variation characteristic of the decision accuracy may be captured and described.



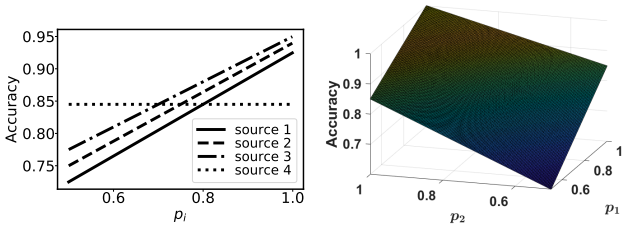
(a) $rest\_p = 0.7$ (b) $n = 10$

Figure 2: Accuracy of $\mathcal{D}_W$ with m sources being identical

4.2 Trustworthiness Sensitivity Analysis

Next, we analyze the cases where trustworthiness and trust are not identical. The decisions are made based on the trust values $\hat{\boldsymbol{p}}$, while the probability of each indicator vector is determined by $\boldsymbol{p}$. The probability of deciding correctly is $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$. In this section, trustworthiness varies with trust value fixed. Recall Equation 3, this means that decisions remain unchanged for given feedback (as the set $\mathbb{D}_W(\hat{\boldsymbol{p}})$ remain unchanged), while decision accuracy $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$ may change with trustworthiness.

If only one parameter p_i varies in $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$, then the resulting decision accuracy is a non-decreasing function, which follows trivially from the proof of Lemma 1. In fact, the line corresponds to one of the line segments from the piece-wise linear graph from the previous section as depicted in Figure 3(a). Besides, we can have multiple variables as before. The surface obtained is non-decreasing and polynomial. The surface corresponds to one of the fragments from the graph discussed in the previous section. A 2d example is depicted in Figure 3(b). The result shows that the decision accuracy has a unique continuous differentiable function, rather than a piecewise function with different functions in different segments.



(a) One Source p_i (b) Two sources p_1,p_2

Figure 3: Sensitivity of $\boldsymbol{p}$ with $\hat{\boldsymbol{p}}$ fixed

4.3 Trust Sensitivity Analysis

Alternatively, we can take trust to be variable and trustworthiness to be fixed. Different from before, the actual probability of each indicator vector ($\mathbb{P}(\boldsymbol{v})$) now remains unchanged, but the decisions may change (as $\mathbb{P}_{\hat{\boldsymbol{p}}}(\boldsymbol{v})$ and accordingly $\mathbb{D}_W(\hat{\boldsymbol{p}})$ may change). Then we can analyze what happens if a trust value used for decision making moves away from the actual trustworthiness in either direction.

First, consider the case where we vary the trust value of only one source in $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$. We get a uni-modal discontinuous staircase function, which is non-decreasing when $\hat{p}_i < p_i$ and non-increasing when $\hat{p}_i > p_i$. Figure 4(a) depicts Example 2 with one variable. Second, when trust values of multiple sources are variable, the resulting surface consists of flat fragments at different heights, with an increasing height with proximity to the point $\hat{\boldsymbol{p}} = \boldsymbol{p}$. Figure 4(b) depicts Example 2 with $\hat{p}_1$ and $\hat{p}_2$ being the variables. Generally:

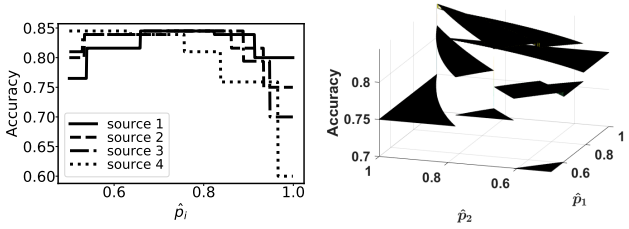
Lemma 3. *Let $f(\hat{\boldsymbol{p}}) = \omega(\hat{\boldsymbol{p}}, \boldsymbol{p})$, where the trustworthiness $\boldsymbol{p}$ is constant. The function $f(\hat{\boldsymbol{p}})$ is a discontinuous staircase function consisting of flat plateaus. Decision accuracy reaches the maximum at the plateau containing the point $\hat{\boldsymbol{p}} = \boldsymbol{p}$.*

Sketch of Proof. The probability that a decision is correct depends on $\boldsymbol{p}$, which is constant. Changing $\hat{\boldsymbol{p}}$ does not affect the probability that a decision is correct, until it reaches a point where it changes the actual decision away from the optimum. Then, there is a discontinuous step to a new platform. □

An insight is that the nearby points are more likely to be on the same plateau. In other words, there is an area of trust values around the trustworthiness values, meaning small estimation deviation

may be unlikely to affect the accuracy. However, it is possible that a certain trustworthiness $\boldsymbol{p}$ is exactly at a border (or corner) of a plateau, meaning that even a tiny difference between trustworthiness and trust can lead to a staircase difference in correctness. The positive news is that the plateaus directly bordering the one containing $\boldsymbol{p}$ are still more often correct than the ones further away.

