ABSTRACT

WANG, YONGHONG. Evidence-Based Trust in Distributed Agent Systems. (Under the direction of Professor Munindar P. Singh).

Trust is a crucial basis for interactions among parties in large, open distributed systems. Yet, the scale and dynamism of such systems make it infeasible for each party to have a direct basis for trusting another party. For this reason, the participants in an open system must share information about trust. Traditional models of trust employ simple heuristics and ad hoc formulas, without adequate mathematical justification. These models fail to properly address the challenges of combining trust from conflicting sources, dealing with malicious agents, and updating trust.

This dissertation understands an agent Alice's trust in an agent Bob in terms of Alice's certainty in her belief that Bob is trustworthy. Unlike previous approaches, this dissertation formulates certainty in terms of a statistical measure defined over a probability distribution of the probability of positive outcomes.

Specifically this dissertation makes the following contributions. It

- Develops a mathematically well-formulated approach for an evidence-based account of trust; proves desirable properties of certainty; and establishes a bijection between evidence and trust.
- Defines a concatenation, an aggregation, and a selection operator to propagate trust, and proves desirable properties of these operators.
- Develops trust update mechanisms and formally analyzes their properties.
- Extends the definition of certainty from binary events to multivalued events. Establishes a bijection between Dempster-Shafer belief space and evidence space, and defines a novel combination operator, which is commutative and associative. In contrast with traditional combination operators, which ignore conflict and sometimes yield counterintuitive results, the proposed operator treats conflict naturally.

Evidence-Based Trust in Distributed Agent Systems

by
Yonghong Wang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina

2009

APPROVED BY:

Dr. Ting Yu	Dr. Greg Byrd
Dr. Munindar Singh Chair of Advisory Committee	Dr. Dennis Bahler

DEDICATION

To my family

BIOGRAPHY

Yonghong Wang was born in a small town in Anhui, China. He received his Bachelor of Engineering degree in Automation, with a minor in Mathematics in 1994 and his Master of Science degree in Mathematics in 1996 from Tsinghua University. He received his Ph.D. degree in Mathematics from North Carolina State University in 2000. He enrolled as a Ph.D. student in the department of Computer Science at North Carolina State University in 2001.

ACKNOWLEDGMENTS

I thank my advisor Dr. Munindar P. Singh for his guidance and help. It was a great privilege for me to study under his guidance. He always understood all the details of my thoughts quickly, then encouraged and directed me to work in the right direction. He has taught me how to survey the literature, to think, to criticize, to write, and to publish. I wish I could stay longer and learn more from him.

I thank my committee members, Dr. Dennis Bahler, Dr. Gregory Byrd, and Dr. Ting Yu, for their useful suggestions and help. I also thank all my teachers who have enriched my knowledge and experience in computer science.

I thank my previous advisors Dr. Xiaobiao Lin and Dr. S. Felix Wu for their guidance, support, and help when I studied and worked in Mathematics Department and the Network Security Lab. Their guidance is invaluable to my learning experience.

I also want to thank all my colleagues at Service-Oriented Computing Laboratory: Dr. Amit Chopra, Chung-Wei Hang, Dr. Bin Yu, Dr. Zhengang Cheng, Dr. Nirmit Desai, Dr. Feng Wan, Dr. Ashok Mallya, Dr. Michael Maximilien, and Dr. Yathi Udupi for their encouragement, discussions, and help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Trust and Autonomous Agent Systems

Trust is a broad concept with many connotations. We concentrate on trust as it relates to beliefs and not, for example, to emotions. We study trust that involves settings wherein independent (i.e., autonomous and adaptive) parties interact with one another, and each party may choose with whom to interact based on how much trust it places in the other. Examples of such applications are social networks and webs of information sources. We can cast each party as providing and seeking services, and the problem as one of service selection in a distributed environment.

The currently dominant computer science approaches for trust emphasize identity and generally take a qualitative stance in determining if a party is to be deemed trustworthy or not. Let us briefly consider the pros and cons of the existing approaches in broad terms. (Chapter 7 discusses the relevant literature in some detail.)

- *Identity.* Traditional approaches address trust primarily with respect to identity. A party attempts to establish its trustworthiness to another party by presenting a certificate. The certificate is typically obtained from a certificate authority or (as in PKI, the Public Key Infrastructure) from another party. The presumption is that the certificate issuer would have performed some offline verification. The best case for such an engagement is that a party truly has the identity that it professes to have. Establishing identity is, of course, central to enabling trust.

  However, identity by itself is inadequate for the problems we discuss here. In particu-

lar, identity does not yield a basis for determining if a given party will serve a desired purpose appropriately. For example, if Amazon presents a valid certificate obtained from Verisign, the most it means is that the presenter of the certificate is indeed Amazon. The certificate does not mean that Alice would have a pleasant shopping experience at Amazon. After all, Verisign's certificate is not based upon any relevant experience: simply put, the certificate does not mean that Verisign purchased goods from Amazon and had a pleasant experience. From the traditional standpoint, this example might sound outlandish, but ultimately if trust is to mean that one party can place its plans in the hands of another, the expected experience is no less relevant than the identity of the provider.

- *All or none.* Traditional approaches model trust qualitatively. This is based on an intuition of hard security. If one cannot definitely determine that a particular party has the stated identity, then that is sufficient reason not to deal with it at all.

  Yet in many cases, requiring an all-or-none decision about trust can be too much to ask for, especially when we think not of identity but more broadly of whether a given party would support one's plans. When we factor in the complexity of the real world and the task to be performed, virtually no one would be able to make a hard guarantee about success. Following the above example, it would be impossible for Bob to guarantee that he will get Alice to the airport on time, recommend only the perfect movies, or introduce her to none other than her potential soul mate.

  Approaches based on reputation management seek to address this challenge. They usually accommodate shades of trust numerically based on ratings acquired from users. However, these approaches are typically formulated in a heuristic, somewhat ad hoc manner. The meaning assigned to the aggregated ratings is not clear from a probabilistic (or some other mathematical) standpoint.

A key intuition about trust as it is applied in the autonomous and adaptive settings is that it reflects the trusting party's belief that the trusted party will support its plans [Castelfranchi and Falcone, 1998]. For example, if Alice trusts Bob to get her to the airport, then this means that Alice is putting part of her plans in Bob's hands. In other words, Alice believes that there will be a good outcome from Bob providing her with the specific service. In a social setting, a similar question would be whether Alice trusts Bob to give her a

recommendation to a movie that she will enjoy watching, or whether Alice trusts Bob to introduce her to a new friend, Charlie, with whom she will have pleasant interactions. In scientific computing on the cloud, Alice may trust a service provider such as Dell that she will receive adequate compute resources.

We take the view that a quantitative account of trust that considers the interactions among parties is crucial for supporting the above kinds of applications. For rational agents, trust in a party should be based substantially on evidence consisting of positive and negative experiences with it. This evidence can be collected by an agent locally or via a reputation agency [Maximilien, 2004] or by following a referral protocol [Sen and Sajja, 2002]. In such cases, the evidence may be implicit in the trust reports obtained that somehow summarize the evidence being shared.

A common way to estimate trustworthiness is to evaluate the probability of the number of good services provided in the past. However, a traditional probability representation cannot distinguish between getting one good service from two interactions, and getting 100 good services from 200 interactions. As a result, modern trust models define trust in terms of both probability and certainty [Gómez et al., 2007, Harbers et al., 2007, Zacharia and Maes, 2000, Teacy et al., 2006]. The certainty is a measure of the confidence that an agent may place in the trust information. Computing the certainty can help an agent filter out those for whom it has insufficient information even if nominally the probability of a good outcome is high. In general, the certainty of a trust value should (a) increase as the amount of information increases with a fixed probability, and (b) decrease as the number of conflicts increases with a fixed total number of experiences.

We develop a mathematically well-formulated evidence-based approach for trust that supports the following four crucial requirements, which arise in multiagent systems applied in important settings such as electronic commerce or information fusion.

Evidence-based. How to model trust based on evidence so that it increases as the amount of information increases with a fixed probability, and decreases as the number of conflicts increases with a fixed total number of experiences.

Dynamism. Practical agent systems face the challenge that trust evolves over time. This may happen because additional information is obtained, the parties alter their behavior, or the needs of the rating party change. An update mechanism is required to update the trust assessments based on new information.

Composition. It is clear that trust cannot be trivially propagated. For example, Alice may trust Bob who trusts Charlie, but Alice may not trust Charlie. However, as a practical matter, a party would not have the opportunity or be able to expend the cost, e.g., in money or time, to obtain direct experiences with every other party. This is the reason that a multiagent approach—wherein agents exchange trust reports—is plausible. Consequently, we need a way to combine trust reports that cannot themselves be perfectly trusted, possibly because of the source of such reports or the way in which they are obtained. And we need to accommodate the requirement that trust is weakened when propagated through such chains.

Conflict. Trust reports can't be trivially combined either. For example, Bob and Alice may have totally contradictory opinions about the services that Dell provides. So trust combination should handle conflict of trust from different sources.

1.2 Combination Rules and Dempster-Shafer Theory

Combining information from different sources can be difficult, especially when the sources are not dependable and the information they provide is conflicting. Dempster-Shafer belief functions have been used in many works to represent and manipulate uncertainty. Many combination rules for Dempster-Shafer theory have been proposed to combine information from different sources.

The traditional DS fusion rule completely ignores conflict. Whereas it is associative and commutative, it is not continuous. Further, it yields counterintuitive results in the face of conflicting information. Zadeh [Zadeh, 1986] gives a well-known example. Suppose a patient, Bob, is diagnosed by two doctors, Alice and Peter. Alice believes that Bob has either meningitis with probability 0.99, or a tumor with probability 0.01. Peter believes that Bob has either concussion with probability 0.99, or a tumor with probability 0.01. According to the traditional DS rule, combining Alice and Peter's beliefs yields that a tumor is certain. This is unintuitive because neither physician considers a tumor likely.

1.3 Previous Work

1.3.1 Trust and Autonomous Agent System

In electronic markets, due to the high uncertainty about quality and reliability of the products and services offered by others, it is crucial for agents to compute the trustworthiness of the other agents before initiating any service request. Similar considerations apply in Web-based information systems, in general: agents must be able to compute how much trust to place in others with whom they might have had no prior experience.

The mechanisms that support finding trust estimations are called reputation systems. There are two kinds of reputation systems: centralized and distributed. Centralized reputation systems include collaborative filtering (social information filtering) systems [Breese et al., 1998],
[Resnick et al., 1994] and [Dellarocas, 2004], and online reputation systems [Dellarocas, 2004]. Distributed reputation systems include peer-to-peer (P2P) systems [Aberer and Despotovic, 2001, Xiong and Liu, 2004] and referral systems [Yu and Singh, 2002, 2003a].

Collaborative filtering approaches select resources based on the relationships between agents as modeled by the similarity or dissimilarity between their subjective judgments. For example, GroupLens [Resnick et al., 1994] helps users find netnews articles based on how their past ratings correlate with ratings given by other users. Online reputation systems [Dellarocas, 2004] build the reputation for an object by aggregating beliefs from agents about a specified agent. One of the most successful commercial examples is eBay. After each transaction, buyers and sellers can rate each other. A participant's reputation is the sum of the ratings he or she has received over the last six months. The centralized approaches do not scale well and they all have a single point of failure, so they are not appropriate in a distributed system.

Peer-to-peer systems and referral systems are built for distributed environments. They enable peers to share their experiences with each other. Trust between peers is captured by the neighborhood relation. Various mathematical approaches have been used to combine trust ratings in these systems. Traditional methods in combining the trust are not efficient in dealing with malicious agents who provide biased information.

Open systems are dynamic and distributed. In other words, an agent often needs to select a service provider with whom its has had no previous interaction. *Referral networks*

enable agents to collect trust information about service providers in a distributed manner [Yu and Singh, 2002]. In a referral network, an agent requests other agents, called referrers, to provide trust information about a service provider. The referrers may refer to other referrers if they have no direct experience with the service provider. In some trust models [Barber and Kim, 2001], trust propagation (or inference) is defined for agents to aggregate trust information from multiple sources (referrals and direct interactions). Also, the referrals can be viewed as services the referrers provide. In other words, the agents can estimate the referrers' trustworthiness based on the quality of the referrals they provide.

1.3.2 Trust and Uncertainty

Jøsang [Jøsang, 2001] models trust via two-dimensional belief, which is triple (b, d, u) where b, d, u are belief, disbelief and uncertainty. (Here $b + d + u = 1$, hence two independent dimensions). Jøsang also defines the evidence space $\langle r, s \rangle$. He defines certainty by $\frac{r+s}{r+s+2}$. Jøsang's approach satisfies our intuition about the effect of evidence but fails our intuition about the effect of conflict (certainty should decrease with increasing conflict provided the amount of evidence is held constant). It falsely predicts that mounting *conflicting* evidence increases certainty—and equally as much as mounting confirmatory evidence. Say Alice deals with Bob four times: in either case, her evidence would be between zero and four positive experiences. It should be uncontroversial that whereas Alice's certainty is greatest when the evidence is all in favor or all against, her certainty is least when the evidence is equally split. Section 2.2.2 shows that Jøsang, in contrast to our approach, assigns the same certainty in each case.

1.3.3 Dempster-Shafer Combination Rules

Yu and Singh [Yu and Singh, 2002] model positive, negative, or neutral evidence, and apply Dempster-Shafer theory to compute trust. Neutral experiences yield uncertainty, but conflicting positive or negative evidence does not increase uncertainty. Further, for conflicting evidence, Dempster-Shafer theory can yield unintuitive results. The following is a well-known example about the Dempster-Shafer theory, and is not specific to Yu and Singh's use of it [Sentz and Ferson, 2002b, Zadeh, 1979]. Say Pete sees two physicians, Dawn and Ed, for a headache. Dawn says Pete has meningitis, a brain tumor, or neither with

probabilities 0.79, 0.20, and 0.01, respectively. Ed says Pete has a concussion, a brain tumor, or neither with probabilities 0.79, 0.20, and 0.01, respectively. Dempster-Shafer theory yields that the probability of a brain tumor is 0.725, which is highly counterintuitive and wrong, because neither Dawn nor Ed thought that a brain tumor was likely. Section 2.2.3 shows that our model of trust yields an intuitive result in this case: the probability of a brain tumor is 0.21, which is close to each individual physician's beliefs.

The problem with the traditional Dempster rule of combination is that it completely ignores conflict.

Many researchers have proposed different combination operators. [Shafer, 1976, Yager, 1987, Inagaki, 2008, Ferson and Kreinovich, 2002]. These operators either ignore conflict or linearly combine conflict. They do not consider the body of evidence behind the beliefs and fail to explore the subtle relationship between conflict and evidence. When they combine beliefs by weighted averaging, they first assign weights to sources, that is, discounting the information based on itself, which is counterintuitive (it is a bit circular). This trust placed in an information source should not depend on the *current* information the source provides, but may be estimated from the *previous* experience with the source.

1.4 Scope of Work

We consider a distributed system of software agents who provide and require services, and who cooperate in helping each other find the most trustworthy service providers. Agents are autonomous and self-interested. Since the agents need to ensure that the service providers they select are trustworthy they help each other determine the trustworthiness of the service providers they are interested in.

1.5 Contributions

This dissertation makes the following contributions. For trust representation, it formulates certainty in terms of evidence based on a statistical measure defined over a probability distribution of the probability of positive outcomes; proves that for a fixed amount of evidence, certainty increases as conflict in the evidence decreases; proves that for a fixed level of conflict, certainty increases as the amount of evidence increases; establishes

a bijection between evidence and trust spaces, enabling robust combination of trust reports; and provides an efficient algorithm for computing this bijection.

For trust propagation, this dissertation applies a probabilistic theory of evidence to represent the trust between agents in terms of referrals and the quality of service obtained; defines three operators, *aggregation*, *concatenation*, and *selection*, for efficiently and accurately propagating trust in social networks; and proves some useful formal properties of the operators.

For trust updating, this dissertation provides three different operators to update trust; shows that one of provides accurate estimates of the trustworthiness of agents that change behavior frequently; and shows that this method captures the dynamic behavior of the agents.

Lastly, this dissertation addresses the challenges of evidence and uncertainty in a multivalued setting related to Dempster-Shafer theory. It extends certainty from binary events to multivalued events; proves important mathematical properties of certainty based on multivalued event; establishes the bijection of the multidimensional belief space and evidence space; and provides a combination operator that applies on Dempster-Shafer belief space and satisfies desired mathematical properties.

Chapter 2

Trust Representation

2.1 Modeling Certainty

The proposed approach is based on the fundamental intuition that an agent can model the behavior of another agent in probabilistic terms. Specifically, an agent can represent the probability of a positive experience with, i.e., good behavior by, another agent. This probability must lie in the real interval $[0, 1]$. The agent's trust corresponds to how strongly the agent believes that this probability is a specific value (whether large or small, we do not care). This strength of belief is also captured in probabilistic terms. To this end, we formulate a probability density function of the probability of a positive experience. Following [Jøsang, 1998], we term this a *probability-certainty density function (PCDF)*. Crucially, in our approach, unlike in Jøsang's, certainty is a statistical measure defined on a PCDF, and thus naturally accommodates both the amount of evidence and the extent of the conflict in the evidence.

2.1.1 Certainty from a PCDF

Because the cumulative probability of a probability lying within $[0, 1]$ must equal 1, all PCDFs must have the mean density of 1 over $[0, 1]$, and 0 elsewhere. Lacking additional knowledge, a PCDF would be a uniform distribution over $[0, 1]$. However, with additional knowledge, the PCDF would deviate from the uniform distribution. For example, knowing that the probability of good behavior is at least 0.5, we would obtain a distribution that is 0 over $[0, 0.5)$ and 2 over $[0.5, 1]$. Similarly, knowing that the probability of good behavior

lies in $[0.5, 0.6]$, we would obtain a distribution that is 0 over $[0, 0.5)$ and $(0.6, 1]$, and 10 over $[0.5, 0.6]$. Notice that although a cumulative probability must equal 1, a probability density can be any nonnegative real number: densities are constrained only to ensure that cumulative probabilities equal 1.

In formal terms, let $p \in [0, 1]$ represent the probability of a positive outcome. Let the distribution of p be given as a function $f : [0, 1] \mapsto [0, \infty)$ such that $\int_0^1 f(p)dp = 1$. The probability that the probability of a positive outcome lies in $[p_1, p_2]$ can be calculated by $\int_{p_1}^{p_2} f(p)dp$. The mean value of f is $\frac{\int_0^1 f(p)dp}{1-0} = 1$. As explained above, when we know nothing else, f is a uniform distribution over probabilities p. That is, $f(p) = 1$ for $p \in [0, 1]$ and 0 elsewhere. This reflects the Bayesian intuition of assuming an equiprobable prior. The uniform distribution has a certainty of 0. As additional knowledge is acquired, the probability mass shifts so that $f(p)$ is above 1 for some values of p and below 1 for other values of p.