Besides, while both underestimation and overestimation cause wrong judgment on $\mathbb{P}(\boldsymbol{v})$ vs. $\mathbb{P}(-\boldsymbol{v})$, the numerical (Figure 4) results imply that overestimation perhaps results in the worse accuracy degradation compared with underestimation. Our intuition is that, if there is a high p-valued source, then that source tends to have a lot of sway on the vote, so any inaccuracies will be noticeable, whereas a low p-valued source tends to only matter in cases where the vote is tight, and thus any inaccuracies tend to matter less. From a micro perspective, the underlying reason might be that overestimation of a trustworthiness value makes the difference $|\mathbb{P}(\boldsymbol{v}) - \mathbb{P}(-\boldsymbol{v})|$ also overestimated, while for underestimation, the difference would be underestimated. Therefore, when the estimation error is typically inevitable, it is better to underestimate trustworthiness.



(a) One Source $\hat{p}_i$ (b) Two sources $\hat{p}_1$,$\hat{p}_2$

Figure 4: Sensitivity of $\hat{p}$ with p fixed

5 STABILITY

The results of the Parameter Sensitivity Section are unsurprising. Increasing trustworthiness typically increases correctness, and cannot decrease correctness. Hence, if a source is believed to decide correctly with probability $\hat{p}$, while its trustworthiness $p < \hat{p}$, then the actual correctness achieved is lower than what the decision maker believes: $\omega(\hat{\boldsymbol{p}}, \boldsymbol{p}) < \omega(\hat{\boldsymbol{p}}, \hat{\boldsymbol{p}})$. Vice versa when $p > \hat{p}$. Our suspicion is that these two effects cancel each other out, if the algorithm that establishes trust is not biased towards overly trusting or being suspicious on average. We call this property *Stability of Correctness*, and prove it absolutely holds for WMV.

A better procedure to obtain trust returns values closer to the trustworthiness values, with little variance, meaning what is believed about the sources is close to the ground truth. The quality of the procedure does not affect the Stability of Correctness at all when it is unbiased, which may initially seem counter-intuitive. However, another property captures the idea that even when it's unbiased, poor trust values still result in worse performance of WMV, *Stability of Optimality*. We prove Stability of Optimality does not hold absolutely, but that drop in the performance is bounded.

Beforehand, we need to formally define what we mean by an algorithm or procedure to establish trust values, and by it being unbiased (on average).

5.1 Parameter Distributions

We introduce random variables for our parameters. For trustworthiness: P_i is a random variable with outcome $p_i \in [0, 1]$, and $\boldsymbol{P}$ is a joint random variable with outcome $\boldsymbol{p} = (p_1, \ldots, p_n)$. Similarly, for trust: $\hat{P}_i$ is a random variable with outcome $\hat{p}_i \in [0, 1]$, and $\hat{\boldsymbol{P}}$ is a joint random variable with outcome $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_n)$. The uncertainty of source trustworthiness may be due to lack of behavior consistency, or experience, so the sources can not provide stable-quality feedback. On the other hand, inadequate interaction with sources or inaccurate modeling by decision maker may incur uncertain trust estimation.

Weighted Majority Voting requires a weight for each source which is determined by $\hat{p}_i$ (the outcome of $\hat{P}_i$). Practical usage of WMV, therefore, must have some algorithms to arrive at values for $\hat{\boldsymbol{p}}$. Depending on the quality of the algorithm, there is a degree of correlation between trust and trustworthiness: $\hat{\boldsymbol{P}}$ and $\boldsymbol{P}$. We consider the procedure to get the trust values $\hat{\boldsymbol{p}}$ as *unbiased* when the expectation of trustworthiness equals the trust value: $\mathbb{E}(\boldsymbol{P}) = \hat{\boldsymbol{p}}$. Hence, if an unbiased trust value $\hat{p}_i$ is 0.7, then the trustworthiness P_i can sometimes be greater or smaller than 0.7. Note that this is a reasonable assumption for various machine learning-based procedures or Bayesian learning in particular. In reality, we cannot guarantee that any machine learning method is completely free of such bias, but the unbiased case is interesting to study, and we expect any residual bias to be fairly small, if the algorithm is configured using sufficient empirical data.

We extend our definition of ω to accept random variables as parameters. In that case, the output of ω is a distribution over accuracy. The expectation of such decision accuracy is:

$$\mathbb{E}(\omega(\hat{\boldsymbol{P}}, \boldsymbol{P})) = \sum_{\hat{\boldsymbol{p}}, \boldsymbol{p}} \mathbb{P}(\hat{\boldsymbol{P}} = \hat{\boldsymbol{p}}, \boldsymbol{P} = \boldsymbol{p}) \omega(\hat{\boldsymbol{p}}, \boldsymbol{p}) \tag{4}$$

Besides, there can be an ideal situation where "magically" the decision maker knows the actual trustworthiness variable (i.e., $\boldsymbol{P}$), and can use it to make decisions. The expected probability of making correct decision is,

$$\mathbb{E}(\omega(\boldsymbol{P}, \boldsymbol{P})) = \sum_{\boldsymbol{p}} \mathbb{P}(\boldsymbol{P} = \boldsymbol{p}) \omega(\boldsymbol{p}, \boldsymbol{p}) \tag{5}$$

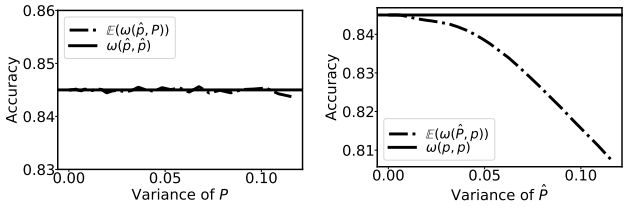
5.2 Stability of Correctness

In this section, we do not care about what the distribution of $\boldsymbol{P}$ actually looks like, as long as $\mathbb{E}(\boldsymbol{P}) = \hat{\boldsymbol{p}}$, meaning the trust values used for decision making are unbiased. The main result is that in this case, the decision accuracy that $\mathcal{D}_W$ is believed to achieve by the decision maker, equals the probability that the decision is actually correct. This is an important positive result, that supports the idea of using WMV in practice. Decision makers are not delusional about the correctness of their decisions. Formally, we define the property of Stability of Correctness (SoC) as:

Theorem 1. *Stability of Correctness (SoC): For WMV, if* $\hat{\boldsymbol{p}} = \mathbb{E}(\boldsymbol{P})$, *then* $\mathbb{E}(\omega(\hat{\boldsymbol{p}}, \boldsymbol{P})) - \omega(\hat{\boldsymbol{p}}, \hat{\boldsymbol{p}}) = 0$.

Sketch of Proof. It follows from the fact in Section 4.2 that the indicator vector set $\mathbb{D}_W(\hat{\boldsymbol{p}})$ where decisions are supposed to be correct remains unchanged, when $\hat{\boldsymbol{p}}$ is unchanged. Also consider the fact that each P_i is independently distributed. □

We show the results of two Monte Carlo simulations with 100, 000 runs over Example 2 to demonstrate the effect of distribution variance on the expected correctness of an unbiased estimate. In Figure 5(a), we depict $\omega(\hat{\boldsymbol{p}}, \hat{\boldsymbol{p}})$ and $\mathbb{E}(\omega(\hat{\boldsymbol{p}}, \boldsymbol{P}))$, where trustworthiness $\boldsymbol{P}$ is a Beta distribution with expected value $\hat{\boldsymbol{p}}$ equal to trust (unbiased) and a variance set by the x-axis. This figure shows the variance of the trustworthiness $\boldsymbol{P}$ has no effect on the correctness on average, which confirms our theorem. In contrast, in Figure 5(b), we depict $\omega(\boldsymbol{p}, \boldsymbol{p})$ and $\mathbb{E}(\omega(\hat{\boldsymbol{P}}, \boldsymbol{p}))$, letting the trust be the quantity being a random variable, distributed around trustworthiness with increasing variance. Unsurprisingly, this figure shows that a more divergent trust distribution leads to lower average correctness $\mathbb{E}(\omega(\hat{\boldsymbol{P}}, \boldsymbol{p}))$ since the trust is more likely to be far away from trustworthiness and results in accuracy degradation. Furthermore, $\mathbb{E}(\omega(\hat{\boldsymbol{P}}, \boldsymbol{p}))$ can never exceed $\omega(\boldsymbol{p}, \boldsymbol{p})$, in line with the conclusion of section 4.3. In the next section, we will study this case further.