Our key intuition is that the agent's trust corresponds to increasing deviation from the uniform distribution. Two of the most established measures for deviation are standard deviation and mean absolute deviation (MAD) [Weisstein, 2003]. MAD is more robust, because it does not involve squaring (which can increase standard deviation because of outliers or "heavy tail" distributions such as the Cauchy distribution). Absolute values can sometimes complicate the mathematics. But, in the present setting, MAD turns out to yield straightforward mathematics. In a discrete setting involving data points $x_1 \ldots x_n$ with mean $\hat{x}$, MAD is given by $\frac{1}{n}\Sigma_{i=1}^{n}|x_i - \hat{x}|$. In the present case, instead of summation we have an integral, so instead of dividing by n we divide by the size of the domain, i.e., 1. Because a PCDF has a mean value of 1, increase in some parts above 1 must yield a matching reduction below 1 elsewhere. Both increase and reduction from 1 are counted by $|f(p) - 1|$. Definition 1 scales the MAD for f by $\frac{1}{2}$ to remove this double counting; it also conveniently places certainty in the interval $[0, 1]$.

Definition 1. *The certainty based on f, c_f, is given by* $c_f = \frac{1}{2}\int_0^1 |f(p) - 1|dp$

In informal terms, certainty captures the fraction of the knowledge that we do have. (Section 7.3 compares this approach to information theory.) For motivation, consider randomly picking a ball from a bin that contains N balls colored white or black. Suppose p is the probability that the ball randomly picked is white. If we have no knowledge about how

many white balls there are in the bin, we cannot estimate p with any confidence. That is, certainty $c = 0$. If we know that exactly m balls are white, then we have perfect knowledge about the distribution. We can estimate $p = \frac{m}{N}$ with $c = 1$. However, if all we know is that at least m balls are white and at least n balls are black (thus $m + n \leq N$), then we have partial knowledge. Here $c = \frac{m+n}{N}$. The probability of drawing a white ball ranges from $\frac{m}{N}$ to $1 - \frac{n}{N}$. We have

$$f(p) = \begin{cases} 0, & [0, \frac{m}{N}) \\ \frac{N}{N-m-n} & p \in [\frac{m}{N}, 1 - \frac{n}{N}] \\ 0 & (1 - \frac{n}{N}, 1]. \end{cases}$$

Using Definition 1, we can confirm that certainty based on the function f as defined above, $c_f = \frac{m+n}{N}$:

$$\begin{aligned} c_f &= \tfrac{1}{2} \textstyle\int_0^1 |f(p) - 1| dp \\ &= \tfrac{1}{2} (\textstyle\int_0^{\frac{m}{N}} 1 \; dp + \int_{\frac{m}{N}}^{1-\frac{n}{N}} (\frac{N}{N-m-n} - 1) dp + \int_{1-\frac{n}{N}}^{1} 1 \; dp \\ &= \tfrac{1}{2} (\tfrac{m}{N} + \tfrac{N-m-n}{N} (\tfrac{N}{N-m-n} - 1) + \tfrac{n}{N}) \\ &= \tfrac{m+n}{N} \end{aligned}$$

2.1.2 Evidence Space

For simplicity, we begin by thinking of a (rating) agent's experience with a (rated) agent as a binary event: positive or negative. Evidence is conceptualized in terms of the numbers of positive and negative experiences. When an agent makes unambiguous direct observations of another, the corresponding evidence could be expressed as natural numbers (including zero). However, our motivation is to combine evidence in the context of trust. As Chapter 1 motivates, for reasons of dynamism or composition, the evidence may need to be discounted to reflect the weakening of the evidence source due to the effects of aging or the effects of imperfect trust having been placed in it. Intuitively, because of such discounting, the evidence is best understood as if there were real (i.e., not necessarily natural) numbers of experiences. Similarly, when a rating agent's observations are not clearcut positive or negative, we can capture the ratings via arbitrary nonnegative real numbers (as long as their sum is positive).

Accordingly, following [Jøsang, 2001], we model the evidence space as $E = \mathbb{R}^+ \times \mathbb{R}^+ \setminus \{\langle 0, 0\rangle\}$, a two-dimensional space of nonnegative reals whose sum is strictly positive. (Here $\mathbb{R}^+$ is the set of nonnegative reals.) The members of E are pairs $\langle r, s\rangle$ corresponding to the numbers of positive and negative experiences, respectively.

Definition 2. *Evidence space* $E = \{\langle r, s\rangle | r \geq 0, s \geq 0, t = r + s > 0\}$

Combining evidence as a result is a trivial operation: simply add up the positive evidence and add up the negative evidence.

Let x be the probability of a positive outcome. The posterior probability of evidence $\langle r, s\rangle$ is the conditional probability of x given $\langle r, s\rangle$ [Casella and Berger, 1990, p. 298].

Definition 3. *The conditional probability of* x *given* $\langle r, s\rangle$ *is*

$$\begin{aligned} f(x|\langle r, s\rangle) &= \frac{g(\langle r,s\rangle|x)f(x)}{\int_0^1 g(\langle r,s\rangle|x)f(x)dx} \\ &= \frac{x^r(1-x)^s}{\int_0^1 x^r(1-x)^s dx} \end{aligned}$$

where $g(\langle r, s\rangle|x) = \begin{pmatrix} r+s \\ r \end{pmatrix} x^r(1-x)^s$

Throughout this dissertation, r, s, and $t = r + s$ refer to positive, negative, and total evidence, respectively. The following development assumes that there is some evidence; i.e., $t > 0$.

Traditional probability theory models the event $\langle r, s\rangle$ by the pair $(p, 1-p)$, the expected probabilities of positive and negative outcomes, respectively, where $p = \frac{r+1}{r+s+2} = \frac{r+1}{t+2}$. The idea of adding 1 each to r and s (and thus 2 to $r + s$) follows Laplace's famous *rule of succession* for applying probability to inductive reasoning [Ristad, 1995]. This rule in essence reflects the assumption of an equiprobable prior, which is common in Bayesian reasoning. Before any evidence, positive and negative outcomes are equally likely, and this prior biases the evidence obtained subsequently.

In practical terms, Laplace's rule of succession, alluded to above, reduces the impact of sparse evidence. It is sometimes termed *Laplace smoothing*. If you only made one observation and it was positive, you would not want to conclude that there would never be a negative observation. As the body of evidence increases, the increment of 1 has a

negligible effect. More sophisticated formulations of rules of succession exist [Ristad, 1995], but Laplace's rule is simple and reasonably effective for our present purposes. Laplace's rule is insensitive to the number of outcomes in that 1 is always added. The effect of this statistical "correction" (the added 1) decreases inversely as the number of outcomes being considered increases. More sophisticated approaches may be thought of as decreasing the effects of their corrections more rapidly.

Importantly, as explained above, total evidence in our approach is modeled as a nonnegative real number. Due to the effect of discounting, the total evidence can appear to be lower than 1. In such a case, the effect of the Laplace smoothing can become dominant. For this reason, this dissertation differs from Wang and Singh [Wang and Singh, 2007] in defining a measure of the conflict in the evidence that is different from the probability to be inferred from the evidence.

2.1.3 Conflict in Evidence

The conflict in evidence simply refers to the relative weights of the negative and positive evidence. Conflict is highest when the negative and positive evidence are equal, and least when the evidence is unanimous one way or the other. Definition 4 characterizes the amount of *conflict* in the evidence. To this end, we define α as $\frac{r}{t}$. Clearly, $\alpha \in [0, 1]$: α being 0 or 1 indicates unanimity, whereas $\alpha = 0.5$ means $r = s$, i.e., maximal conflict in the body of evidence. Definition 4 captures this intuition.

Definition 4. $\mathit{conflict}(r, s) = \min(\alpha, 1 - \alpha)$

2.1.4 Certainty in Evidence

In our approach, as Definition 1 shows, certainty depends on a PCDF. The particular PCDF we consider is the one of Definition 3, which generalizes over binary events. It helps in our analysis to combine these so as to define certainty based on evidence $\langle r, s \rangle$, where r and s are the positive and negative bodies of evidence, respectively. Definition 5 merely writes certainty as a function of r and s.

Definition 5. $c(r, s) = \frac{1}{2} \int_0^1 |\frac{(x^r(1-x)^s}{\int_0^1 x^r(1-x)^s dx} - 1| dx$

Recall that $t = r + s$ is the total body of evidence. Thus $r = t\alpha$ and $s = t(1 - \alpha)$. We can thus write $c(r, s)$ as $c(t\alpha, t(1 - \alpha))$. When α is fixed, certainty is a function of t,

and is written $c(t)$. When t is fixed, certainty is a function of α, and is written $c(\alpha)$. And, $c'(t)$ and $c'(\alpha)$ are the corresponding derivatives.

2.1.5 Trust Space

The traditional probability model outlined above ignores uncertainty. Thus it predicts the same probability whenever r and s have the same ratio (correcting for the effect of the Laplace smoothing) even though the total amount of evidence may differ significantly. For example, we would obtain $p = 0.70$ whether $r = 6$ and $s = 2$ or $r = 69$ and $s = 29$. However, the result would be intuitively much more certain in the second case because of the overwhelming evidence: the good outcomes hold up even after a large number of interactions. For this reason, we favor an approach that accommodates certainty.

Following [Jøsang, 2001], a trust space consists of *trust reports* modeled in a three-dimensional space of reals in $[0, 1]$. Each point in this space is a triple $\langle b, d, u\rangle$, where $b+d+u = 1$, representing the weights assigned to belief, disbelief, and uncertainty, respectively. Certainty c is simply $1-u$. Thus $c = 1$ and $c = 0$ indicate perfect knowledge and ignorance, respectively. Definition 6 states this formally.

Definition 6. *Trust space* $T = \{\langle b, d, u\rangle | b \geq 0, d \geq 0, b + d > 0, u > 0, b + d + u = 1\}$

Combining trust reports is nontrivial. Our proposed definition of certainty is key in accomplishing a bijection between evidence and trust reports. The problem of combining independent trust reports is reduced to the problem of combining the evidence underlying them. Section 2.1.6 further explains how evidence and trust space are used in this approach.

2.1.6 From Evidence to Trust

As remarked above, it is easier to aggregate trust in the evidence space and to discount it in trust space. As trust is propagated, each agent involved would map the evidence it obtains to trust space, discount it, map it back to evidence space, and aggregate it as evidence. We cannot accomplish the above merely by having the agents perform all their calculations in either the evidence space or the trust space. Therefore, we need a function to map evidence space to trust space. This function should be (uniquely) invertible.

Definition 7 shows how to map evidence to trust. This mapping relates positive and negative evidence to belief and disbelief, respectively, but with each having been discounted

by the certainty. Definition 7 generalizes the pattern of [Jøsang, 1998] by identifying the degree of conflict α and certainty $c(r, s)$. The development below describes two important differences with Jøsang's approach.

Definition 7. *Let* $Z(r, s) = \langle b, d, u \rangle$ *be a transformation from* E *to* T *such that* $Z = \langle b(r, s), d(r, s), u(r, s) \rangle$*, where*

1. $b(r, s) = \alpha c(r, s)$
2. $d(r, s) = (1 - \alpha)c(r, s)$
3. $u(r, s) = 1 - c(r, s)$

where $\alpha = \frac{r}{t}$ *and* $c(r, s)$ *is as defined in 5.*

One can easily verify that $c(0, 1) > 0$. In general, because $t = r+s > 0$, $c(r, s) > 0$. Moreover, $c(r, s) < 1$: thus, $1 - c(r, s) > 0$. This ensures that $b + d > 0$, and $u > 0$. Notice that $\alpha = \frac{b}{b+d}$.

Jøsang [Jøsang, 1998] maps evidence $\langle r, s \rangle$ to a trust triple $(\frac{r}{t+1}, \frac{s}{t+1}, \frac{1}{t+1})$. Two main differences with our approach are:

- Our definition of certainty depends not only on the amount of evidence but also on the conflict, which Jøsang ignores.
- Our definition of certainty incorporates a subtle characterization of the probabilities whereas, in essence, Jøsang defines certainty as $\frac{t}{t+1}$. He offers no mathematical justification for doing so. The underlying intuition seems to be that certainty increases with increasing evidence. We finesse this intuition to capture that increasing evidence yields increasing certainty but *only if* the conflict does not increase.

Section 2.2.2 shows a counterintuitive consequence of Jøsang's definition.

In passing, we observe that discounting as defined by Jøsang [Jøsang, 1998] and Wang and Singh [Wang and Singh, 2006] reduces the certainty but does not affect the probability of a good outcome. Discounting in their manner involves multiplying the belief and disbelief components by the same constant, $\gamma \neq 0$. Thus a triple $\langle b, d, u \rangle$ is discounted by γ to yield $\langle b\gamma, d\gamma, 1 - b\gamma - d\gamma \rangle$. Recall that the probability of a good outcome is given by $\alpha = \frac{b}{b+d}$. The probability of a good outcome from a discounted report is $\frac{b\gamma}{b\gamma+d\gamma} = \frac{b}{b+d}$, which is the same as α.

Let us consider a simple example. Suppose Alice has eight good and two bad transactions with a service provider, Charlie, yielding a trust triple of $\langle 0.42, 0.1, 0.48\rangle$. Suppose Bob has one good and four bad transactions with the Charlie, yielding a trust triple of $\langle 0.08, 0.33, 0.59\rangle$. Suppose Alice and Bob report their ratings of Charlie to Ralph. Suppose that Ralph's trust in Alice is $\langle 0.2, 0.3, 0.5\rangle$ and his trust in Bob is $\langle 0.9, 0.05, 0.05\rangle$. Ralph then carries out the following steps.

- Ralph discounts Alice's report by the trust he places in Alice (i.e., the belief component of his triple for Alice, 0.2), thus yielding $\langle 0.084, 0.02, 0.896\rangle$. Ralph discounts Bob's report in the same way by 0.9, thereby yielding $\langle 0.072, 0.297, 0.631\rangle$.
- Ralph transforms the above two discounted reports into the evidence space, thus obtaining $\langle 0.429, 0.107\rangle$ from Alice's report and $\langle 0.783, 3.13\rangle$ from Bob's report.
- Ralph combines these in evidence space, thus obtaining a total evidence of $\langle 1.212, 3.237\rangle$.
- Transforming these back to trust space, Ralph obtains that he trusts Charlie to $\langle 0.097, 0.256, 0.645\rangle$.

Notice how, in the above, since Ralph places much greater credibility in Bob than in Alice, Ralph's overall assessment of Charlie is closer to Bob's than to Alice's.

2.2 Important Properties and Computation

We now show that the above definition yields important formal properties and how to compute with this definition.

2.2.1 Increasing Experiences with Fixed Conflict

Consider the scenario where the total number of experiences increases for fixed $\alpha = 0.50$. For example, compare observing 5 good episodes out of 10 with observing 50 good episodes out of 100. The expected value, α, is the same in both cases, but the certainty is clearly greater in the second. In general, we would expect certainty to increase as the amount of evidence increases. Definition 5 yields a certainty of 0.46 from $\langle r, s\rangle = \langle 5, 5\rangle$, but a certainty of 0.70 for $\langle r, s\rangle = \langle 50, 50\rangle$.

Figure 2.1 plots how certainty varies with t both in our approach and in Jøsang's approach. Notice that the specific numeric values of certainty in our approach should not

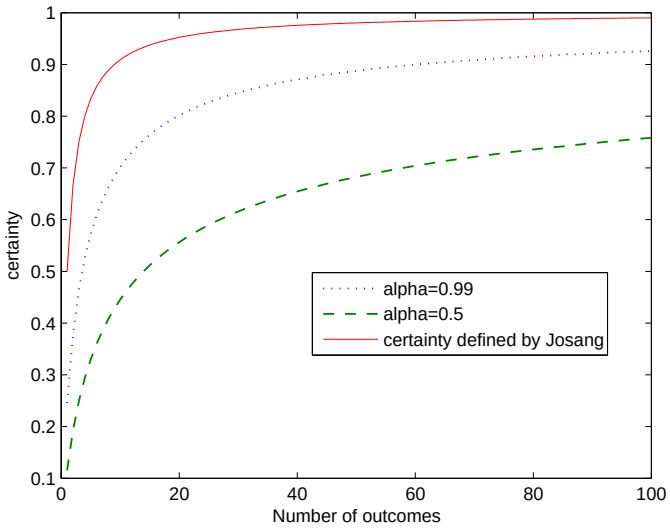


Figure 2.1: Certainty increases with t both in Jøsang's approach and in our approach when the level of conflict is fixed; for our approach, we show $\alpha = 0.5$ and $\alpha = 0.99$; in Jøsang's approach, certainty is independent of the level of conflict; X-axis: t, the amount of evidence; Y-axis: $c(t)$, the corresponding certainty

be compared to those in Jøsang's approach. The trend is monotonic and asymptotic to 1 in both approaches. The important observation is that our approach yields a higher certainty curve when the conflict is lower.

Theorem 1 captures this property in general for our approach.

Theorem 1. *Fix α. Then $c(t)$ increases with t for $t > 0$.*

Proof sketch: The proof of this theorem is built via a series of steps. The main idea is to show that $c'(t) > 0$ for $t > 0$. Here $f(r, s, x)$ is the function of Definition 3 viewed as a function of r, s, and x.

1. Let $f(r, s, x) = \frac{(x^r(1-x)^s}{\int_0^1 x^r(1-x)^s dx}$. Then $c(r, s) = \frac{1}{2}\int_0^1 |f(r, s, x) - 1|dx$. We can write c and f as functions of t and α. That is, $c = c(t, \alpha)$ and $f = f(t, \alpha, x)$.
2. Eliminate the absolute sign. By Lemma 22, we can define A and B where $f(A) = f(B) = 1$ so that $c(t, \alpha) = \frac{1}{2}\int_0^1 |f(t, \alpha, x) - 1|dx = \int_A^B (f(t, \alpha, x) - 1)dx$ A and B are also functions of t and α.
3. When α is fixed, $c(t, \alpha)$ is a function of t and we can differentiate it by t. Notice that: $\frac{d}{dt}\int_{A(t)}^{B(t)}(f(t, x) - 1)dx = B'(t)(f(t, B) - 1) - A'(t)(f(t, A) -$

$1) + \int_{A(t)}^{B(t)} (\frac{\partial}{\partial t} f(t,x) - 1)dx$. The first two terms are 0 by the definition of A and B.

4. Using the formula, $\frac{d}{dx} a^{f(x)} = \ln a f'(x) a^{f(x)}$ we can calculate $\frac{\partial}{\partial t} f(t, \alpha, x)$.
5. Then we break the result into two parts. Prove the first part to be positive by Lemma 22, and the second part to be 0 by exploiting the symmetry of the terms.

Hence, $c'(t) > 0$, as desired. □

The appendix includes full proofs of this and the other theorems.

2.2.2 Increasing Conflict with Fixed Experience

Another important scenario is when the total number of experiences is fixed, but the evidence varies to reflect different levels of conflict by using different values of α. Clearly, certainty should increase as r or s dominates the other (i.e., α approaches 0 or 1) but should reduce as r and s are in balance (i.e., α approaches 0.5). Figure 2.2 plots certainty for fixed t and varying conflict.

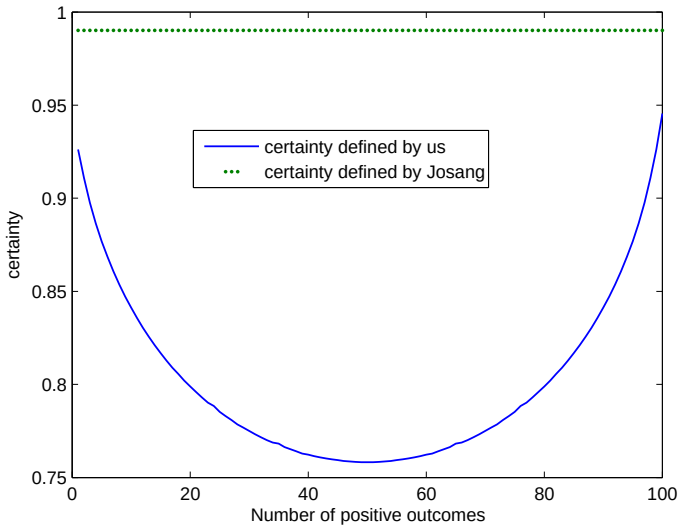


Figure 2.2: Certainty is concave when t is fixed at 100; X-axis: $r+1$; Y-axis: $c(\alpha)$; minimum occurs at $r = s = 5$; certainty according to Jøsang is constant and is shown for contrast

More specifically, consider Alice's example from Chapter 1.3. Table 5.1 shows the effect of conflict where $t = 4$. Briefly, Yu and Singh [Yu and Singh, 2002] base uncertainty

Table 2.1: Certainty computed by different approaches for varying levels of conflict

	$\langle 0,4 \rangle$	$\langle 1,3 \rangle$	$\langle 2,2 \rangle$	$\langle 3,1 \rangle$	$\langle 4,0 \rangle$
Our approach	0.54	0.35	0.29	0.35	0.54
Jøsang	0.80	0.80	0.80	0.80	0.80
Yu & Singh	1.00	1.00	1.00	1.00	1.00

not on conflict, but on intermediate (neither positive not negative) outcomes. If there is no intermediate value, the certainty is maximum.

Let's revisit Pete's example of Section 1.3. In our approach, Dawn and Ed's diagnoses would correspond to two b, d, u triples (where b means "tumor" and d means "not a tumor"): $(0.20, 0.79, 0.01)$ and $(0.20, 0.79, 0.01)$, respectively. Combining these we obtain the b, d, u triple of $(0.21, 0.78, 0.01)$. That is, the weight assigned to a tumor is 0.21 as opposed to 0.725 by Dempster-Shafer theory, which is unintuitive, because a tumor is Dawn and Ed's least likely prediction.

Theorem 2 captures the property that certainty increases with increasing unanimity.

Theorem 2. *The function $c(\alpha)$ is decreasing when $0 < \alpha \leq \frac{1}{2}$, and increasing when $\frac{1}{2} \leq \alpha < 1$. Thus $c(\alpha)$ is minimized at $\alpha = \frac{1}{2}$.*

Proof sketch: The main idea is to show that $c'(\alpha) < 0$ when $\alpha \in (0, 0.5)$ and $c'(\alpha) > 0$ when $\alpha \in [0.5, 1.0)$. This is accomplished via steps similar to those in the proof of Theorem 1. First remove the absolute sign, then differentiate, then prove the derivative is negative in the interval $(0, 0.5)$ and positive in the interval $(0.5, 1)$. □

Putting the above results together suggests that the relationship between certainty on the one hand and positive and negative evidence on the other hand is nontrivial. Figure 2.3 confirms this intuition by plotting certainty against r and s as a surface. The surface rises on the left and right corresponding to increasing unanimity of negative and positive evidence, respectively, and falls in the middle as the positive and negative evidence approach parity. The surface trends upward going from front to back corresponding to the increasing evidence at a fixed level of conflict.

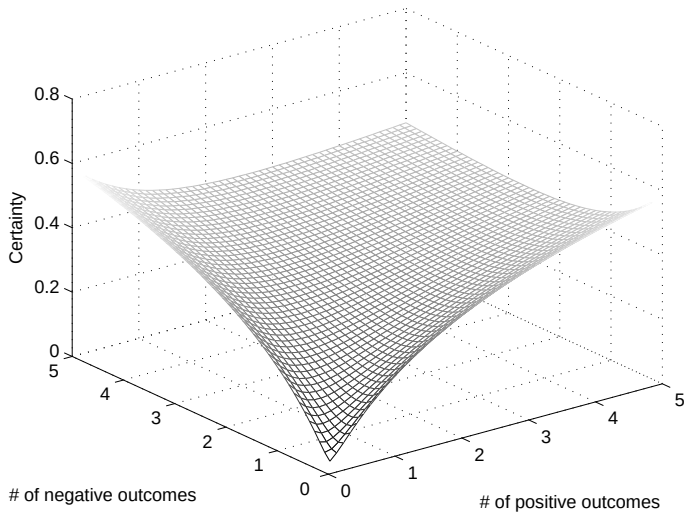


Figure 2.3: X-axis: r, number of positive outcomes; Y-axis: s, number of positive outcomes; Z-axis: certainty $c(r, s)$, the corresponding certainty

It is worth emphasizing that certainty does *not* necessarily increase even as the evidence grows. When additional evidence conflicts with the previous evidence, a growth in evidence can possibly yield a loss in certainty. This accords with intuition because the arrival of conflicting evidence can shake one's beliefs, thus lowering one's certainty.

Figure 2.4 demonstrates a case where we first acquire negative evidence, thereby increasing certainty. Next we acquire positive evidence, which conflicts with the previous evidence, thereby lowering certainty. In Figure 2.4, the first ten transactions are all negative; the next ten transactions are all positive. Certainty grows monotonically with unanimous evidence and falls as we introduce conflicting evidence. Because of the dependence of certainty on the size of the total body of evidence, it doesn't fall as sharply as it rises, and levels off as additional evidence is accrued.

2.2.3 Bijection Between Evidence and Trust

A major motivation for modeling trust and evidence spaces is that each space facilitates computations of different kinds. Discounting trust is simple in the trust space whereas aggregating trust is simple in the evidence space.

Recall that, as Theorem 1 shows, we associate greater certainty with larger bodies

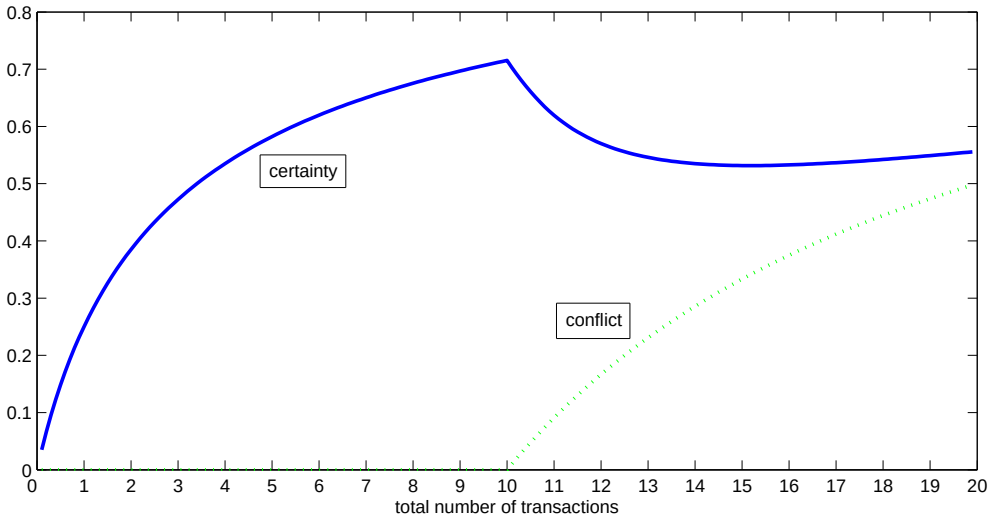


Figure 2.4: Certainty increases as unanimous evidence increases; the addition of conflicting evidence lowers certainty; X-axis: number of transactions. Y-axis: c certainty.

of evidence (assuming conflict is fixed). Thus the certainty of trust reports to be combined clearly matters: we should place additional credence where the certainty is higher (generally meaning the underlying evidence is stronger). Consequently, we need a way to map a trust report to its corresponding evidence in a manner that higher certainty yields a larger body of evidence.

The ability to combine trust reports effectively relies on being able to map between the evidence and the trust spaces. With such a mapping in hand, to combine two trust reports, we would simply perform the following steps:

1. Map trust reports to evidence.
2. Combine the evidence.
3. Transform the combined evidence to a trust report.

The following theorem establishes that Z has a unique inverse, Z^{-1}.

Theorem 3. *The transformation Z is a bijection.*

Proof sketch: Given $\langle b, d, u\rangle \in T$, we need $(r, s) \in E$ such that $Z(r, s) = \langle b, d, u\rangle$. As explained in Section 2.1.6, $\alpha = \frac{b}{b+d}$. Thus, we only need to find t such that $c(t) = 1 - u$. The existence and uniqueness of t is proved by showing that

1. $c(t)$ is increasing when $t > 0$ (Theorem 1)
2. $\lim_{t\to\infty} c(t) = 1$ (Lemma 24)
3. $\lim_{t\to 0} c(t) = 0$ (Lemma 25)

Thus there is a unique t that corresponds to the desired level of certainty. □

2.2.4 Computing the Bijection: Algorithm and Complexity

Theorem 3 shows the existence of Z^{-1}. However, no closed form is known for Z^{-1}. For this reason, we develop an iterative, approximate algorithm for computing Z^{-1}.

As explained in Section 2.1.6, the ratio α depends solely on b and d. Thus given $\langle b, d, u\rangle$, we can determine α immediately as $\frac{b}{b+d}$. Since $r = t\alpha$ and $s = t(1-\alpha)$, in this manner, we know the relationships between r and t, and between s and t. But we do not immediately know t. In essence, no closed form for Z^{-1} is known because no closed form is known for its third component, namely, t.

The intuition behind our algorithm to compute t is that after fixing α, we should find t that would yield the desired certainty of $(1-u)$. This works because, as remarked in the proof sketch for Theorem 3, $c(t)$ ranges between 0 and 1. Further, Theorem 1 shows that for fixed α, $c(t)$ is monotonically increasing with t. In general, t being the size of the body of evidence is not bounded. However, as a practical matter, an upper bound can be placed on t. Thus, a binary search is an obvious approach. (When no bound is known, a simple approach would be to (1) guess exponentially increasing values for t until a value is found for which the desired certainty is exceeded; and then (2) conduct binary search between that and the previously guessed value.)

For binary search, since we are dealing with real numbers, it is necessary to specify $\epsilon > 0$, the desired precision to which the answer is to be computed.

Algorithm 2.2.4 calculates Z^{-1} via binary search on $c(t)$ to a specified precision, $\epsilon > 0$. Here $t_{max} > 0$ is the maximum size of the body of evidence considered. (Recall that lg means logarithm to base 2.)

Theorem 4. *The complexity of Algorithm 2.2.4 is* $\Omega(-\lg \epsilon)$.

Proof: After the **while** loop iterates i times, $t_2 - t_1 = t_{max}2^{-i}$. Eventually, $t_2 - t_1$ falls below ϵ, thus terminating the loop. Assume the loop terminates in n iterations. Then,

$\alpha = \frac{b}{b+d}$
$c = 1 - u$
$t_1 = 0$
$t_2 = t_{max}$
while $t_2 - t_1 \geq \epsilon$ **do**
    $t = \frac{t_1+t_2}{2}$;
    if $c(t) < c$ **then**
        $t_1 = t$
    else
        $t_2 = t$
    end if
end while
return $r = (t\alpha)$, $s = t - r$

$t_2 - t_1 = t_{max}2^{-n} < \epsilon \leq t_{max}2^{-n+1}$. This implies $2^n > \frac{t_{max}}{\epsilon} \geq 2^{n-1}$. That is, $n > (\lg t_{max} - \lg \epsilon) \geq n - 1$.

Chapter 3

Propagating Trust

Say an agent A_r does not have adequate evidence to determine whether to trust an agent A_g. A_r could obtain evidence from other agents, i.e., *witnesses*, who have experience with A_g. The mechanism of finding witnesses is irrelevant here: a possible approach is based on referrals [Yu and Singh, 2002]. A_r can evaluate the trustworthiness of A_g by combining the trust in A_g placed by those witnesses. In general, A_r would not have direct experience with the witnesses, and would need witnesses for them, and so on: so this would be a case of propagating trust.

We use the term "reference" to generalize over a referral or another means for an agent to indicates its level of trust in another agent. A reference could correspond to a URI on a web page or even to a reference in a specialized application or vocabulary such as Friend of a Friend (FOAF).

Definition 8. *A reference r to A_j by A_i is represented by $\langle A_i, A_j, M_{ij} \rangle$, where M_{ij} represents the trust in A_j placed by A_i.*

Definition 9. *A trust network TN(A_r, A_g, A, R, W) is an acyclic directed graph rooted at A_r, where A is a finite set of agents $\{A_1, A_2, \ldots, A_N\}$, R is a finite set of references $\{r_1, r_2, \ldots, r_n\}$, and W is a set of witnesses for A_g.*

A referral network, which is a kind of a trust network, can be constructed as described in [Yu and Singh, 2002].

Building on the above model of a social network as a graph, we propose a model called *CertProp* that handles the propagation of trust. CertProp is based on three operators.

Concatenation or $\otimes$ deals with the propagation of trust ratings along a path. Selection or Ⓢ chooses the most trustworthy path to each witness, whereas aggregation of $\oplus$ deals with the combination of trust ratings from paths between the same source and target. The aggregation operator is due to Jøsang [Jøsang, 1998]; the others are new here.

3.1 Path Algebra

We use the idea of *path algebra* from the generalized transitive closure literature [Richardson et al., 2003] for our computation of the merged trust. Path algebra provides a means to talk about trust propagation and aggregation. Below we formalize it in our terms.

Definition 10. *Let $G = (V, E)$ be a graph. For each edge $e(i, j) \in E$, where i is the source and j is the destination, define the label associated with the edge as $L(e(i, j)) \in L$. Here L is the range of the labels, for example, R for real numbers, R^2 for vectors. The edge $e(i, j)$ together with the label $L(e(i, j))$ is called an* edge tuple.

Definition 11. *A path from node i to node j is denoted by $P(i, j)$, which is the concatenation of an ordered set of labeled edges $e_k(v_k, v_{k+1})$, for $k = 1, \cdots, n$, where $i = v_1$ and $j = v_{n+1}$. Define the label associated with the path as $L(P(i, j)) \in L$. The label associated with the path is computed as a function of the labels associated with the edges in the path through the concatenation operator $\otimes$. The path together with its label is called a* path tuple.

Definition 12. *A* path-set *from node i to node j, denoted by $\psi(i, j)$, is the set of all paths from i to j in the given graph $G = (V, E)$. $\psi(i, j) = \{P_k(i, j)\}$, for $k = 1, \cdots, m$. Define the label for the* path-set *as $L(\psi(i, j)) \in L$, which should be computed as a function of all the labels of the paths in $\psi(i, j)$ through an aggregation operator $\oplus$. A* path-set *together with its label is called a* path-set *tuple.*

Definition 13. *Concatenation operator $\otimes : L \times L \mapsto L$. This accommodates the propagation of trust along a path.*

Definition 14. *Aggregation operator $\oplus : L \times L \mapsto L$. This accommodates combining trust from different paths.*

3.2 Merge and Combine Trust Ratings

Given a trust network, we adapt the two operators used in the path algebra to merge the trust.

Assume agent A has a trust M_1 in agent B's references and B has a trust M_2 in agent C. Suppose $M_1 = (b_1, d_1, u_1)$ and $M_2 = (b_2, d_2, u_2)$. We use a different approach from Jøsang [Jøsang, 1998], in which A disbelieves B's references means that A thinks B is uncertain about agent C's trustworthiness. In our approach, A's belief in B determines how much A discounts the information B provides, which is the number of transactions B has with C. So we define A's trust in C due to the reference from B to be $M = M_1 \otimes M_2$. Here $\otimes$ is the *concatenation operator*.

Motivation: We treat the number of positive and negative transactions provided by agent B as the amount of information provided by B. We discount this information by how much A trusts in B, thus we have the following definition:

Definition 15. *Concatenation operator $\otimes$ [Jøsang, 1998]. Let $Z = (B, D, U)$ be the transformation from evidence space to belief space as defined in Definition 7 and $Z^{-1} = (R, S)$ is the inverse of Z. Suppose $M_1 = (b_1, d_1, u_1)$ and $M_2 = (b_2, d_2, u_2)$ are two belief functions, we define $M = M_1 \otimes M_2 = (b, d, u)$ as:*

$$\begin{aligned} b &= B(b_1 r_2, b_1 s_2) \\ d &= D(b_1 r_2, b_1 s_2) \\ u &= U(b_1 r_2, b_1 s_2) \\ r_2 &= R(b_2, d_2, u_2) \\ s_2 &= S(b_2, d_2, u_2) \end{aligned} \tag{3.1}$$

Assume agents A and B have trust M_1 and M_2, respectively, in A_g. The combined trust of A and B in A_g is captured via the *aggregation operator* $\oplus$, as in $M_1 \oplus M_2$. Suppose A has r_1 positive experiences and s_1 negative experiences with A_g, and B has r_2 positive experiences and s_2 negative experiences with A_g. Then the combined evidence will be $r_1 + r_2$ positive experiences and $s_1 + s_2$ negative experiences with A_g.

In order to combine the beliefs M_1 and M_2, we first transform the two beliefs to two bodies of evidence in the evidence space, combine them, then map the combined evidence back to belief space.

Motivation: Suppose the number of transactions of A with A_g is r and s respectively, the number of transactions of B with A_g is r' and s' respectively. If we look at this in a different way, it is to say A_g has provided $r+r'$ positive transactions and $s+s'$ negative transactions to A and B in total. Thus we have the following definition:

Definition 16. *Aggregation operator* $\oplus$. *Let* $Z = (B, D, U)$ *be the transformation from evidence space to belief space as defined in Definition 7 and* $Z^{-1} = (R, S)$ *is the inverse of* Z. *Suppose* $M_1 = (b_1, d_1, u_1)$ *and* $M_2 = (b_2, d_2, u_2)$. *Then* $M_1 \oplus M_2 = M = (b, d, u)$ *where*

$$\begin{aligned} b &= B(r_1 + r_2, s_1 + s_2) \\ d &= D(r_1 + r_2, s_1 + s_2) \\ u &= U(r_1 + r_2, s_1 + s_2) \\ r_1 &= R(b_1, d_1, u_1), r_2 = R(b_2, d_2, u_2) \\ s_1 &= S(b_1, d_1, u_1), s_2 = S(b_2, d_2, u_2) \end{aligned} \tag{3.2}$$

By using path algebra, we can enumerate all paths from the originating agent to the target. Using concatenate operator to combine all beliefs along each path, then using aggregate operators to combine all beliefs from all paths. But there is a problem by doing this way: For example, suppose agent A has two neighbors N_1 and N_2. N_1 refers witness W_1. N_2 refers witness W_1 and W_2. W_1 reported $\langle 5, 5\rangle$ and W_2 reported $\langle 1, 1\rangle$. Suppose A's trust in N_1, N_2, and N_1's trust in W_1 are the same , $(0.9, 0.05, 0.05)$. N_2's trust in W_1 is $(0.8, 0.1, 0.1)$ and N_2's trust in W_2 is $(0.85, 0.1, 0.05)$. There are 3 paths from A to the target, two from W_1 and one from W_2. By using the above method, the aggregated trust $\langle R, S\rangle = \langle 6.477, 6.477\rangle$. This does not make sense , since the total number of transactions made by W_1 and W_2 is $\langle 6, 6\rangle$. The problem comes from double counting.