(a) $\hat{\boldsymbol{p}}$ fixed, $\boldsymbol{P}$ probabilistic (b) $\boldsymbol{p}$ fixed, $\hat{\boldsymbol{P}}$ probabilistic

Figure 5: Effect of variance on Stability of Correctness

5.3 Stability of Optimality

Although for Stability of Correctness the shape (and variance) of the trustworthiness distribution was irrelevant, intuitively a distribution with less variance should still be better for the decision maker. We introduce another stability property in this section to capture this idea: *Stability of Optimality* (SoO). Formally, it means whether decisions made with the trustworthiness variables revealed are as good as those made only with the trust values available. We formally capture this gap with the definition below:

$$SoO(\boldsymbol{P}) = \mathbb{E}(\omega(\boldsymbol{P}, \boldsymbol{P})) - \mathbb{E}(\omega(\hat{\boldsymbol{p}}, \boldsymbol{P})) \tag{6}$$

In other words, it also measures compared with using trust to decide, how much the decision accuracy can be improved, if trustworthiness values are available. Note that an equivalent (via Theorem 1) formulation is: $\mathbb{E}(\omega(\boldsymbol{P}, \boldsymbol{P})) - \omega(\hat{\boldsymbol{p}}, \hat{\boldsymbol{p}})$, when $\hat{\boldsymbol{p}} = \mathbb{E}(\boldsymbol{P})$.

To analyze Stability of Optimality formally, we introduce some definitions. Assume the trustworthiness p_i is bounded in some range, e.g. $\forall i, a_i \leq p_i \leq b_i$. Denote the value space of $\boldsymbol{p}$ as Hypercube $\mathbb{H}$, $\boldsymbol{p} \in \mathbb{H}$. The set of vertexes of the Hypercube is denoted as Vertex Space $\mathbb{Q}$, where each vertex $\boldsymbol{q} \in \mathbb{Q}$ and q_i is either a_i or b_i. Defined in the hypercube, the distribution of $\boldsymbol{P}$ with expectation $\hat{\boldsymbol{p}}$ can be arbitrary. We name an *extreme distribution* for random variables $\boldsymbol{P}$ in the vertex space $\mathbb{Q}$ of the hypercube, where $\mathbb{P}(P_i = a_i) = \frac{b_i - \hat{p}_i}{b_i - a_i}$, $\mathbb{P}(P_i = b_i) = \frac{\hat{p}_i - a_i}{b_i - a_i}$.

When trustworthiness is revealed for decision-making, we observe that a high variance in trustworthiness is *good* for accuracy, especially the extreme distribution. That is, when a source is more trustworthy than the average, increasing its weight enhances overall decision accuracy. Conversely, when a source is less trustworthy than the average, it can degrade decision quality to some extent, but the impact is mitigated by reducing the weight of this source. In other words, it's better to have a 50% chance for a source with $P = 0.9$ and 50% for $P = 0.5$, than a source with $p = 0.7$. Formally,

Lemma 4. *Take random variables $\boldsymbol{P}$ defined in a Hypercube with $\mathbb{E}(\boldsymbol{P}) = \hat{\boldsymbol{p}}$. The correctness of $\mathbb{E}(\omega(\boldsymbol{P}, \boldsymbol{P}))$ is bounded by the extreme distribution:*

$$\mathbb{E}(\omega(\boldsymbol{P}, \boldsymbol{P})) \leq \sum_{\boldsymbol{q} \in \mathbb{Q}} \left(\omega(\boldsymbol{q}, \boldsymbol{q}) \prod_{i=1}^{n} \mathbb{P}(P_i = q_i) \right) \tag{7}$$

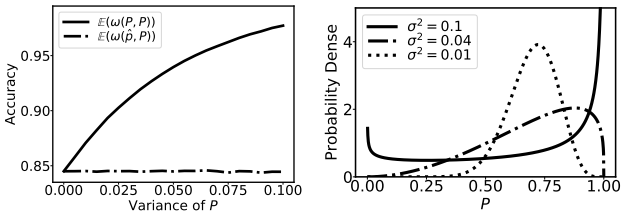
Sketch of Proof. Per Lemma 1, $\omega(\boldsymbol{p}, \boldsymbol{p})$ is convex in one dimension, and the extreme distribution maximizes $\mathbb{E}(\omega(\boldsymbol{P}, \boldsymbol{P}))$ in that dimension. The Lemma follows by independence of the trustworthiness variables. □

Lemma 4 demonstrates that the decision accuracy is bounded (not always 100%), even in the ideal situation where trustworthiness is given, and it is determined by the distribution of trustworthiness. This is intuitive as more trustworthy sources should lead to better decisions. Further, if trustworthiness is a constant rather than a random variable, Lemma 4 still holds. That is:

Corollary 1. *For any point $\boldsymbol{p}$ in the Hypercube, $\omega(\boldsymbol{p}, \boldsymbol{p})$ is bounded by a linear combination of the correctness of the vertexes of the hypercube.*

$$\omega(\boldsymbol{p}, \boldsymbol{p}) \leq \frac{1}{\prod_{i=1}^{n}(b_i - a_i)} \sum_{\boldsymbol{q} \in \mathbb{Q}} \omega(\boldsymbol{q}, \boldsymbol{q}) \prod_{i:q_i = a_i} (b_i - p_i) \prod_{i:q_i = b_i} (p_i - a_i) \tag{8}$$

Proof. Let $\mathbb{P}(\boldsymbol{P} = \boldsymbol{p}) = 1$ in Lemma 4. □



(a) $\hat{p}$ fixed, P probabilistic (b) Example of Beta Distribution

Figure 6: Effect of Variance on Stability of Optimality

Stability of Optimality does not strictly hold, as the gap (Equation 6) is typically non-zero. We prove upper bounds on the gap, which goes to 0 as the distribution of trustworthiness becomes tighter. Let δ quantify the size of the support of the distribution.

Theorem 2. *Stability of Optimality: If $\hat{\boldsymbol{p}} = \mathbb{E}(\boldsymbol{P})$ and all P_i have support $[\hat{p}_i - \delta_i, \hat{p}_i + \delta_i]$, then*

$$SoO(\boldsymbol{P}) \leq (1 - \omega(\hat{\boldsymbol{p}}, \hat{\boldsymbol{p}})) \cdot \left(1 - \prod_{i=1}^{n} \left(1 - \frac{1}{2} \cdot \frac{\delta_i}{1 - \hat{p}_i} \right) \right) \tag{9}$$

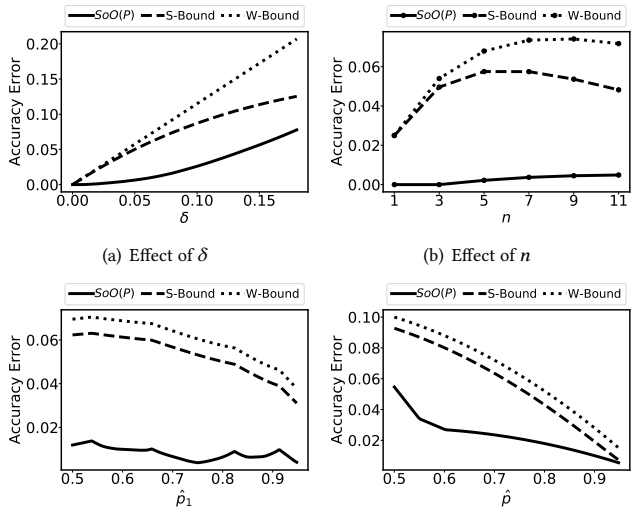
A weaker but more intuitive bound is also derived using the Bernoulli Inequality,

$$SoO(\boldsymbol{P}) \leq \frac{1-\omega(\hat{\boldsymbol{p}},\hat{\boldsymbol{p}})}{2}\sum_{i=1}^{n}\frac{\delta_i}{1-\hat{p}_i} \tag{10}$$

Sketch of Proof. Via Lemma 4, we know the extreme distribution that maximizes $\mathbb{E}(\omega(\boldsymbol{P},\boldsymbol{P}))$, relying on the correctness of the vertexes. Via Corollary 1, the upper bounds for the correctness of vertexes can be obtained, only relying on $\omega(\hat{\boldsymbol{p}},\hat{\boldsymbol{p}})$. With some algebra, both bounds (9) and (10) can be obtained. □

While there is a gap between making decisions based on unbiased trust and based on trustworthiness, Theorem 2 proves that this gap is bounded by a relatively small threshold, implying that the unbiased trust would not reduce the decision quality too much. The upper bound is influenced by the distribution of trustworthiness, and converges towards zero with that variance reducing.

To illustrate the effect of distribution variance on $SoO(\boldsymbol{P})$, we provide a Monte Carlo simulation with 100,000 runs over Example 2. In Figure 6(a), we measure $\mathbb{E}(\omega(\boldsymbol{P},\boldsymbol{P}))$ and $\mathbb{E}(\omega(\hat{\boldsymbol{p}},\boldsymbol{P}))$, where $\hat{\boldsymbol{p}}$ is constant, and $\boldsymbol{P}$ follows Beta distribution with increasing variance. It presents that the larger the variance is, the larger $SoO(\boldsymbol{P})$ is, which validates the result of Lemma 4. To put the quantity of the variance in context, we provide examples of trustworthiness being Beta distributions with a certain variance in Figure 6(b).