So how to combine beliefs which come from the same source, but propagate from different paths? For example, witness W_1 tells agent A that he has 1 positive and 1 negative experience with the target. He tells the same information to agent B. Both A and B are agent A_r's neighbors. A_r can not use the aggregation operator $\oplus$ to combine the beliefs from A and B, since their beliefs come from the same source and it will create double counting. So we define a new operator, which is called selection operator, Ⓢ. To select one path from multiple paths which end with the same point. In this example, we have two paths: $P_1 = A_r \to A \to W_1$, and $P_2 = A_r \to B \to W_1$. Let A_r's trust in A and B is 0.8 and 0.5 respectively. A and B's trust in W_1 is 0.5 and 0.6 respectively. We pick the most

reliable path from A_r to W_1. Since $0.8 \times 0.5 > 0.5 \times 0.6$, so P_1 is the most reliable path. So $P_1 Ⓢ P_2 = P_1$.

Motivation: Why do we select the most reliable path? For example, if A_r has perfect trust in A and A has a perfect trust in W_1, then A_r should not consider the opinion from other agent since it has perfect trust in A. So only the path through A counts which is the most reliable path.

Definition 17. *Selection operator Ⓢ. Let* $P_1 = A_r \to A_1 \to A_2 \cdots \to A_m \to W$ *and* $P_2 = A_r \to B_1 \to B_2 \cdots \to B_n \to W$ *are two paths from* A_r *to* W*. Let* A_r*'s trust in* A_1 *and* B_1 *are* (b_0, d_0, u_0) *and* (b'_0, d'_0, u'_0)*. Let* A_i*'s trust in* A_{i+1} *be* (b_i, d_i, u_i) *and* B_i*'s trust in* B_{i+1} *be* (b'_i, d'_i, u'_i)*.* A_m*'s trust in* W *is* (b_m, d_m, u_m) *and* b_n*'s trust in* W *is* (b'_n, d'_n, u'_n)*. Then if* $\prod_{i=0}^{m} b_i <= \prod_{i=0}^{n} b'_i$*, then* $P_1 Ⓢ P_2 = P_2$*, otherwise* $P_1 Ⓢ P_2 = P_1$

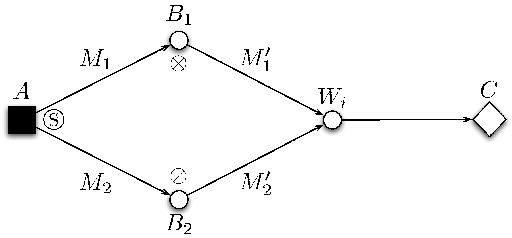


Figure 3.1: Selection operator Ⓢ

For a given trust network, to propagate trust of A with C, we combine beliefs as follows. Using selection, we find the best path (to a fixed depth) from A to each witness. Then we concatenate beliefs along this path. The resulting belief is the consolidated belief supported by that witness. Then we combine all beliefs supported by all witnesses by using aggregation.

For example, in Figure 3.1, by applying all three operators, the propagated trust is $M = ((M_1 \otimes M'_1) Ⓢ (M_2 \otimes M'_2) Ⓢ (M_3 \otimes M'_3)) \otimes M_5) \oplus (M_4 \otimes M'_4)$.

To illustrate the double counting problem without selection in Figure 3.2, let $M_1 \otimes M'_1 = \langle 0.8, 0.1, 0.1 \rangle$, $M_2 \otimes M'_2 = \langle 0.6, 0.2, 0.2 \rangle$, $M_3 \otimes M'_3 = \langle 0.6, 0.3, 0.1 \rangle$, $Z^{-1}(M_5) = \langle 50, 5 \rangle$, $M_4 = \langle 0.9, 0.05,$
$0.05 \rangle$, and $Z^{-1}(M'_4) = \langle 20, 0 \rangle$. The actual trust of C is M, where $Z^{-1}(M) = Z^{-1}(M_5) \oplus Z^{-1}(M'_4) = \langle 70, 5 \rangle$. However, by applying only concatenation and aggregation, the esti-

mated trust $Z^{-1}(M) = Z^{-1}(((M_1 \otimes M_1') \oplus (M_2 \otimes M_2') \oplus (M_3 \otimes M_3')) \otimes M_5) \oplus (M_4 \otimes M_4')) = \langle 118, 10 \rangle$, which double-counts M_5. This shows that multiple paths via the same witness can lead to double-counting. This is especially obvious when the aggregated belief of the paths from the source to a witness is greater than one.

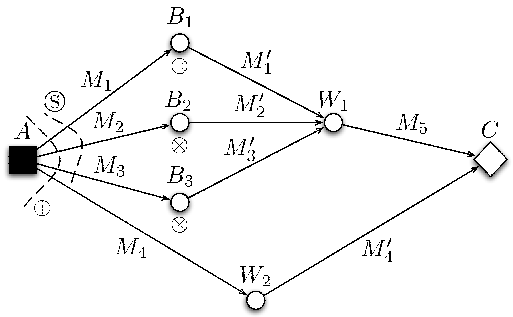


Figure 3.2: Trust propagation in a social network

3.3 Properties of the Operators

In order to combine the trust ratings in a meaningful way and to deal with deception caused by malicious agents, we desire certain properties of our approach. We motivate and prove that the concatenation operator $\otimes$ and aggregation operator $\oplus$ satisfy the following properties. The proofs of the theorems are given in the appendix.

Theorem 5. *The concatenation operator $\otimes$ is associative.*

This property enables us to merge the trust in a bottom up fashion. For example, consider the path $A_r \to B \to C \to A_g$, (see Figure 3.3). Suppose it is the only path from A_r to A_g in the trust network. The trust of A_r in B is M_1, the trust of B in C is M_2 and the trust of C in A_g is M_3. If we merge M_1 and M_2, we will get the trust of A_r in C, which is $M_1 \otimes M_2$, when we merge it with M_3, we get the trust of A_r in A_g, which is $(M_1 \otimes M_2) \otimes M_3$. If we do it bottom up, then we merge M_2 and M_3 and get the trust of B in A_g and then merge it with M_1 and get the trust of A_r in A_g which is $M_1 \otimes (M_2 \otimes M_3)$. Certainly we expect these two results are the same, that is $(M_1 \otimes M_2) \otimes M_3 = M_1 \otimes (M_2 \otimes M_3)$. This means $\otimes$ should be associative.

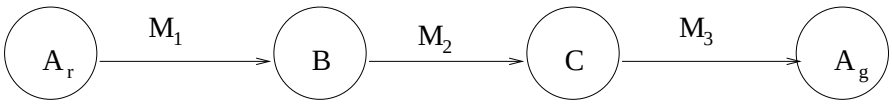


Figure 3.3: Path in a trust network

Theorem 6. *The aggregation operator $\oplus$ is associative.*

This is required to aggregate trust ratings across independent paths. Suppose A_r has three neighbors B, C and D who are the only witnesses of A_g, (see Figure 3.4). The trust ratings of B, C and D in A_g are M_1', M_2' and M_3', respectively. Suppose the trust ratings of A_r in B, C and D are M_1, M_2 and M_3, respectively. Then according to the algorithm to find the trust of A_r to A_g. We first concatenate the trust ratings, $M_1 \otimes M_1'$, $M_2 \otimes M_2'$ and $M_3 \otimes M_3'$. Then we aggregate them. There are different ways to aggregate them. We can aggregate $M_1 \otimes M_1'$ with $M_2 \otimes M_2'$ and then with $M_3 \otimes M_3'$, or we aggregate $M_2 \otimes M_2'$ with $M_3 \otimes M_3'$, and then with $M_1 \otimes M_1'$. These should be the same. So we require $((M_1 \otimes M_1') \oplus (M_2 \otimes M_2')) \oplus (M_3 \otimes M_3') = (M_1 \otimes M_1') \oplus ((M_2 \otimes M_2') \oplus (M_3 \otimes M_3'))$. That is, $\oplus$ should be associative.

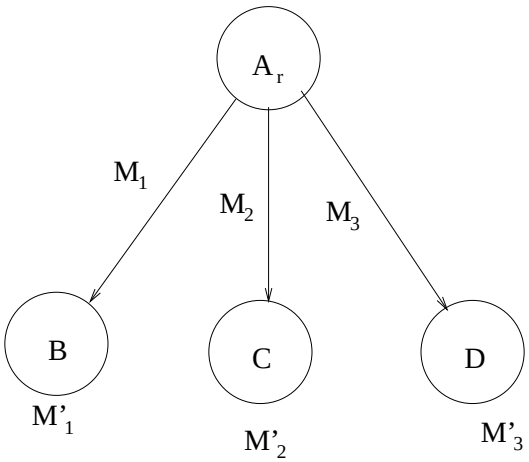


Figure 3.4: Associativity of $\oplus$

Theorem 7. *The selection operator Ⓢ is associative.*

Proof. We have to show M_1 Ⓢ $(M_2$ Ⓢ $M_3) = (M_1$ Ⓢ $M_2)$ Ⓢ M_3. Let $M_1 = \langle b_1, d_1, u_1 \rangle$, $M_2 = \langle b_2, d_2, u_2 \rangle$, and $M_3 = \langle b_3, d_3, u_3 \rangle$.

Case 1 : Suppose b_1 is larger than b_2 and b_3. If $b_2 \geq b_3$. Then $M_1 \circledS (M_2 \circledS M_3) = M_1 = M_1 \circledS M_3 = (M_1 \circledS M_2) \circledS M_3$. If $b_2 < b_3$, then $M_1 \circledS (M_2 \circledS M_3) = M_1 \circledS M_3 = M_1 = M_1 \circledS M_3 = (M_1 \circledS M_2) \circledS M_3$.

Case 2 : Suppose b_2 is larger than b_1 and b_3. Then $M_1 \circledS (M_2 \circledS M_3)$ $= M_1 \circledS M_2 = M_2 = M_2 \circledS M_3 = (M_1 \circledS M_2) \circledS M_3$.

Case 3 : Suppose b_3 is larger than b_1 and b_2. Then $M_1 \circledS (M_2 \circledS M_3)$ $= M_1 \circledS M_3 = M_3 = (M_1 \circledS M_2) \circledS M_3$.

Thus, $M_1 \circledS (M_2 \circledS M_3) = (M_1 \circledS M_2) \circledS M_3$. □

Theorem 8. *The aggregation operator $\oplus$ is commutative.*

This is required to aggregate trust ratings across paths that are not ordered. For example, suppose A_r has two neighbors B and C, who are the only witnesses. The trust ratings of B and C in A_g are M_1' and M_2'. (see Figure 3.5). Suppose the trust ratings of A_r in B, and C are M_1, and M_2, respectively. After concatenating the trust ratings, we obtain $M_1 \otimes M_1'$, and $M_2 \otimes M_2'$. We have two ways to aggregate them: $(M_1 \otimes M_1') \oplus (M_2 \otimes M_2')$, or $(M_1 \otimes M_1') \oplus (M_2 \otimes M_2')$, these should be the same. That is, $(M_1 \otimes M_1') \oplus (M_2 \otimes M_2') = (M_1 \otimes M_1') \oplus (M_2 \otimes M_2')$. So $\oplus$ should be commutative.

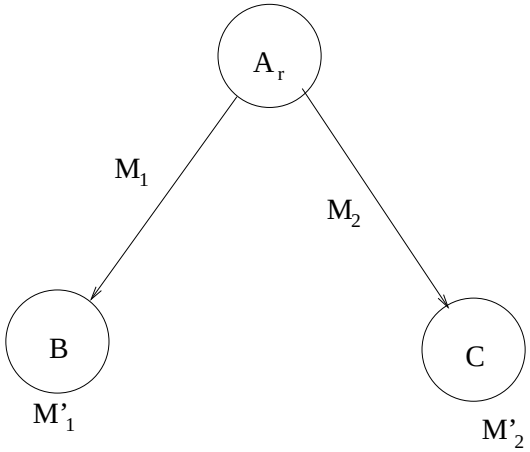


Figure 3.5: Commutativity of $\oplus$

Theorem 9. *The aggregation operator $\circledS$ is commutative.*

Proof. We have to show $M_1 Ⓢ M_2 = M_2 Ⓢ M_1$. Let $M_1 = \langle b_1, d_1, u_1 \rangle$, $M_2 = \langle b_2, d_2, u_2 \rangle$, and $M_1 Ⓢ M_2 = M_1$, which means $b_1 > b_2$. Then $M_2 Ⓢ M_1 = M_1$. □

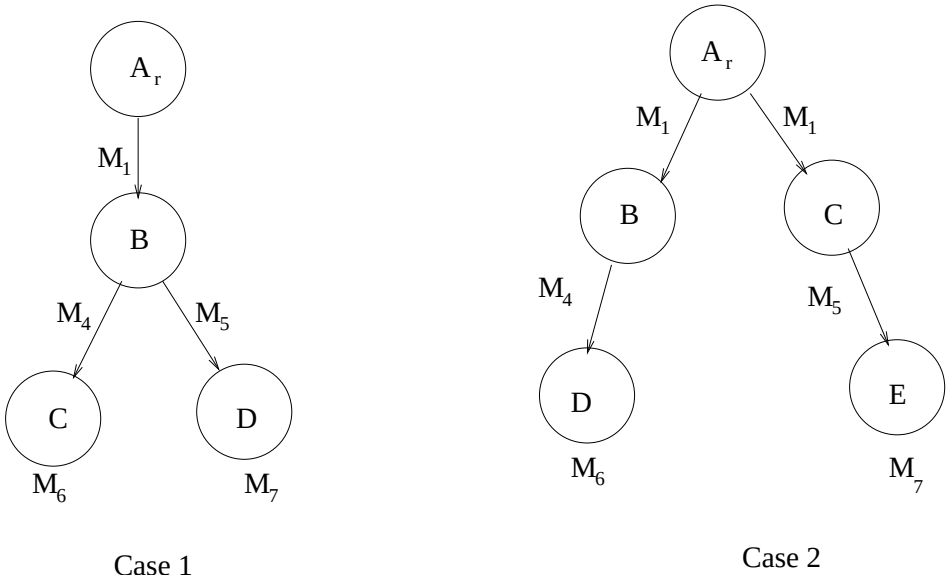


Figure 3.6: Nondistributivity of $\otimes$ over $\oplus$

Theorem 10. *The concatenation operator $\otimes$ distributes over the aggregation operator $\oplus$.*

Proof. We need to prove that $M_1 \otimes (M_2 \oplus M_3) = (M_1 \otimes M_2) \oplus (M_1 \otimes M_3)$. Assume the equivalent evidence corresponding to M_2 and M_3 is $\langle r_2, s_2 \rangle$ and $\langle r_3, s_3 \rangle$, respectively. Let $M_1 = \langle b_1, d_1, u_1 \rangle$. Then the equivalent evidence corresponding to $M_1 \otimes (M_2 \oplus M_3)$ is $\langle b_1(r_2 + r_3), b_1(s_2 + s_3) \rangle$. The evidence corresponding to $(M_1 \otimes M_2) \oplus (M_1 \otimes M_3)$ is $\langle b_1(r_2), b_1(s_2) \rangle \oplus \langle b_1(r_3), b_1(s_3) \rangle = \langle b_1(r_2 + r_3), b_1(s_2 + s_3) \rangle$, which completes the proof. □

Chapter 4

Updating Trust

4.1 Max-Certainty: Adaptive Trust Model

4.1.1 Evidence-Based Trust Model

This dissertation takes as its underlying model of trust the one proposed by W&S [Wang and Singh, 2006, 2007, Wang, 2004], which is based on a probability-certainty distribution function (PCDF) [Jøsang, 1998]. We provide a mechanism to update the trust values of agents, based on the referrals they provide (Section 4.1.2). Experimental results are shown by Hang *et al.* Chung-Wang-Singh-08,Chung-Wang-Singh-09

For example, an agent A estimates the trustworthiness of agent B by maintaining trust values of B. The trust value of B comes from both agent A's direct experience with agent B, and any referrals provided by a third-party, say, agent C. We model trust values in both the evidence space and the belief space. In evidence space, a trust value of an agent B is in the form $\langle r, s\rangle$, where $r+s > 0$. Here, $r \geq 0$ is the number of positive experiences with the agent B and $s \geq 0$ is the number of negative experiences with the agent B ($r + s \geq 0$). We define the probability $\alpha = \frac{r}{r+s}$, the expected value of the probability of a positive outcome. In belief space, a trust value is modeled as a triple $\langle b, d, u\rangle$, where $b, d, u > 0$ and $b + d + u = 1$. The values of $\langle b, d, u\rangle$ can be interpreted as the weights of belief, disbelief, and uncertainty, respectively. The certainty $c = 1 - u$ represents the confidence of the probability. A bijective trust transformation from evidence to belief space is defined in [Wang and Singh, 2007].

W&S provide aggregation and concatenation operators (similar to Jøsang's con-

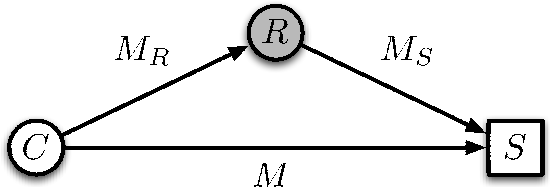


Figure 4.1: C calculates its trust in S by combining two parts

sensus and recommendation operators [Jøsang, 1998]) that enable agent A to combine the trust values of B from different sources (e.g., direct experience with B and a referral provided by agent C).

The concatenation operator $\otimes$ is used when agent A collects a referral of agent B from agent C. The aggregation operator $\oplus$ is used when agent A combines trust values $M_1 = \langle r_1, s_1 \rangle$ and $M_2 = \langle r_2, s_2 \rangle$ of B from different sources [Wang and Singh, 2006].

For example, in Figure 4.1, agent C estimates the trustworthiness of the target S by combining its direct experience M and the referral M_S from referrer R. The combined trust is $M \oplus (M_R \otimes M_S)$, where M_R is the trustworthiness of R.

4.1.2 Adaptive Trust Update Mechanism

A limitation of current trust models is that they lack support for trust update. Thus current systems perform trust updates in an ad hoc manner. The trustworthiness of an agent is the prediction of its future behavior, built on the experience of past interactions. However, for most trust models in referral networks, they do not provide appropriate trust update methods for the agents to update the trustworthiness of the referrers, although referrals are also a kind of interaction. In our model, the trustworthiness of the referrers is estimated based on how accurate their referrals are. The accuracy is determined by comparing the actual trustworthiness of the source and the referrals.

Trust Update Based on Max Certainty

Let the trustworthiness of R before the update is $\langle r_R, s_R \rangle$ in evidence space. Suppose the actual trust of the service provider S is $\langle r, s \rangle$, and R reported $\langle r', s' \rangle$. We define the trustworthiness of R after update, $\langle r'_R, s'_R \rangle$ by

$$q = \frac{\alpha'^r(1-\alpha')^s}{\alpha^r(1-\alpha)^s} \tag{4.1}$$

$$r'_R = \delta r_R + (1-\beta) r_R \tag{4.2}$$

$$s'_R = \delta s_R + (1-\beta) s_R \tag{4.3}$$

$$\delta r_R = c'q \tag{4.4}$$

$$\delta s_R = c'(1-q) \tag{4.5}$$

Here $\alpha = \frac{r}{(r+s)}$ and $\alpha' = \frac{r'}{(r'+s")}$. The accuracy q is the difference of C's aggregated trustworthiness $M = \langle r, s\rangle$ of S and the referral $M_S = \langle r', s'\rangle$ provided by agent R in the PCDF, as shown in Figure 4.1. Actually q represents how likely that the referral M' is correct. We will explain that in the following.

When agent C receives a report from an agent R, say, the quality of service provided by S is α', how does C interpret the report as good or bad? That depends on the true quality of the service provided by S. Instead of defining the correctness of a referral based on how likely α' represents the true quality provided by S, we define it to be how likely that the quality of the service provided by S is α'.

The probability that the quality of the service provided by S is p, is defined by the PCDF $\frac{p^r(1-p)^s}{\int_0^1 x^r(1-x)^s dx}$. This PCDF maximizes at α, which means that the quality of the service provided by S is most likely to be α. If we assume $\frac{\alpha^r(1-\alpha)^s}{\int_0^1 x^r(1-x)^s dx} = 1$, and if a referral said the quality of the service provided by S is α', then $q = \frac{\alpha'^r(1-\alpha')^s}{\int_0^1 x^r(1-x)^s dx}$, which is the probability that the quality of the service provided by S is α'. So q measures how likely the referral's report is to be correct.

Trust Update Based on Sensitivity

Though the update mechanism based on Max Certainty can tell how likely the referral tells the truth, it is not sensitive against malicious referral. It takes a long time for the agent to pinpoint a malicious referral. For example, if the actual trust of the service provider is $\langle 2, 1\rangle$, the referral reported $\langle 5, 5\rangle$. According to the formula above, $\langle \delta r_R, \delta S_R\rangle = \langle 0.37, 0.07\rangle$. If the reported trust is $\langle 1000, 1000\rangle$, then $\langle \delta r_R, \delta S_R\rangle = \langle 0.79, 0.15\rangle$. Since the effect of the bad report of $\langle 1000, 1000\rangle$ is very misleading and may overweight many good

reports, so the report of $\langle 1000, 1000\rangle$ should be considered a bad report. So we have the following update mechanism

Let the trustworthiness of R before the update is $\langle r_R, s_R\rangle$ in evidence space. Suppose the actual trust of the service provider S is $\langle r, s\rangle$, and R reported $\langle r', s'\rangle$. We define the trustworthiness of R after update, $\langle r'_R, s'_R\rangle$ by

$$
\begin{aligned}
q &= \frac{\alpha^{r'}(1-\alpha)^{s'}}{\alpha'^{r'}(1-\alpha')^{s'}} && (4.6)\\
r'_R &= \delta r_R + (1-\beta) r_R && (4.7)\\
s'_R &= \delta s_R + (1-\beta) s_R && (4.8)\\
\delta r_R &= c'q && (4.9)\\
\delta s_R &= c'(1-q) && (4.10)
\end{aligned}
$$

Here $\alpha = \frac{r}{(r+s)}$ and $\alpha' = \frac{r'}{(r'+s")}$. Actually q represents how likely that the referral thinks the quality of the service provider is α. In referral's view, the probability that the quality of the service provider is p, is defined by $l(p) = \frac{n(p)}{d}$, where $n(p) = p^{r'}(1-p)^{s'}$ and $d = \int_0^1 x^{r'}(1-x)^{s'}dx$.

$l(p)$ maximizes at α', which means that the referral thinks the quality of the service provided by S is most likely to be α'. If we normalize $l(\alpha')$ to be 1, then $\frac{\alpha^{r'}(1-\alpha)^{s'}}{\alpha'^{r'}(1-\alpha')^{s'}}$ measures how likely that the referral thinks the quality of the service provider is α.

Returning to the previous example, $\langle \delta r_R, \delta S_R\rangle = \langle 0.01, 0.93\rangle$, which means $\langle 1000, 1000\rangle$ is a bad report.

Trust Update Formula Based Average Accuracy with No Uncertainty

Both of the above mechanism has sensitivity problems when the number of transactions is too large. In the update formula based on max certainty, let the trust of the service provider is $\langle r, s\rangle$. When $r+s$ is very big, then the formula is very sensitive, since the PCDF will be at maximum when $x = \alpha = \frac{r}{r+s}$, but decreases very quickly when x deviates from α. For example, Let trust of the service provider is $\langle 800, 200\rangle$, when the agent report $\langle 19, 6\rangle$, it predicts that the quality of the service provider is 0.76. it is very close to the actual quality of the service provider which is 0.80, but the formula gives $\langle \delta r_R, \delta s_R\rangle = \langle 0.06, 0.58\rangle$ which indicates it is a very bad referral.

Similar problem happens for the update formula based on sensitivity when the number of transactions in the reported trust is large.

We we propose the following update mechanism based on average of accuracy.

Suppose the actual trustworthiness of the source is $(\alpha, 1-\alpha, 0)$. It is the case when we know for sure that the source can provide quality of service or referral at probability α, here the uncertainty is 0. Assume that the referred trust about the source is (b', d', u') which corresponds to the event $\langle r', s' \rangle$.

The updated trustworthiness $\langle r_R, s_R \rangle$ of R is defined as

$$
\begin{aligned}
q &= \frac{\int_0^1 x^{r'}(1-x)^{s'}|x-\alpha|dx}{\max(\alpha, 1-\alpha)\int_0^1 x^{r'}(1-x)^{s'}dx} && (4.11)\\
r'_R &= \delta r_R + (1-\beta) r_R && (4.12)\\
s'_R &= \delta s_R + (1-\beta) s_R && (4.13)\\
\delta r_R &= c'q && (4.14)\\
\delta s_R &= c'(1-q) && (4.15)
\end{aligned}
$$

Here $c' = c(r', s')$ is the certainty based on $langler', s' \rangle$. q measures how much error the referral is, if we look at the PCDF function corresponding to the referral $\langle r_R, s_R \rangle$. it's an average of accuracy of the referral. The reason we normalize it by $\max(\alpha, 1-\alpha)$ is that $\frac{\int_0^1 x^{r'}(1-x)^{s'}|x-\alpha|dx}{\int_0^1 x^{r'}(1-x)^{s'}dx}$ is between 0 and $\max(\alpha, 1-\alpha)$. After normalization, the error is between 0 and 1.

Theorem 11. *Let q be defined in 4.16. Then $0 \leq q \leq 1$.*

Theorem 12. *Let $(\langle r, s \rangle)$ be the referral*

Trust Update Formula Based on Average Accuracy with Uncertainty

When the trust of the service provider is a belief function with uncertainty, i.e. $M = (b, d, u)$ which corresponding to $\langle r, s \rangle$ in evidence space, we have the following formula

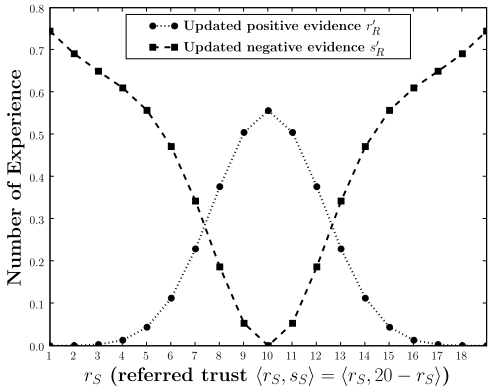


Figure 4.2: The new trust $M'_R = \langle r'_R, s'_R \rangle$ when referred trust is $M_S = \langle r_S, s_S \rangle$ where $1 \leq r_S \leq 20$ and $t_S = 20$, and actual trust is $M = \langle r, s \rangle = \langle 10, 10 \rangle$

$$q = \frac{\int_0^1 x^{r'}(1-x)^{s'}|x-\alpha|dx}{\max(\alpha, 1-\alpha)\int_0^1 x^{r'}(1-x)^{s'}dx} \tag{4.16}$$

$$r'_R = \delta r_R + (1-\beta) r_R \tag{4.17}$$

$$s'_R = \delta s_R + (1-\beta) s_R \tag{4.18}$$

$$\delta r_R = cc'q \tag{4.19}$$

$$\delta s_R = cc'(1-q) \tag{4.20}$$

Here $c = c(r, s)$ and $c' = c(r', s')$

the only difference is $\delta r_R = cc'qS$ and $\delta s_R = cc'(1-q)$, the reason we consider the certainty of $M = (b, d, u) = langler, s\rangle$ is that when we are not certain of the actual quality of the service provider, we are not certain to evaluate the referral, so we discount the update by extra term c.

Return to our previous example, let the trust of the service provider is $\langle 800, 200 \rangle$, when the agent reports $\langle 19, 6 \rangle$, the new formula gives $\langle \delta r_R, \delta s_R \rangle = \langle 0.53, 0.06 \rangle$ which indicates it is a very good referral.

4.1.3 Trust in History

To accommodate trust updates in a natural manner, we introduce the idea of *trust in history*. An agent's current level of trust in a service provider is treated as if it were a

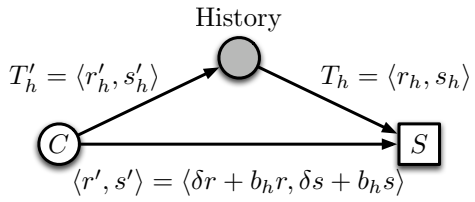


Figure 4.3: Illustration of trust in history

referral from the agent's past assessments of the provider. The idea is that the agent tries to estimate how much trust to place in the past. In this manner, we can avoid hard-coding a parameter, such as a discount factor, to determine the weight placed in the past. The trust in history gives an estimate of how much to weigh the past.

We formalize trust in history on par with other uses of trust, as $T_h = \langle r_h, s_h \rangle$. We can update our trust in history based on how accurate the transactions in history match our new transaction. So when a new service is received, T_h is updated as the following

$$r'_h = q + r_h \tag{4.21}$$

$$s'_h = 1 - q + s_h \tag{4.22}$$

Here $T'_h = \langle r'_h, s'_h \rangle$ is the updated trust in history; and q represents how close the new service quality is to the history. Recall that q is calculated in Section 4.1.2.

Let C's trust in S after new service be $\langle r', s' \rangle$. Then

$$r' = \delta r + b_h r \tag{4.23}$$

$$s' = \delta s + b_h s \tag{4.24}$$

Here $\langle b_h, d_h, u_h \rangle$ is equivalent to $\langle r_h, s_h \rangle$. The difference of our approach is that T_h is not a fixed discounting scalar; it is based on how much the history matches the subsequent transactions. If the provider's behavior changes a lot and cannot be predicted from the history, then the trust in history becomes low, and the history is discounted to a greater extent. Figure 4.3 illustrates the idea of trust in history.

Chapter 5

Certainty Based on Multivalued Events and Dempster-Shafer Theory

5.1 Introduction

We study the challenges of combining information from unreliable and potentially conflicting sources. Applications of this problem arise in multiagent systems, sensor networks, and trust and security. Dempster-Shafer (DS) theory provides an elegant basis for representing and manipulating uncertain information. However, despite DS theory's foundations in probability theory, its existing *fusion rules* are somewhat ad hoc, and yield counterintuitive results in circumstances where there are conflicts in the combined information or the certainty that one may place in it is low.

Sentz and Ferson [Sentz and Ferson, 2002a] argue that a fusion rule should satisfy four algebraic properties: *commutativity:* so that the fusion rule does not depend on the order in which information is received; *associativity:* to combine new information with the already combined old information [Yager, 1987]; *continuity:* meaning that if the result varies gradually if one of the inputs is changed gradually, so as to avoid sudden jumps; and *idempotence:* meaning that the result of two identical reports equals any of the reports.

Of these properties, we reject idempotence. Like most previous work on this topic, we make the closed-world assumption [Smets, 1990], under which combining information

from two independent sources may increase the total amount of information. For example, consider a box with 100 balls, each black or white. The probability that a randomly picked ball is white would vary from 0 to 1 depending on how many white balls there are in the box. Suppose Alice knows there are 10 white balls and 10 black balls, and has no information about the remaining 80 balls. Suppose Bob knows there are 10 white and 10 black balls, and does not know anything about the other 80 balls. If these two sets of information are completely independent, e.g., if Alice and Bob each put 20 balls into the box, the sets of balls they refer to are disjoint. Combining their information will yield the information that 20 black and 20 white balls are in the box, with the remaining 60 balls unknown.

We introduce a new property. A null belief, $\perp$, is totally uninformative (i.e., it has uncertainty 1.0). The property *null* says that any belief b that combines with $\perp$ yields b. For example, if Alice says the probability that Dell sells a good computer is 0.9 (with uncertainty 0.1), and Bob says he knows nothing about Dell, then the combined belief is the same as Alice's belief. Bob's null belief is ignored.

5.2 Certainty Based on Multivalued Events

5.2.1 Certainty vs Information: An example

Certainty measures how much information we have. Wang and Singh [Wang and Singh, 2007] show how to calculate certainty in a binary setting. In their example, a bin contains total n balls, each ball being either white or black. Let's estimate the probability that a randomly drawn ball is white. If we have complete information, i.e., that w balls are white and $n-w$ balls are black, then we know $\frac{w}{n}$ is the probability. Otherwise, we can't give a certain answer. However we can still give a reasonable answer with some certainty. Say we know that half of the balls are white, but don't know the color of the rest, which could include from 0 to $\frac{n}{2}$ white balls. That is, we only have half the possible information. Thus we can guess a probability between 0.5 and 1.0, but with certainty 0.5. Wang and Singh define certainty to be the percentage of total information. In this example, this equals $\frac{w+b}{n}$ where we know there are w white and b black balls in the bin, but have no information about the rest of the $(n-w-b)$ balls.

From the definition of certainty in the case of finitely many balls, Wang and Singh [Wang and Singh, 2007] define certainty in general, where the number of balls might

be infinity. Their certainty definition is based on the posterior certainty density function (PCDF).

Certainty defined for binary events has limited use, because many settings require multivalued events. But we need to clarify some notions first before we talk more about the certainty based on multivalued events.

Suppose there are a total of n balls in a bin, with four colors: pink, blue, green, and red. Again, the certainty we place in the probability that a randomly drawn ball is pink depends on what we know of the color distribution of some or all of the n balls. If we know there are p pink, b blue, g green, and r red balls, then we can define certainty as the percentage of information we have, or $\frac{p+b+g+r}{n}$. However, this is not adequate for multivalued settings, because the information on an event can be partial. Suppose we know that one ball is either pink or blue, and we don't know the colors of the other $n-1$ balls. The information we have in this case should be less than knowing the exact color of a ball.

We model multivalued situations by mapping them to the binary case. The usual intuitions from information theory apply here. One ball with four possible colors is equivalent (in information) with two balls, each of two possible colors. For example, we can map pink, blue, green, and red from the four-color setting to two balls in the binary setting. The original ball being pink means the corresponding two balls have the specified configuration of colors. The original ball being pink or blue means the corresponding two balls may be of two possible configurations. Similarly, in an eight-color setting, we can map one ball to three binary balls, and each of its colors to a configuration of the three binary balls.

We generalize the above discussion to m different colors. The knowledge of the color of one ball carries $\lg m$ bits of information ($\lg$ is $\log_2$). We can think of this as a unit of information in the given setting. When we have partial knowledge, i.e., that a ball can only be one of k colors where $1 \leq k \leq m$, then the information we have is $\lg m - \lg k$ bits, which equals $\frac{\lg m - \lg k}{\lg m} = \log_m \frac{m}{k}$ units of information.

5.2.2 Posteriori Probability Based on Multivalued Events

Let's introduce our key concepts and notation. Let $O = \{o_1, \cdots, o_m\}$ be the set of m possible mutually exclusive and exhaustive values. A multivalued event e is an m-tuple $\langle r_1, \cdots, r_m \rangle$. Let x_i represent the probability of the outcome with value o_i. Notice that $\sum_{i=1}^{m} x_i = 1$ or $x_m = 1 - \sum_{i=1}^{m-1} x_i$. We define the space of possibilities as $\Omega =$

$\{(x_1, \cdots, x_{m-1}) | \sum_{i=1}^{m-1} x_i \leq 1\}$. $V(\Omega)$, the volume of Ω, is $\frac{1}{(m-1)!}$, and is used below to define the prior.

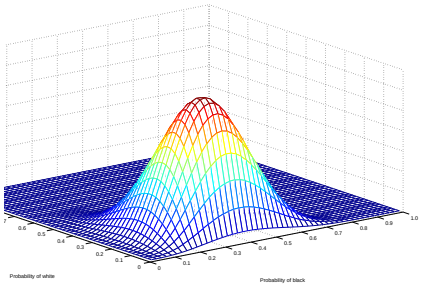
The probability-certainty distribution function (PCDF) of $x = (x_1, \cdots, x_{m-1})$ is a function $f : \Omega \to \Re$ where $\int_\Omega f \, dx = 1$. It doesn't need to be less than 1, but the integral of f over Ω should always be 1. We assume that the prior $f_\perp(x)$ is a uniform distribution, i.e., $f_\perp(x) = \frac{1}{V(\Omega)} = (m-1)!$, for any $x \in \Omega$.

The posterior distribution of $x = (x_1, \cdots, x_{m-1})$, $f_e(x) = f(x|e)$, given a multivalued event $e = \langle r_1, \cdots, r_m \rangle$ is a conditional distribution [Casella and Berger, 1990, p. 298]. Below, $g(e|x) = \delta \prod_{i=1}^{m-1} x_i^{r_i} (1 - \sum_{i=1}^{m-1} x_i)^{r_m}$.

$$\begin{aligned} f(x|e) &= \frac{g(e|x)f(x)}{\int_\Omega g(e|x)f(x)dx} \\ &= \frac{\delta(\prod_{i=1}^{m-1} x_i^{r_i}(1-\sum_{i=1}^{m-1} x_i)^{r_m}(m-1)!}{\int_\Omega \delta \prod_{i=1}^{m-1} x_i^{r_i}(1-\sum_{i=1}^{m-1} x_i)^{r_m}(m-1)!dx} \\ &= \frac{\prod_{i=1}^{m-1} x_i^{r_i}(1-\sum_{i=1}^{m-1} x_i)^{r_m}}{\int_\Omega \prod_{i=1}^{m-1} x_i^{r_i}(1-\sum_{i=1}^{m-1} x_i)^{r_m} dx} \end{aligned} \tag{5.1}$$

Notice that $\delta = \frac{(\sum_{i=1}^{m} r_i)!}{\prod_{i=1}^{m} r_i!} \neq 0$ appears in both the numerator and the denominator, so it cancels out from $f(x|e)$.

Figure 5.2.2 is an example of a three-dimension PCDF function. $f = f(x_1, x_2)$ where $x_1 + x_2 \leq 1$.



5.2.3 Certainty from a Multivalued PCDF

Wang and Singh [Wang and Singh, 2007] define certainty for binary events. We extend their definition to multivalued events. When more information is obtained, the

density function f deviates from $f_\perp$. To preserve $\int_\Omega f\ dx = 1$, f is greater than $f_\perp$ for some values and less than $f_\perp$ for other values. As in Wang and Singh, we take the mean absolute deviation of f and $f_\perp$ (dividing by 2 to normalize to 1).