(a) Effect of δ (b) Effect of n

(c) Effect of single $\hat{p}_1$ (d) Effect of identical $\hat{\boldsymbol{p}}$

Figure 7: Parameter Analysis on Stability of Optimality

In Figure 7, we provide a parameter analysis with numerical experiments to demonstrate how they influence $SoO(\boldsymbol{P})$ and the bounds. Example 2 is used in Figures 7(a) and 7(c); Figures 7(b) and 7(d) needed some adaptation, where the sources have identical $\hat{p} = 0.7$. And $\delta = 0.05$ is the default for all the sources.

Figure 7(a) represents that with δ decreasing, $SoO(\boldsymbol{P})$ and its bounds also decrease to zero. There is a linear bound on the effect of δ (via the weaker bound). This implies that the uncertainty level of sources plays a significant role in determining the Stability of Optimality. In Figure 7(b), with the number of identical sources n increasing ($\hat{p} = 0.7$), $SoO(\boldsymbol{P})$ and the bounds change little, which implies source number perhaps influences little on the accuracy gap $SoO(\boldsymbol{P})$ (i.e., the gap in accuracy between decision making using unbiased trust and using trustworthiness). This means that the number of sources may barely influence how valuable it is to know the sources' combined trustworthiness.

In Figure 7(c), only $\hat{p}_1$ is variable and it shows that $SoO(\boldsymbol{P})$ always remains low level and it is a piecewise function with local maximization. As studied in Section 4.1, it becomes evident that the local maximization results from the piece-wise nature of the trustworthiness effect on decision accuracy. In Figure 7(d), where all $\hat{p}$ are equal and increase to 1 simultaneously, $SoO(\boldsymbol{P})$ almost decreases to 0. It makes sense because with the trustworthiness increasing, the decision accuracy increases more slowly due to the concavity of WMV with identical sources (See Lemma 2).

To conclude, the gap of Stability of Optimality $SoO(\boldsymbol{P})$ is somewhat sensitive to parameter δ, which depicts the range and variance of sources trustworthiness, but not sensitive to the number of sources and other parameters. Overall, the optimality of WMV has a high degree of stability, meaning $SoO(\boldsymbol{P})$ tends to be close to 0.

6 CONCLUSION AND FUTURE WORK

The common dependence on an estimate or trust of source trustworthiness brings out the need to analyze whether WMV is stable, meaning having tolerant decision inaccuracy with the difference between trust and trustworthiness bounded.

We first analyze how sensitive WMV is to the changes in trust and trustworthiness. We find that small deviation between trust and trustworthiness does not affect accuracy, and also underestimation usually harms less than overestimation. We then introduced two statistical properties of WMV, *Stability of Correctness* and *Stability of Optimality*. Assuming that on average the estimation procedure has no bias towards over or underestimating, we proved that *Stability of Correctness* holds absolutely, regardless of which estimation procedure is used or how well it estimates. This guarantees that relying on an unbiased estimate of source trustworthiness is safe, which is also common in practice. However, the amount of inefficiency introduced by relying on an estimate instead of the trustworthiness itself is limited, as we prove a linear bound on *Stability of Optimality*. The proposed formal framework and the two types of stability properties can be generalized to analyze other types of decision mechanisms or scenarios (e.g., where sources are dependent).

For future work, beyond the bounded assumption, it's valuable to explore a more precise characterization of the impact of the trustworthiness distribution on SoO in the unbiased setting. Besides, it is also worth studying the stability of WMV in a more general case, namely when trust is a biased estimate of trustworthiness. Some researchers have found that although some sources are assigned weights, they have no influence on the decision result [1, 6]. In other words, we may distribute more estimate error on such sources.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (NSFC) under Grant 62106223 and (NSFC) Grant 62293511.

REFERENCES

[1] Tahar Allouche, Bruno Escoffier, Stefano Moretti, and Meltem Öztürk. 2021. Social ranking manipulability for the cp-majority, Banzhaf and lexicographic excellence solutions. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 17–23.

[2] Jamal Bentahar, Nagat Drawel, and Abdeladim Sadiki. 2022. Quantitative group trust: A two-stage verification approach. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 100–108.

[3] Daniel Berend and Aryeh Kontorovich. 2013. A sharp estimate of the binomial mean absolute deviation with applications. *Statistics & Probability Letters* 83, 4 (2013), 1254–1259. https://doi.org/10.1016/j.spl.2013.01.023

[4] Daniel Berend and Aryeh Kontorovich. 2015. A finite sample analysis of the Naive Bayes classifier. *J. Mach. Learn. Res.* 16, 1 (2015), 1519–1545.

[5] Philip J. Boland. 1989. Majority Systems and the Condorcet Jury Theorem. *Journal of the Royal Statistical Society: Series D (The Statistician)* 38, 3 (1989), 181–189.

[6] Larry Bowen. 2009. Weighted voting systems.

[7] Javier Carbo and Jose M Molina. 2023. Promoting cooperation of agents through aggregation of services in trust models. *Knowledge-Based Systems* 277 (2023), 110804.

[8] marquis de Condorcet, Jean-Antoine-Nicolas de Caritat. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie royale. 1743–1794 pages.

[9] Alexander Philip Dawid and Allan M Skene. 1979. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28, 1 (1979), 20–28.

[10] Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-based trust: estimating the trustworthiness of web sources. *Proceedings of the VLDB Endowment* 8, 9 (2015), 938–949.

[11] Nagat Drawel, Jamal Bentahar, Amine Laarej, and Gaith Rjoub. 2022. Formal verification of group and propagated trust in multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 36, 1 (2022), 19.

[12] Nagat Drawel, Hongyang Qu, Jamal Bentahar, and Elhadi Shakshuki. 2020. Specification and automatic verification of trust-based multi-agent systems. *Future Generation Computer Systems* 107 (2020), 1047–1060.

[13] David A Freedman. 1963. On the asymptotic behavior of Bayes' estimates in the discrete case. *The Annals of Mathematical Statistics* 34, 4 (1963), 1386–1403.

[14] Chao Gao, Yu Lu, and Dengyong Zhou. 2016. Exact exponent in optimal rates for crowdsourcing. In *International Conference on Machine Learning*. PMLR, 603–611.

[15] Yan Ge, Jun Ma, Li Zhang, Xiang Li, and Haiping Lu. 2023. Trustworthiness-aware knowledge graph representation for recommendation. *Knowledge-Based Systems* 278 (2023), 110865.

[16] Pascal Germain, Alexandre Lacasse, Francois Laviolette, Mario March, and Jean-Francis Roy. 2015. Risk Bounds for the Majority Vote: From a PAC-Bayesian Analysis to a Learning Algorithm. *Journal of Machine Learning Research* 16, 26 (2015), 787–860.

[17] Bernard Grofman, Guillermo Owen, and Scott L Feld. 1983. Thirteen theorems in search of the truth. *Theory and decision* 15, 3 (1983), 261–278.

[18] Melody Guan, Varun Gulshan, Andrew Dai, and Geoffrey Hinton. 2018. Who Said What: Modeling Individual Labelers Improves Classification. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (Apr. 2018).

[19] Zhaori Guo. 2023. Multi-Advisor Dynamic Decision Making. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 2949–2951.

[20] JQ James. 2020. Sybil attack identification for crowdsourced navigation: A self-supervised deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* 22, 7 (2020), 4622–4634.

[21] James Kotary, Vincenzo Di Vito, and Ferdinando Fioretto. 2023. Differentiable model selection for ensemble learning. In *Proceedings of the Fifteen International Joint Conference on Artificial Intelligence, IJCAI-23*.

[22] Alexandre Lacasse, François Laviolette, Mario Marchand, Pascal Germain, and Nicolas Usunier. 2006. PAC-Bayes bounds for the risk of the majority vote and the variance of the Gibbs classifier. In *NIPS*. 769–776.

[23] Hongwei Li and Bin Yu. 2014. Error rate bounds and iterative weighted majority voting for crowdsourcing. *arXiv preprint arXiv:1411.4086* (2014).

[24] N. Littlestone and M.K. Warmuth. 1994. The Weighted Majority Algorithm. *Information and Computation* 108, 2 (1994), 212–261.

[25] Yuan Luo. 2023. Incentivizing Sequential Crowdsourcing Systems. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 2697–2699.

[26] Edoardo Manino, Long Tran-Thanh, and Nicholas Jennings. 2019. Streaming Bayesian inference for crowdsourced classification. *Advances in Neural Information Processing Systems* 32 (2019), 12782–12792.

[27] Edoardo Manino, Long Tran-Thanh, and Nicholas R Jennings. 2019. On the efficiency of data collection for multiple Naïve Bayes classifiers. *Artificial Intelligence* 275 (2019), 356–378.

[28] Irene Martín-Morató and Annamaria Mesaros. 2023. Strong labeling of sound events using crowdsourced weak labels and annotator competence estimation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 31 (2023), 902–914.