Definition 18. *Let $f : \Omega \mapsto \Re$. The certainty based on f, c_f is given by $c_f = \frac{1}{2}\int_\Omega |f(x) - f_\perp| dx$*

Combining with (5.1), the certainty based on multivalued event $e = \langle r_1, \cdots, r_m \rangle$ is given by the following.

$$c_e = \frac{1}{2}\int_\Omega \left| \frac{\prod_{i=1}^{m-1} x_i^{r_i}(1-\sum_{i=1}^{m-1} x_i)^{r_m}}{\int_\Omega \prod_{i=1}^{m-1} x_i^{r_i}(1-\sum_{i=1}^{m-1} x_i)^{r_m} dx} - (m-1)! \right| dx \tag{5.2}$$

5.2.4 Fuzzy Multivalued Events

The foregoing only considers an observation with a specific value o_i where $1 \leq i \leq m$. However, a definite value may not be available. For example, suppose there are three possible values: blue, red, and pink. If an observation only tells you that a ball is not blue, and yields no information about if it is red or pink, then it is a *fuzzy* observation, since it has more than one possible value. So we extend the notion of multivalued events to *fuzzy multivalued* events.

Definition 19. *A fuzzy outcome A is defined to be a subset of O where $A \neq O$ and $A \neq \emptyset$.*

For example, $A = \{o_1, o_2\}$ means the outcome is either the value o_1 or o_2, but not any of the others. When $A = \{o_1\}$, it means the outcome has the value of o_1. The reason we require $A \neq O$ is that an outcome with O means the outcome could be any value, which does not provide any new information, and is the same as a null experiment.

Definition 20. *A fuzzy multivalued event e is a function $e : 2^O \backslash \{\emptyset, O\} \mapsto \Re^+ \bigcup \{0\}$, where $e(A) \geq 0$, and $\sum_{A \subset O, A \neq \emptyset} e(A) > 0$.*

That is, e maps any subset of O to a nonnegative real number. For example, consider drawing a ball from a bin with balls of any of three colors, i.e., $O = \{red, pink, blue\}$. An outcome $\{red, pink\}$ means that the ball drawn is either red or pink. Suppose $\hat{e}$ is an event where $\hat{e}(\{red\}) = r_1$, $\hat{e}(\{pink\}) = r_2$, $\hat{e}(\{blue\}) = r_3$, $\hat{e}(\{red, pink\}) = r_4$, $\hat{e}(\{red, blue\}) = 0$, and $\hat{e}(\{pink, blue\}) = 0$. Then $\hat{e}$ is an event with r_1 red balls, r_2 pink balls, r_3 blue balls, and r_4 either red or pink balls.

5.2.5 Certainty Based on Fuzzy Multivalued Events

To define certainty based on a fuzzy multivalued event, we transform the fuzzy multivalued event to a classical multivalued event, in which each outcome only has one possible value. Consider again the example of $\hat{e}$ above. As Section 5.2.1 explains, knowing one ball with color either red or blue is equivalent to $\log_3 \frac{3}{2}$ units of information. Since this information tells us that the ball is either red or blue, we assume the probability is uniformly distributed between these two colors. In other words, it is equivalent to saying that $\frac{\log_3 (3/2)}{2}$ balls are red and $\frac{\log_3 (3/2)}{2}$ balls are blue. So r_4 balls being red or blue is equivalent to $\frac{r_4}{2} \log_3 \frac{3}{2}$ balls being red and $\frac{r_4}{2} \log_3 \frac{3}{2}$ balls being blue. Thus we can transform the fuzzy multivalued event to a classical multivalued event $e = \langle r_1 + \frac{r_4}{2} \log_3 \frac{3}{2}, r_2 + \frac{r_4}{2} \log_3 \frac{3}{2}, r_3 \rangle$. For the general case, we have the following transformation.

Definition 21. *Let e be a fuzzy multivalued event. The equivalent classical multivalued event $e' = \langle r_1, \cdots, r_m \rangle$ is*

$$r_i = \sum_{A \subset O, A \neq \emptyset, o_i \in A} \frac{e(A)}{|A|} \log_m \frac{m}{|A|}, 1 \leq i \leq m$$

In order to understand the above formula, consider an outcome A that contains value o_i. A contributes $\log_m \frac{m}{|A|}$ unit information, which is uniformly distributed among $|A|$ possible values, and o_i is one of them. As Section 5.2.1 explains, A contributes $\frac{\log_m \frac{m}{|A|}}{|A|}$ to r_i, which is the total number of outcomes with value o_i. Since there are $e(A)$ outcomes A, the contribution to r_i is $\frac{e(A) \log_m \frac{m}{|A|}}{|A|}$.

Now we can simply define certainty based on the equivalent classical multivalued event, as given in (5.2).

5.3 DS Belief Space and Fuzzy Events

Below, let θ be *frame of discernment* or universe of discourse. θ is a set of mutually exclusive alternatives. Definitions 22 and 23 are from DS theory [Dempster, 1968, Shafer, 1976].

Definition 22. *A DS basic probability assignment (BPA) is a function $b : 2^\theta \mapsto [0, 1]$, where (1) $b(\emptyset) = 0$, and (2) $\sum_{A \subseteq \theta} b(A) = 1$*

Definition 23. *Let b_1 and b_2 be two independent BPAs on θ. The orthogonal sum of b_1 and b_2, $b = b_1 \oplus b_2$, is:*

$$b(\emptyset) = 0$$
$$b(A) = \frac{\sum_{X \cap Y = A} b_1(X) b_2(Y)}{1 - \sum_{X \cap Y = \emptyset} b_1(X) b_2(Y)}$$

5.4 Mapping Fuzzy Events to DS Belief Functions

We map fuzzy events to DS belief functions by translating fuzzy events to the crisp representation of the previous section. Taking O as the frame of discernment in DS theory, we can define a basic probability assignment on any $A \subset O$.

For the binary case, where $O = (\mathsf{B}, \mathsf{W})$, for each binary event $\langle r_\mathsf{B}, r_\mathsf{W} \rangle$, Wang and Singh [Wang and Singh, 2007] define the belief $\langle b(\mathsf{B}), b(\mathsf{W}), u \rangle$ by $b(\mathsf{B}) = c\frac{r_\mathsf{B}}{r_\mathsf{B}+r_\mathsf{W}}$, $b(\mathsf{W}) = c\frac{r_\mathsf{W}}{r_\mathsf{B}+r_\mathsf{W}}$, and $u = 1 - c(r_\mathsf{B}, r_\mathsf{W})$, where $c(r_\mathsf{B}, r_\mathsf{W})$ is the certainty based on $\langle r_\mathsf{B}, r_\mathsf{W} \rangle$. Here u is $b(O)$, the total uncertainty. Notice that the ratio of r_B and r_W is the same as the ratio of $b(\mathsf{B})$ and $b(\mathsf{W})$. Similarly, for fuzzy multivalued events, we can keep the ratio of r_A and r_B the same as for $b(A)$ and $b(\mathsf{B})$, for any $A, B \subset O$.

Definition 24. *We define the DS Belief Space as* $B = \{b | b : 2^O \mapsto [0,1], b(\emptyset) = 0, \sum_{A \subseteq O} b(A) = 1\}$

Definition 25. *We define the fuzzy multivalued evidence space as* $E = \{e | e : 2^O \setminus \{\emptyset, O\} \mapsto \Re^+ \bigcup \{0\},$ *where* $e(A) \geq 0$ *and* $\sum_{A \subset O, A \neq \emptyset} e(A) > 0\}$

Definition 26. *With B and E as in Definition 24 and 25, let function $Z : E \mapsto B$ be a map from fuzzy multivalued evidence space to DS Belief Space. Let $e \in E$ be represented as in Definition 20. When $A \subseteq O$, define $b = Z(e)$*

$$\begin{aligned}
r_i &= \textstyle\sum_{A \subset O, A \neq \emptyset, o_i \in A} \frac{e(A)}{|A|} \log_m \frac{m}{|A|}, \forall 1 \leq i \leq m \\
c_e &= \tfrac{1}{2} \textstyle\int_\Omega \left| \frac{\prod_{i=1}^{m-1} x_i^{r_i} (1 - \sum_{i=1}^{m-1} x_i)^{r_m}}{\int_\Omega \prod_{i=1}^{m-1} x_i^{r_i} (1 - \sum_{i=1}^{m-1} x_i)^{r_m} dx} - (m-1)! \right| dx \\
b(A) &= c_e \tfrac{e(A)}{\sum A \subset O, A \neq \emptyset e(A)} \\
b(O) &= 1 - c_e \\
b(\emptyset) &= 0
\end{aligned} \tag{5.3}$$

It is easy to see that Z is a total function: it yields a unique value for each e.

5.4.1 Key Mathematical Properties

We would like to have Z be a bijection between E and B. In order to prove Z is a bijection, we need to prove that for each $b \in B$, there is unique $e \in E$ such that $Z(e) = b$.

Let e be a fuzzy event. Then $\frac{e(A)}{e(B)}$ equals $\frac{b(A)}{b(B)}$, for any $A, B \subset O$ and $A, B \neq \emptyset$. The certainty c_e depends on the amount of evidence. If we assume $e(A) = b(A)t$ for $A \subset O, A \neq \emptyset$, then c_e is a function of t, i.e., $c(t)$.

$$\begin{aligned} b_i &= \textstyle\sum_{A, A \subset O, A \neq \emptyset, o_i \in A} \frac{b(A)}{|A|} \log_m \frac{m}{|A|}, 1 \leq i \leq m \\ c(t) &= \tfrac{1}{2} \textstyle\int_\Omega \left| \frac{\prod_{i=1}^{m-1} x_i^{b_i t} (1 - \sum_{i=1}^{m-1} x_i)^{b_m t}}{\int_\Omega \prod_{i=1}^{m-1} x_i^{b_i t} (1 - \sum_{i=1}^{m-1} x_i)^{b_m t} dx} - (m-1)! \right| dx \end{aligned} \tag{5.4}$$

We only need to show that there is a unique t such that $c(t) = 1 - b(O)$. Figure 5.1 plots how certainty varies with the total amount of evidence $r_1 + r_2 + r_3$, where the ratio $r_1 : r_2 : r_3$ is fixed at $1 : 2 : 3$. The trend is monotonic and asymptotic to 1.

Thus we have the following theorems. Theorem 13 shows that $c(t)$ is bounded. That is, the certainty is in $[0, 1]$ regardless of the amount of evidence.

Theorem 13. $0 < c(t) < 1$ *for any* $t > 0$.

Theorem 14 shows that $c(t)$ is monotonic with the amount of evidence provided that conflict in the data is held constant.

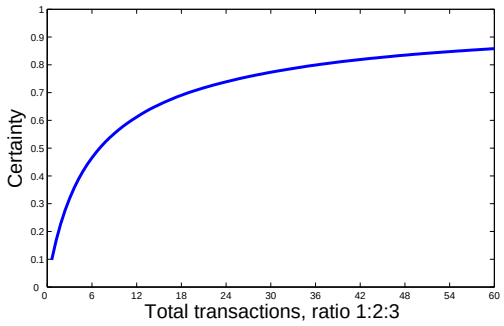


Figure 5.1: Certainty increases with evidence

Theorem 14. $c(t)$ *is strictly increasing,* $t > 0$.

Proof idea: The proof of this theorem is built via a series of steps. The main idea is to show that $c'(t) > 0$ for $t > 0$.

1. Eliminate the absolute sign by showing that there exists a region Ω_t such that $c(t)$ is an integral on Ω_t.

2. Differentiate $c(t)$ to obtain an integral on the boundary of Ω_t plus an integral on Ω_t.

3. Show that the integral on the boundary of Ω_t is 0.

4. Break the integral on Ω_t into two terms, with first term to be 0 by the symmetric property and the second term greater than 0. Hence $c'(t) > 0$.

Another important scenario is when the total amount of evidence is fixed, but the evidence varies to reflect different levels of conflict. Certainty should decrease where there is more conflict in the evidence. For the binary case, we can define conflict as ratio of positive and negative outcomes. For the multivalued case too, the conflict intuitively relates to the ratio of data. For example, for a three-dimensional event $\langle 9, 0, 0\rangle$ indicates no conflict, whereas $\langle 9, 9, 9\rangle$, means the conflict is maximal. Figure 5.2 plots certainty for $r_1 + r_2 + r_3 = 10$, and varying conflict.

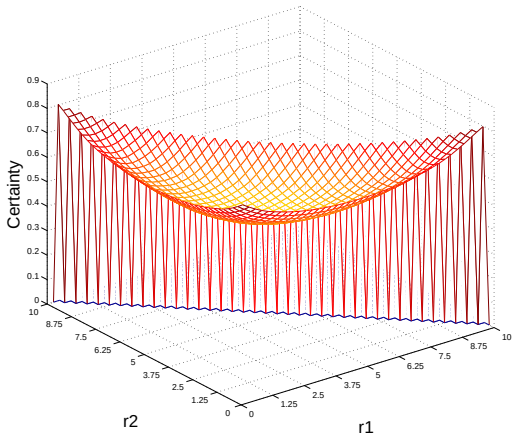


Figure 5.2: Certainty increases with evidence and decreases with conflict

Theorem 15. *$Z : E \mapsto B$ as in Definition 26 is a bijection.*

Proof sketch: From Definition 26, Z yields a unique value for each event in E. Conversely, to establish existence and uniqueness of Z^{-1}, observe that the evidence has the same ratios as the belief, and the size of the evidence depends on the certainty in the belief. For an amount of evidence, t, let $c(t)$ be as in Section 5.4.1. By Theorem 14, there is a unique t, corresponding to any certainty. By Theorem 13, t exists for any valid certainty (in $[0, 1]$). Thus, given b, it suffices to calculate the correct amount of evidence, t.

Let $c(t)$ be defined in equation 5.4. Algorithm 5.4.1 calculates Z^{-1} via binary search on $c(t)$ to a specified precision, $\epsilon > 0$. Here $t_{max} > 0$ is the maximum size of the body of evidence considered. (Recall that lg means logarithm to base 2.)

$\alpha = \frac{b}{b+d}$
$c = 1 - u$
$t_1 = 0$
$t_2 = t_{max}$
while $t_2 - t_1 \geq \epsilon$ **do**
    $t = \frac{t_1+t_2}{2}$;
    if $c(t) < c$ **then**
        $t_1 = t$
    else
        $t_2 = t$
    end if
end while
return t

5.5 Proposed Fusion Rules

Combining evidence is trivial. We just sum up the corresponding numbers. Let the evidence fusion rule be $\odot$.

Definition 27. *With E from Definition 20, let $e_1, e_2 \in E$. Let $e = e_1 \odot e_2$. Then $e(A) = e_1(A) + e_2(A)$ $\forall A \subset O, A \neq \emptyset$.*

Theorem 16. *The evidence fusion rule $\odot$ is commutative, associative, and continuous.*

We introduce $\oplus$, a new DS fusion rule. This rule maps the beliefs to evidence space, combines them in the evidence space, then maps the combined evidence back to DS belief space.

Definition 28. *Let B be as in Definition 24, and let $b_1, b_2 \in B$. Then $b_1 \oplus b_2 = Z(Z^{-1}(b_1) \odot Z^{-1}(b_2))$*

Since Z is a continuous bijection between E and B, we have the following theorem.

Theorem 17. *The belief fusion rule $\oplus$ is commutative, associative, and continuous.*

5.6 Zadeh's Example

The above properties provide a formal evaluation of the soundness of our approach. A more practical and conceptual evaluation involves showing that our approach produces different and better results than existing approaches. To this end, we compare our belief fusion rule with Dempster's rule and Jøsang and Pope's (J&P) [Jøsang and Pope, 2008] cumulative and averaging rules. We consider a series of examples to demonstrate that our approach produces intuitive results in a variety of settings whereas the above approaches are intuitive only in some cases. Below following Zadeh's [Zadeh, 1986] example, M, T, and C represent meningitis, tumor, and concussion, respectively.

J&P propose two fusion rules: *cumulative* and *averaging*. These rules rely on the bijection between Dirichlet distribution and belief functions. Since J&P treat each outcome equally and fail to consider conflict when calculating certainty, both rules yield unintuitive results in some cases.

Table 5.1: Zadeh's example using different approaches

	$m(M)$	$m(T)$	$m(C)$
Dempster's rule	0.000	1.00	0.000
J&P cumulative	0.495	0.01	0.495
J&P averaging	0.495	0.01	0.495
Our approach	0.495	0.01	0.495

Table 5.1 shows that our approach behaves the same as J&P's approaches on Zadeh's example, since both doctors have no uncertainty in their beliefs. Our approach and J&P's approaches produce the average of two beliefs.

Let's modify the example by introducing uncertainty to both doctors' beliefs, as J&P propose. Suppose Alice believes that Bob has either meningitis with probability 0.98 or a tumor with probability 0.01 (with uncertainty 0.01). Peter believes that Bob has either a concussion with probability 0.98, or a tumor with probability 0.01 (with uncertainty 0.01).

Table 5.2: Adding uncertainty to Zadeh's example

	$m(M)$	$m(T)$	$m(C)$	$m(\Theta)$
Dempster's rule	0.49000	0.01500	0.49000	0.0050
J&P cumulative	0.49246	0.01005	0.49246	0.0053
J&P averaging	0.49000	0.01000	0.49000	0.0100
Our approach	0.48390	0.00980	0.48390	0.0224

In Table 5.2, the $m(\Theta)$ column indicates uncertainty. Our rule gives a slight smaller probability for a tumor than each doctor's belief, since we take into account the conflict between the two beliefs. Our approach correctly captures the fact that conflict decreases certainty. J&P ignore the conflict when calculating certainty, so their cumulative rule yields slightly higher probability than both doctors' beliefs that Bob has a tumor. The averaging rule yields the same probability as each doctor's belief.

When uncertainty is higher, J&P's cumulative rule yields unintuitive results. Suppose Alice believes that Bob has either meningitis with probability 0.30, or tumor with probability 0.10 (thus, uncertainty is 0.60). Peter believes that Bob has either a concussion with probability 0.30, or a tumor with probability 0.10 (thus, uncertainty is 0.60).

As Table 5.3 shows, Dempster's and J&P's cumulative rules yield lower uncertainty than expected, and about 50% higher probability that Bob has a tumor. Since the two doctors provide contradictory information, even though they both claim that the probability that Bob has a tumor is 0.1, the combined belief should not be much higher than 0.1. When conflict is severe, the combined belief might be even lower, since we treat the information obtained as unreliable. Our rule yields a probability of 0.095, which is slightly below 0.10.

Table 5.3: Zadeh's example with high uncertainty

	$m(M)$	$m(T)$	$m(C)$	$m(\Theta)$
Dempster's rule	0.2118	0.1529	0.2118	0.4325
J&P cumulative	0.2143	0.1429	0.2143	0.4285
J&P averaging	0.1500	0.1000	0.1500	0.6000
Our approach	0.1448	0.0965	0.1448	0.6139

The uncertainty is 0.6139: slightly above 0.60 due to the conflict.