[29] Lucas Maystre, Nagarjuna Kumarappan, Judith Bütepage, and Mounia Lalmas. 2021. Collaborative Classification from Noisy Labels. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1639–1647.

[30] Alessio Mazzetto, Dylan Sam, Andrew Park, Eli Upfal, and Stephen Bach. 2021. Semi-supervised aggregation of dependent weak supervision sources with performance guarantees. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 3196–3204.

[31] Reshef Meir, Ofra Amir, Omer Ben-Porat, Tsviel Ben Shabat, Gal Cohensius, and Lirong Xia. 2023. Frustratingly easy truth discovery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 6074–6083.

[32] Bassam Moslem, Mohamad Diab, Mohamad Khalil, and Catherine Marque. 2012. Combining data fusion with multiresolution analysis for improving the classification accuracy of uterine EMG signals. *EURASIP Journal on Advances in Signal Processing* 2012, 1 (2012), 1–9.

[33] Tim Muller, Dongxia Wang, and Jun Sun. 2020. Provably Robust Decisions based on Potentially Malicious Sources of Information. In *2020 IEEE 33rd Computer Security Foundations Symposium (CSF)*. IEEE, 411–424.

[34] Shmuel Nitzan and Jacob Paroush. 1982. Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review* (1982), 289–297.

[35] Vikas C Raykar, Shipeng Yu, Linda H Zhao, Gerardo Hermosillo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. 2010. Learning from crowds. *Journal of Machine Learning Research* 11, 4 (2010).

[36] Theodoros Rekatsinas, Manas Joglekar, Hector Garcia-Molina, Aditya Parameswaran, and Christopher Ré. 2017. Slimfast: Guaranteed results for data fusion and source reliability. In *Proceedings of the 2017 ACM International Conference on Management of Data*. 1399–1414.

[37] Noel Sardana, Robin Cohen, Jie Zhang, and Shuo Chen. 2018. A Bayesian multiagent trust model for social networks. *IEEE Transactions on Computational Social Systems* 5, 4 (2018), 995–1008.

[38] Amartya Sen. 1977. Social choice theory: A re-examination. *Econometrica: journal of the Econometric Society* (1977), 53–89.

[39] Pankaj Telang, Munindar P Singh, and Neil Yorke-Smith. 2023. Maintenance commitments: Conception, semantics, and coherence. *Artificial Intelligence* 324 (2023), 103993.

[40] Zhijun Tong and Richard Y Kain. 1991. Vote assignments in weighted voting mechanisms. *IEEE Trans. Comput.* 40, 05 (1991), 664–667.

[41] Elizabeth Walter. 2008. *Cambridge advanced learner's dictionary*. Cambridge university press.

[42] Gongqing Wu, Xingrui Zhuo, Xianyu Bao, Xuegang Hu, Richang Hong, and Xindong Wu. 2023. Crowdsourcing Truth Inference via Reliability-Driven Multi-View Graph Embedding. *ACM Transactions on Knowledge Discovery from Data* 17, 5 (2023), 1–26.

[43] Yihong Wu and Pengkun Yang. 2016. Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Transactions on Information Theory* 62, 6 (2016), 3702–3720.

[44] Yi-Shan Wu, Andres Masegosa, Stephan Lorenzen, Christian Igel, and Yevgeny Seldin. 2021. Chebyshev-Cantelli PAC-Bayes-Bennett Inequality for the Weighted Majority Vote. *Advances in Neural Information Processing Systems* 34 (2021).

[45] Yi-Shan Wu, Andres Masegosa, Stephan Lorenzen, Christian Igel, and Yevgeny Seldin. 2021. Chebyshev-Cantelli PAC-Bayes-Bennett inequality for the weighted majority vote. *Advances in Neural Information Processing Systems* 34 (2021), 12625–12636.

[46] Bin Yu, Munindar P Singh, and Katia Sycara. 2004. Developing trust in large-scale peer-to-peer systems. In *IEEE First Symposium onMulti-Agent Security and Survivability, 2004.* IEEE, 1–10.

[47] Leonit Zeynalvand, Tie Luo, Ewa Andrejczuk, Dusit Niyato, Sin G. Teo, and Jie Zhang. 2021. A Blockchain-Enabled Quantitative Approach to Trust and Reputation Management with Sparse Evidence. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems* (Virtual Event, United Kingdom) *(AAMAS '21)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1707–1708.

[48] Yuchen Zhang, Xi Chen, Dengyong Zhou, and Michael I Jordan. 2016. Spectral methods meet EM: A provably optimal algorithm for crowdsourcing. *The Journal of Machine Learning Research* 17, 1 (2016), 3537–3580.