Table 5.4: When two beliefs are the same

	$m(M)$	$m(T)$	$m(C)$	$m(\Theta)$
Dempster's rule	0.2632	0.2632	0.2632	0.2104
J&P cumulative	0.2500	0.2500	0.2500	0.2500
J&P averaging	0.2000	0.2000	0.2000	0.4000
Our approach	0.2436	0.2436	0.2436	0.2692

J&P's averaging rule takes the average of the two beliefs. Though it is close to our intuition, the averaging rule does not make sense under the closed-world assumption, since averaging is idempotent, i.e., $b \oplus b = b$. If two doctors provide the same beliefs b, the averaging rule will yield the same belief b, whereas our intuition tells that the combined belief should have more certainty, or equivalently, less uncertainty. Table 5.4 illustrates an example. Suppose both Alice and Peter believe that Bob has a tumor with probability 0.2, meningitis with probability 0.2, concussion with probability 0.2, uncertainty 0.4. Under the closed-world assumption no other diseases are possible [Smets, 1990]. Our approach, Dempster's rule, and J&P's cumulative rule all yield more certainty in the fused beliefs, but J&P's averaging rule keeps the certainty the same.

Table 5.5 illustrates another shortcoming of J&P's averaging rule. Suppose Alice believes that Bob has a tumor with probability 0.5 (and uncertainty 0.5). Peter provides no information, i.e., his belief is $\perp$ with uncertainty 1.0. J&P's averaging rule yields a highly unreasonable result, namely, that Bob has a tumor with probability 0.33 and uncertainty

Table 5.5: Example where one of beliefs is null

	$m(M)$	$m(T)$	$m(C)$	$m(\Theta)$
Dempster's rule	0.00	0.50	0.00	0.50
J&P cumulative	0.00	0.50	0.00	0.50
J&P averaging	0.00	0.33	0.00	0.67
Our approach	0.00	0.50	0.00	0.50

0.67. This contrasts with the *null* property that Section 5.1 introduces: we should have $b \oplus \perp = b$. Our approach, Dempster's rule, and J&P's cumulative rule all obey this property.

Chapter 6

Discussion

This dissertation offers a theoretical development of trust that would underlie a variety of situations where trust reports based on evidence are combined. In particular, it contributes to a mathematical understanding of trust, especially as it underlies a variety of multiagent applications. These include social networks understood via referral systems and webs of trust, in studying which we identified the need for this research. Such applications require a natural treatment of composition and discounting in an evidence-based framework. As Chapter 1 shows, these applications broaden to service-oriented computing in general.

Further, an evidence-based notion of trust must support important properties regarding the effects of increasing evidence (for constant conflict) and of increasing conflict (for constant evidence). Theoretical validation, as provided here, is valuable for a general-purpose conceptually driven mathematical approach such as ours. The main technical insight of this dissertation is how to manage the duality between trust and evidence spaces in a manner that provides a rigorous basis for combining trust reports.

A payoff of this approach is that an agent who wishes to achieve a specific level of certainty can compute how much evidence would be needed at different levels of conflict. Or, the agent can iteratively compute certainty to see if the certainty of its beliefs or disbeliefs has reached an acceptably high level.

It is worth considering briefly the benefits of treating trust as a well-defined concept. Potentially, instead of exchanging trust reports, the agents could exchange probability distributions based upon the evidence. However, discounting such evidence would require going through the trust report representation that we described. Because of the bijection

that Theorem 3 establishes, the trust and evidence representations are equivalent, so the choice between them is arbitrary. However, trust is an important concept for both conceptual and practical reasons. In conceptual terms, trust represents a form of *relational* capital [Castelfranchi et al., 2006] among the agents. From a practical standpoint, trust summarizes the prospects for an interaction in way that makes intuitive sense to people and fits into practical agent architectures. Making intuitive sense to people is crucial from the standpoint of effective requirements elicitation and for explaining outcomes to users. Moreover, in open architectures where the agents are implemented heterogeneously, a numeric treatment of trust can provide a simple means of facilitating inter-operation without requiring that the implementations agree on their internal representations.

The broad family of applications that this approach targets includes social networks and service-oriented computing. These applications rely on the parties concerned acquiring evidence in order to make reasoned judgments about interacting with others. As a practical matter, it is inevitable that there will be conflicts in the trust reports received from others. There is agreement that certainty is crucial in combining trust reports. However, previous approaches calculate certainty naïvely in a manner that disregards conflict. Thus our results are a significant advance even though our approach begins from the same PCDF framework as applied by Jøsang in his treatment of trust.

This dissertation treats reputation management as a well-defined path algebra problem. Each agent only needs to maintain the trust ratings for its neighbors and to use referral networks and path algebra to merge and combine the trust ratings and obtain the trust in the service provider with which it has no direct interactions. Our algorithm scales well, since the trust ratings are merged in a bottom up fashion.

The underlying notion of a reference associated with a trust rating is extremely general. It accommodates referrals, neighborhood relationships in peer-to-peer computing, symbolic references across information resources, and potentially even physical neighborhood relationships such as on certain kinds of ad hoc networks.

This dissertation proposes 3 different mechanisms to update referrers' trustworthiness. Further, the discounting rate parameter offers the flexibility to capture different dynamic behavior of agents.

Further directions include further experimental evaluation. More different types of agents, say, discriminative agents, should be also considered in our future work. Besides,

the decision-making strategies for agents to compare trust estimates for service selection should be also discussed.

The broad family of applications that this approach targets includes social networks and service-oriented computing. These applications rely on the parties concerned acquiring evidence in order to make reasoned judgments about interacting with others. As a practical matter, it is inevitable that there will be conflicts in the trust reports received from others. There is agreement that certainty is crucial in combining trust reports. However, previous approaches calculate certainty naïvely in a manner that disregards conflict. Thus our results are a significant advance even though our approach begins from the same PCDF framework as applied by Jøsang in his treatment of trust.

Chapter 7

Literature

A huge amount of research has been conducted on trust. We now review some of the most relevant literature from our perspective of an evidential approach.

7.1 Distributed Trust

Trust models are widely studied in various domains. Trust propagation methods can be categorized as the models that consider the witness information sources [Sabater and Sierra, 2005], or retrieve and aggregate ratings from the social network [Ramchurn et al., 2004]. Here, we focus on trust propagation approaches not covered in above surveys.

Richardson *et al.*[Richardson et al., 2003] discuss an abstract framework for trust propagation. Each user maintains trust in a small number of other users. A user's trust in any other user can be computed by using the existing web of trust recursively. Richardson *et al.* first enumerate all paths between the user and every other user who has a local belief in a given statement. Next they calculate the belief associated with each path by using a concatenation operator along each path, and combine the beliefs associated with all paths using a predefined aggregation operator.

Yu and Singh [Yu and Singh, 2002, 2003b] study distributed reputation management in a social network whose member agents cooperate with each other to find the trustworthiness of other agents. When an agent wants to find the trustworthiness of a service provider, it uses the referral network to find witnesses, and combines the beliefs of

those witnesses regarding the service provider. Yu and Singh's approach is based on the Dempster-Shafer theory of evidence. It treats a service with medium quality as of unknown quality, not as a known medium quality. More importantly, it does not completely address double counting in that multiple paths through the same agent are treated as if they were independent. Although the approach identifies witnesses who have direct experience, so there is no double counting at that level.

Gray *et al.* [Gray et al., 2003] solve trust propagation problem in mobile ad hoc networks. In their approach, the concatenation is calculated as the average of trust of each edge along the referral path, where the trust of edges is discounted by the depth in the referral path. Trust from different paths is aggregated by choosing the most trusted of the available paths. Quercia *et al.*
[Quercia et al., 2007] also propagates trust in mobile network context. Based on the web of trust, they build *relationship graphs* where nodes are relationships. Two *related* nodes are linked if they are either from the same rater or rating the same person. At the beginning, some of nodes are rated at the first place. Then trust is propagated from rated nodes to unrated ones by computing a *predictive* function. However, the number of relationships in real-world networks is much larger than the number of nodes. For example, there are $1,234$ edges (538 nodes) in FilmTrust, and $317,979$ edges ($39,246$ nodes) in PGP. Propagating trust in the corresponding relationship graphs would be more computationally expensive.

Trust management in peer-to-peer systems have been also widely studied
[Kamvar et al., 2003]
[Zhou et al., 2008]. Spectral decomposition is used on the adjacency matrix of the network graph to estimate global reputation, which is also called the *global group* trust metric [Ziegler and Lausen, 2004].
Guha *et al.*[Guha et al., 2004] associate trust relations to matrix operations. For example, the commutativity of trust is associated to matrix transpose, while direct propagation is associated to matrix multiplication. Different from global group methods, our approach models trust from a personal perspective. Propagated trust takes into account personal bias. Besides, matrix approaches require several iterations to converge.

Advogate [Levien, 2003] and Appleseed [Ziegler and Lausen, 2004] are two *local group* trust metrics, which propagate global trust only on the local subgraph. Advogate applies network flow on a modified graph, where capacities are assigned to edges based on

the depth in the referral path. The deeper the edge is the less capacity it has. Conversely, Appleseed adopts *spreading activation*. It spreads *energy* across the graph, and, when propagating through a node, divides energy among successors based on the edge weights. The idea of Appleseed is similar to the spectral decomposition, which requires several iterations to converge. CertProp propagates trust in local subgraphs. It provides *local* trust, rather than group-level (local group) or global trust. In other words, the propagated trust of Alice from Bob is different from the trust of Alice from Charlie. Richardson *et al.* [Richardson et al., 2003] have studied the trust management for the semantic web. Each user maintains trust in a small number of other users. A user's trust in any other user can be computed by using the existing web of trust recursively. Richardson *et al.* first enumerate all paths between the user and every other user who has a local based belief in a given statement, then calculate the belief associated with each path by using a predefined *concatenation function* along each path and the belief held by the final user, and those beliefs associated with all paths can be combined by using a predefined *aggregation function*. Since Richardson *et al.* use the path algebra interpretation on the whole network, it is not appropriate in a distributed network where agents in the network are autonomous, since the topology may change constantly and some agents may not be cooperative.

Yu and Singh [Yu and Singh, 2002], [Yu and Singh, 2003a] have studied distributed reputation management in a social network of agents where the agents cooperate with each other in finding trustworthiness of the other agents. Yu and Singh build their work on referral networks. They use Dempster-Shafer theory to represent the agent's trust on the service provider. When an agent wants to find the trustworthiness of a service provider, it uses the referral network to find witnesses, and combines the beliefs of those witnesses regarding the service provider. Yu and Singh assign weights to each witness to detect and penalize deceptive agents. There are some limitations to this approach. First, a service with medium quality is treated as unknown quality, not as a known medium quality. Second, suppose some witnesses are found. Let's consider two cases. In one case, all witnesses are referred by the same agent, while in the other case, all witnesses are referred by different agents. We should expect more evidence from the second case. By assigning weights to each witness these two cases are not differentiated.

Despotovic and Aberer [Despotovic and Aberer, 2005] propose a simple probabilistic method, maximum likelihood estimation, to consolidate ratings. Though it reduces the

calculation overhead and admits a clear statistical interpretation, however, their model is overly simplified. The agents rate a target as either good or bad, Thus their method cannot be used where the agents need to give precise ratings, secondly, their method does not consider the uncertainty of a rating, or equivalently, the number of transactions on which a rating might be based. Since only a small number of witnesses are chosen and each agent only knows a small number of all agents, most of the time, the agent cannot compute the trust to the necessary witnesses, which causes problem to scale up.

Huynh *et al.* [Huynh et al., 2004] introduced a trust model, FIRE, which has four components: interaction trust, role-based trust, witness reputation, and certified reputation. FIRE incorporates all those components to provide a combined trust. Referrals were used to obtain the witnesses, each witness is assigned a weight. But we have not seen any work on the biased information referred by the malicious acquaintances. Further, it does not scale well since a weight for each witness needs to be maintained.

Jøsang *et al.* [Jøsang et al., 2003] analyzed transitive trust topologies, specified three basic topology dimensions: trust origin, trust target and trust purpose. They also described principles for recommendation, measuring and computing trust based on those topologies. Whereas detailed mathematical computation formula for computation of the combined trust has not been found in his work. Jøsang *et al.* used cryptography to provide authenticity and integrity of referrals, but key management is still a major and largely unsolved problem on the Internet today.

Carbone *et al.* [Carbone et al., 2003] study trust formally in a setting based on distributed computing systems. They propose a two-dimensional representation of trust consisting of (1) trustworthiness and (2) certainty placed in the trustworthiness. Carbone *et al.* introduce the notion of ordering for each dimension of the trust. But they do not specify how to calculate the certainty or the properties of it.

Weeks [Weeks, 2001] introduces a mathematical framework for distributed trust management systems. He uses the least fixed point in a lattice to define the semantics of trust computations. Importantly, Weeks only deals with the so-called hard trust among agents that underlies traditional authorization and access control approaches. He doesn't deal with evidential trust, as studied in this dissertation.

Quercia *et al.* [Quercia et al., 2007] design a system to propagate trust among mobile users. They use a graph-based learning technique. Quercia *et al.*'s model is lightweight,

distributed and efficient for mobile users. However, their approach only uses a very simple similarity between agents to predict the trust, it is not based on evidence.

Ziegler and Lausen [Ziegler and Lausen, 2004] classify trust metrics among various axes: global, local, centralized, distributed, group, and scalar trust. They apply the idea of Appleseed to activate and distribute energy among the nodes. The connection between energy and trust is not justified. In their work, trust is subjective, which is only based on agent's opinions instead of evidence.

7.2 Trust and Evidence

Abdul-Rahman and Hailes [Abdul-Rahman and Hailes, 2000] present an early model for computing trust. However, their approach is highly ad hoc and limited. Specifically, various weights are simply added up without any mathematical justification. Likewise, the term *uncertainty* is described but without being given any mathematical foundation.

Sierra and Debenham [Sierra and Debenham, 2007] define an agent strategy by combining the three dimensions of utility, information, and semantic views. Their framework defines trust, reputation, and uncertainty. Their definition is rather complex. It is justified based on the authors' intuitions and is experimentally evaluated. Thus it is plausible in a conceptual way. However, it lacks an explicit mathematical justification, such as we have sought to develop in this work.

The Regret system combines several aspects of trust, notably the social aspects [Sabater and Sierra, 2002]. It involves a number of formulas, which are given intuitive, but not mathematical, justification. A lot of other work, e.g.,
[Huynh et al., 2006], involves heuristics that combine multiple information sources to judge trust. It would be an interesting direction to combine a rigorous approach such as ours with the above heuristic approaches to capture a rich variety of practical criteria well.

Teacy *et al.* [Teacy et al., 2005] proposed TRAVOS, the Trust and Reputation model for Agent-based Virtual OrganisationS. TRAVOS uses a probabilistic treatment of trust. Teacy *et al.* model trust in terms of confidence that the expected value lies within a specified error tolerance. An agent's confidence increases with the error tolerance. Teacy *et al.* study combinations of probability distributions to which the evaluations given by different agents might correspond. They do not formally study certainty. Further, Teacy *et*

al.'s approach does not yield a probabilistically valid method for combining trust reports, which is the focus of this dissertation.

Despotovic and Aberer [Despotovic and Aberer, 2005] propose a simple probabilistic method, maximum likelihood estimation, to aggregate ratings. This method dramatically reduces the calculation overhead of propagating and aggregating trust information. Further, the aggregated trust admits a clear statistical interpretation. However, Despotovic and Aberer's model is overly simplified: the agents rate a target as either good or bad. Thus their approach cannot be used where the agents are required to give more accurate ratings, e.g., even a scalar (as a real value) from 0 to 1. Further, Despotovic and Aberer's method does not consider the uncertainty of a rating or equivalently the number of transactions on which a rating might be based. Since witnesses can be any agents, in order to estimate the maximum likelihood, each agent needs to record the trustworthiness of all possible witnesses, thus increasing the complexity of scaling up. More importantly, since only a small number of witnesses are chosen and each agent only knows a small number of all agents, most of the time, the agent cannot compute how much trust to place in the necessary witnesses.

Reece *et al.* [Reece et al., 2007] develop a method to consolidate an agent's direct experience with a service provider and trust reports about that service provider received from other agents. Reece *et al.* calculate a covariance matrix based on the Dirichlet distribution that describes the uncertainty and correlations between different dimensional probabilities. The covariance matrix can be used to communicate and fuse ratings. The Dirichlet distribution considers only the ratio of positive and negative transactions. It does not depend on the number of transactions, so the resulting uncertainty or certainty estimates are independent of the total number of transactions. As we explained above, this is contrary to intuition because certainty does increase with mounting evidence if the ratio of positive and negative transactions is kept constant. Lastly, Reece *et al.* neglect the trustworthiness of the agent who provides the information. This dissertation provides a basis for accommodating the trustworthiness of agents who provide information: this aspect is studied by Wang and Singh [Wang and Singh, 2006], which applies the present approach to specify operators for propagating trust.

Halpern and Pucella [Halpern and Pucella, 2006] consider evidence as an operator that maps prior beliefs to posterior beliefs. Similar to our certainty function, they use a

confirmation function to measure the strength of the evidence. However, there are many kinds of confirmation functions available, and it is not clear which one to use. Halpern and Pucella use the log-likelihood ratio. They do not give a mathematical justification for its optimality, only that it avoids requiring a prior distribution on hypotheses.

Fullam and Barber [Fullam and Barber:, 2006] describe trust-related decisions based on agent role (trustee or truster) and transactions (fundamental transaction or reputation transaction). They propose applying Q-learning and explain why the learning is complicated when reputation transaction is used. Fullam and Barber use the Agent Reputation and Trust (ART) test-bed to evaluate their learning techniques. Fullam and Barber [Fullam and Barber, 2007] study different sources of trust information: direct experience, referrals from peers, and reports from third parties. They propose a dynamical learning technique to identify the best sources of trust, based on some parameters. Both the above works do not consider the uncertainty of trust information and do not offer any mathematical justification for their approach.

The following work is not closely related to our work but worth discussing because it deals with service discovery based on uncertainty. Li *et al.*
[Li et al., 2008] describe ROSSE, a search engine for grid service discovery. They introduce the notion of "property uncertainty" when matching services. A property is uncertain when it is used in some but not all advertisements for services in the same category. Thus, for Li *et al.*, uncertainty means how unlikely a service has the property in question. This is quite different from our meaning based on the probability of the probability of a particular outcome. Li *et al.* use rough set theory to deal with property uncertainty and select the best matched services.

Other approaches study systems in which agents alter their relationships or their behaviors based on calculations of each other's trustworthiness. Jurca and Faltings
[Jurca and Faltings, 2007] describe a mechanism that uses side-payment schemes to provide incentives for agents so that it becomes rational for the agents to report ratings of other agents truthfully. Jurca and Faltings use a simplistic trust model. They express trust as a scalar from 0 to 1, and do not consider uncertainty. As a result, one bad experience out of ten yields the same trust as 1,000 bad experiences out of 10,000. By contrast, our approach finds the two cases to yield different measures of certainty. Overall, though, our approach is complementary to theirs. Jurca and Faltings are interested in obtaining individual ratings;

we are interested in aggregating the ratings into measures of belief of certainty, which can then be propagated [Wang and Singh, 2006].

Sen and Sajja [Sen and Sajja, 2002] also address the problem of deceptive agents. They study reputation-based trust with an emphasis on the problem of estimating the number of raters (some of whom may be liars) to query in order to obtain a desired likelihood threshold about the quality of a service provider. They model the agents' performance as drawn from two distributions (high and low); agents use reputation to determine if some of the raters are liars. Sen and Sajja experimentally study the effect on performance of systems wherein some of the raters are liars with a view to identifying thresholds beyond which the number of liars can disrupt a system. Yu and Singh [Yu and Singh, 2003a] show how agents may adaptively detect deceptive agents. Yolum and Singh [Yolum and Singh, 2003] study the emergent graph-theoretic properties of referral systems. This dissertation complements such works because it provides an analytical treatment of trust that they do not have whereas they address system concerns that this dissertation does not study.

7.3 Information Theory and Dempster-Shafer Theory

Shannon entropy [Shannon, 1948] is the best known information-theoretic measure of uncertainty. It is based on a discrete probability distribution $p = \langle p(x) | x \in X \rangle$ over a finite set X of alternatives (elementary events). Shannon's formula encodes the number of bits required to obtain certainty: $S(p) = -\sum_{x \in X} p(x) \lg p(x)$. Here $S(p)$ can be viewed as the weighted average of the conflict among the evidential claims expressed by p. Jaynes [Jaynes, 2003] provides examples, intuitions, and precise mathematical treatment of the entropy principle. More complex, but less well-established, definitions of entropy have been proposed for continuous distributions as well, e.g., [Smith, 2001].

Entropy, however, is not suitable for the present purposes of modeling evidential trust. Entropy captures (bits of) missing information and ranges from 0 to ∞. At one level, this disagrees with our intuition that, for the purposes of trust, we need to model the confidence placed in a probability estimation. Moreover, the above definitions cannot be used in measuring the uncertainty of the probability estimation based on past positive and negative experiences.

Shafer [Shafer, 1976] introduces *tradeoff* combination rule. It first discounts the

information according to information source, then combines them. The weighted average of the beliefs is the combined belief. Though it is continuous and commutative, it is not associative. Shafer does not consider volume of evidence behind the beliefs or the subtlety of the conflict in the information. The calculation of discount parameter is ad hoc at best.

Yager [Yager, 1987] attributes conflict to the universal set X through the conversion of basic probability assignment to ground probability assignment. The combination rule is continuous, commutative, but not associative. It does not consider body of evidence underlying the beliefs.

Inagaki [Inagaki, 2008] defines a continuous parameterized class of combination rules, which subsumes both Dempster's rule and Yager's rule. His rule is commutative but not associative. Inagaki wrongly claims that his rule subsumes all combination rules. He does not realize that he only considers linear combination rules where weights of averaging only depend on beliefs linearly. He does not consider the body of evidence either. As we showed above, the body of evidence depends on evidence nonlinearly.

Ferson and Kreinovich [Ferson and Kreinovich, 2002] introduce a generalized averaging rule. It slightly differs from Shafer's *tradeoff* method by the definition of weights. It is continuous, commutative, but not associative. Since their method is linear, it does not capture the body of evidence or the extent of conflict contained in the beliefs.

Jøsang and Diaz [Jøsang and Elouedi, 2007] define certainty based on dirichlet distribution for multivalued events. They argue that certainty increase as evidence increase. But there are two major problems with their approach. First, they map each outcome to each singleton in the dirichlet state space, thus treating all outcomes equally. Secondly, they define certainty to be a function of the total amount of evidence, without considering conflict.

Jøsang and Pope[Jøsang and Pope, 2008] propose two fusion rules: cumulative fusion of evidence and averaging fusion rule. They all based on bijection between dirichlet distribution and belief distribution functions. Since they treat each outcome equally and fail to consider conflict when calculate certainty, their rules yields counterintuitive results.

Bibliography

Alfarez Abdul-Rahman and Stephen Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on Systems Science*, Los Alamitos, 2000. IEEE Computer Society Press.

Karl Aberer and Zoran Despotovic. Managing trust in a peer-2-peer information system. *CIKM*, pages 310–317, 2001.

K. Suzanne Barber and Joonoo Kim. Belief revision process based on trust: Agents evaluating reputation of information sources. In *Proceedings of the workshop on Deception, Fraud, and Trust in Agent Societies held during the Autonomous Agents Conference*, pages 73–82, London, UK, 2001. Springer-Verlag.

John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, pages 43–52, 1998.

Marco Carbone, Mogens Nielsen, and Vladimiro Sassone:. Formal model for trust in dynamic networks. In *Proceedings of 1st International Conference on Software Engineering and Formal Methods (SEFM 2003),*, pages 54–63, Brisbane, Australia, 2003. IEEE Computer Society.

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, 1990.

Cristiano Castelfranchi and Rino Falcone. Principles of trust for MAS: cognitive anatomy, social importance, and quantification. In *Proceedings of the 3rd International Conference on Multiagent Systems*, pages 72–79, 1998.

Cristiano Castelfranchi, Rino Falcone, and Francesca Marzo. Being trusted in a social network: Trust as relational capital. In *Trust Management: Proceedings of the iTrust Workshop*, volume 3986 of *LNCS*, pages 19–32, Berlin, 2006. Springer.

Chris Dellarocas. Online reputation mechanisms. chapter 20. Chapman Hall & CRC Press, Baton Rouge, 2004.

Arthur P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society*, 40:205–247, 1968.

Zoran Despotovic and Karl Aberer. Probabilistic prediction of peers' performances in P2P networks. *International Journal of Engineering Applications of Artificial Intelligence*, 18 (7):771–780, 2005.

Scott Ferson and Vladik Kreinovich. Representation, propagation, and aggregation of uncertainty, 2002. Sandia National Laboratories.

Karen Fullam and K. Suzanne Barber:. Learning trust strategies in reputation exchange networks. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1241–1248, New York, 2006. ACM Press.

Karen Fullam and K. Suzanne Barber. Dynamically learning sources of trust information: Experience vs. reputation. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1062–1069, Columbia, SC, 2007. IFAAMAS.

Mario Gómez, Javier Carbó, and Clara Benac Earle. Honesty and trust revisited: the advantages of being neutral about other's cognitive models. *Autonomous Agents and Multi-Agent Systems*, 15(3):313–335, December 2007.

Elizabeth Gray, , Jean-Marc Seigneur, Yong Chen, and Christian Jensen. Trust propagation in small worlds. *Lecture Notes in Computer Science*, 2692:239–254, 2003.

R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412. ACM Press, 2004.

Joseph Y. Halpern and Reccardo Pucella. A logic for reasoning about evidence. *Journal of AI Research*, 26:1–34, 2006.

Maaike Harbers, Rineke Verbrugge, Carles Sierra, and John Debenham. The examination of an information-based approach to trust. In *MALLOW Workshop on Coordination, Organization, Institutions and Norms in agent systems (COIN)*, September 2007.

Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. Developing an integrated trust and reputation model for open multi-agent systems. In *Proceedings of the Workshop on Trust in Agent Societies at AAMAS*, 2004.

Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and MultiAgent Systems*, 13(2):119–154, September 2006.

Toshiyuki Inagaki. Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory. *IEEE Transactions on Reliability*, 40:182–188, 2008.

Edwin Thompson Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

Audun Jøsang. A logic for uncertain probabilities. *Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:279–311, 2001.

Audun Jøsang. A subjective metric of authentication. In *Proceedings of the 5th European Symposium on Research in Computer Security (ESORICS)*, 1998.

Audun Jøsang and Zied Elouedi. Interpreting belief functions as Dirichlet distributions. In *Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, Hammamet, Tunisia, 2007.

Audun Jøsang and Simon Pope. Dempster's rule as seen by little coloured balls. *Artificial Intelligence*, 2008. Preprint.

Audun Jøsang, Elizabeth Gray, and Michael Kinateder. Analysing topologies of transitive trust. In *Proceedings of the Workshop of Formal Aspects of Security and Trust (FAST)*, 2003.

Rudu Jurca and Boi Faltings. Obtaining reliable feedback for sanctioning reputation mechanisms. *Journal of Artificial Intelligence Research (JAIR)*, 29:391–419, 2007.

Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The EigenTrust algorithm for reputation management in p2p networks. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 640–651. ACM Press, 2003.

Raph Levien. *Attack Resistant Trust Metrics*. PhD thesis, UC Berkeley, 2003.

Maozhen Li, Bin Yu, Omer Rana, and Zidong Wang. Grid service discovery with rough sets. *IEEE Transactions on Knowledge and Data Engineering*, 20(6):851–862, 2008.

E. Michael Maximilien. *Toward Autonomic Web Services Trust and Selection*. PhD thesis, Department of Computer Science, North Carolina State University, 2004.

Daniele Quercia, Stephen Hailes, and Licia Capra. Lightweight distributed trust propagation. In *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM)*, pages 282–291, Omaha, NE, 2007. IEEE Computer Society.

Sarvapali D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *Knowledge Engineering Review*, 19(1):1–25, 2004.

Steven Reece, Alex Rogers, Stephen Roberts, and Nicholas R. Jennings. Rumours and reputation: Evaluating multi-dimensional trust within a decentralised reputation system. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1063–1070, 2007.

Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186. ACM Press, 1994.

Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web. In *Proceedings of the Second International Semantic Web Conference*, 2003.

Eric Sven Ristad. A natural law of succession. TR 495-95, Department of Computer Science, Princeton University, July 1995.

Jordi Sabater and Carles Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 475–482, New York, July 2002. ACM Press.

Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.

Sandip Sen and Neelima Sajja. Robustness of reputation-based trust: Boolean case. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 288–293, New York, July 2002. ACM Press.

Kari Sentz and Scott Ferson. Combination of evidence in Dempster-Shafer theory, 2002a. Sandia National Laboratories 2002-0835.

Karl Sentz and Scott Ferson. Combination of evidence in Dempster Shafer theory. TR 0835, Sandia National Laboratories, Albuquerque, New Mexico, 2002b.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

Claude E. Shannon. The mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

Carles Sierra and John K. Debenham. Information-based agency. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1513–1518, Detroit, 2007. IJCAI.

Philippe Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:447–458, 1990.

Jonathan D. H. Smith. Some observations on the concepts of information-theoretic entropy and randomness. *Entropy*, 3:1–11, 2001.

Luke Teacy, Jigar Patel, Nicholas Jennings, and Michael Luck. Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 997–1004. ACM Press, July 2005.

W. T. Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.

Yonghong Wang. A theoretical approach to the problem of distributed reputation management. In *Proceedings of the Workshop on Trust in Agent Societies at AAMAS*, 2004.

Yonghong Wang and Munindar P. Singh. Trust representation and aggregation in a distributed agent system. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, pages 1425–1430, Menlo Park, 2006. AAAI Press.

Yonghong Wang and Munindar P. Singh. Formal trust model for multiagent systems. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1551–1556, Detroit, 2007. IJCAI.

Stephen Weeks. Understanding trust management systems. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 94–105, 2001.

Eric W. Weisstein. Mean deviation, 2003. http://mathworld.wolfram.com/MeanDeviation.html.

Li Xiong and Ling Liu. PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities. *IEEE Transactions on Knowledge and Data Engineering*, 16(7): 843–857, 2004.

Ronald R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences: an International Journal*, 41:93–137, 1987.

Pınar Yolum and Munindar P. Singh. Emergent properties of referral systems. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 592–599, New York, July 2003. ACM Press.

Bin Yu and Munindar P. Singh. Distributed reputation management for electronic commerce. *Computational Intelligence*, 18(4):535–549, November 2002.

Bin Yu and Munindar P. Singh. Detecting deception in reputation management. In *Proceedings of Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 73–80, 2003a.

Bin Yu and Munindar P. Singh. Searching social networks. In *Proceedings of Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 65–72, 2003b.

Giorgos Zacharia and Pattie Maes. Trust management through reputation mechanisms. *Applied Artificial Intelligence*, (14):881–907, 2000.

Lotfi A. Zadeh. On the validity of Dempsters rule of combination. TR 79/24, Department of Computer Science, University of California, Berkeley, 1979.

Lotfi A. Zadeh. Review of shafer's a mathematical theory of evidence. *AI Magazine*, 5(3): 81–83, Fall 1986.

Runfang Zhou, Kai Hwang, and Min Cai. GossipTrust for fast reputation aggregation in peer-to-peer networks. *IEEE Transactions on Knowledge and Data Engineering*, 20(9): 1282–1295, 2008.

Cai-Nicolas Ziegler and Georg Lausen. Spreading activation models for trust propagation. In *IEEE International Conference on e-Technology, e-Commerce, and e-Service (EEE)*, pages 83–97, Taibei, Taiwan, 2004. IEEE Computer Society.

Appendices

Lemma 18. *$f_{r,s}(x)$ is increasing when $x \in [0, \frac{r}{r+s})$ and decreasing when $x \in (\frac{r}{r+s}, 1]$ $f_{r,s}(x)$ is maximized at $x = \frac{r}{r+s}$.*

Proof: To show monotonicity, it is adequate to assume r and s are integers and $r + s > 0$. The derivative

$$\begin{aligned} \frac{df_{r,s}(x)}{dx} &= \frac{x^{r-1}(1-x)^{s-1}}{\int_0^1 x^r(1-x)^s dx}(r(1-x) - sx) \\ &= \frac{x^{r-1}(1-x)^{s-1}}{\int_0^1 x^r(1-x)^s dx}(r - (r+s)x) \end{aligned}$$

Since $r - (r + s)x > 0$ when $x \in [0, \frac{r}{r+s})$ and $r - (r + s)x < 0$ when $x \in (\frac{r}{r+s}, 1]$, we have $\frac{df_{r,s}(x)}{dx} > 0$ when $x \in [0, \frac{r}{r+s})$ and $\frac{df_{r,s}(x)}{dx} < 0$ when $x \in (\frac{r}{r+s}, 1]$. Then $f_{r,s}(x)$ is increasing when $x \in [0, \frac{r}{r+s})$ and $f_{r,s}(x)$ is decreasing when $x \in (\frac{r}{r+s}, 1]$ $f_{r,s}(x)$ has maximum at $x = \frac{r}{r+s}$. □

The motivation behind Lemma 19 is, in essence, to remove the absolute value function that occurs in the definition of certainty. Doing so enables differentiation.

Lemma 19. *Given A and B defined by $f_{r,s}(A) = f_{r,s}(B) = 1$, $0 < A < \frac{r}{r+s} < B < 1$, we have $c_f = \int_A^B (f_{r,s}(x) - 1)dx$*

Proof: As in Definition 2, $r+s > 0$ throughout this paper. By Lemma 18, $f_{r,s}(x)$ is strictly increasing for $x \in [0, \frac{r}{r+s})$, strictly decreasing for $x \in (\frac{r}{r+s}, 1]$ and maximized at $x = \frac{r}{r+s}$. Since the average of $f_{r,s}(x)$ in $[0, 1]$ is 1.0, we have that $f_{r,s}(\frac{r}{r+s}) > 1.0$. Since $f_{r,s}(0) = f_{r,s}(1) = 0$, there are A and B such that $f_{r,s}(A) = f_{r,s}(B) = 1$ and $0 < A < \frac{r}{r+s} < B < 1$
From Lemma 18, we have $f_{r,s}(x) < 1$ when $x \in [0, A)$ or $x \in (B, 1]$ and $f_{r,s}(x) > 1$ when $x \in (A, B)$. By Definition 3, we have $\int_0^1 (f_{r,s}(x) - 1)dx = 0$.
Therefore, $\int_0^A (f_{r,s}(x) - 1)dx + \int_B^1 (f_{r,s}(x) - 1)dx + \int_A^B (f_{r,s}(x) - 1)dx = 0$
and $\int_0^A (1 - f_{r,s}(x))dx + \int_B^1 (1 - f_{r,s}(x))dx$
$= \int_A^B (f_{r,s}(x) - 1)dx$.
Thus $\int_0^1 |f_{r,s}(x) - 1|dx = \int_0^A 1 - (f_{r,s}(x))dx + \int_B^1 (1 - f_{r,s}(x))dx + \int_A^B (f_{r,s}(x) - 1)dx$
and $\frac{1}{2}\int_0^1 |f_{r,s}(x) - 1|dx = \int_A^B (f_{r,s}(x) - 1)dx$. □

Lemma 20.

$$\int_0^1 x^r(1-x)^s dx = \frac{1}{r+s+1}\prod_{i=1}^{r}\frac{i}{r+s+1-i}$$

Proof: We use integration by parts.
$\int_0^1 x^r(1-x)^s dx = \int_0^1 x^r d(\frac{-1}{s+1}(1-x)^{s+1})$

$= -\frac{x^r(1-x)^{s+1}}{s+1}|_0^1 + \frac{r}{s+1}\int_0^1 x^{r-1}(1-x)^{s+1}dx$
$= \frac{r}{s+1}\int_0^1 x^{r-1}(1-x)^{s+1}dx$
$= \cdots$
$= \frac{r\cdot(r-1)\cdots 1}{(r+s)\cdot(r+s-1)\cdots(s+1)}\int_0^1(1-x)^{r+s}dx$
$= \frac{1}{r+s+1}\prod_{i=1}^{r}\frac{i}{r+s+1-i}.$ □

Lemma 21.

$$\lim_{r\to\infty}\sqrt[r]{\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}} = \frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}} \tag{.1}$$

Where r is a positive integer.

Proof: This lemma is used in the next lemma, to show that the right side of an equation approaches a constant, where the equation has duplicated roots, and then the two roots of the equation approach that duplicated root.

$\lim_{r\to\infty}\frac{1}{r}\ln\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}$
$= \lim_{r\to\infty}\frac{1}{r}\ln(\prod_{i=1}^{r} i\prod_{i=1}^{r}\frac{1}{\alpha r+r+1-i})$
$= \lim_{r\to\infty}\frac{1}{r}\ln(\prod_{i=1}^{r} i\prod_{i=1}^{r}\frac{1}{\alpha r+i})$
$= \lim_{r\to\infty}\frac{1}{r}\sum_{i=1}^{i=r}\ln\frac{i}{\alpha r+i}$
$= \lim_{r\to\infty}\frac{1}{r}\sum_{i=1}^{i=r}\ln\frac{\frac{i}{r}}{\alpha+\frac{i}{r}}$
$= \int_0^1\ln\frac{x}{\alpha+x}dx$
$= \ln\frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}}$

Therefore,

$\lim_{r\to\infty}\sqrt[r]{\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}} = \frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}}.$ □

Lemma 22.

$$\lim_{r\to\infty} A(r) = \lim_{r\to\infty} B(r) = \frac{1}{1+\alpha} \tag{.2}$$

Where r is a positive integer.

Proof: The idea is to show that $A(r)$ and $B(r)$ are two roots of an equation $g(x) = \beta(r)$. If $\lim_{r\to\infty}\beta(r) = \beta$ and the equation $g(x) = \beta$ has duplicated roots of α, then we have $\lim_{r\to\infty} A(r) = \lim_{r\to\infty} B(r) = \alpha$

By Lemma 19, $A(r)$ and $B(r)$ are two roots for the equation
$x(1-x)^\alpha = \sqrt[r]{\int_0^1 x^r(1-x)^{\alpha r}dx}$
since
$\lim_{r\to\infty} \sqrt[r]{\int_0^1 x^r(1-x)^{\alpha r}dx}$
$= \lim_{r\to\infty} \sqrt[r]{\frac{1}{\alpha r+r+1}\prod_{i=1}^{r}\frac{i}{\alpha r+r+1-i}}$ (by Lemma 20)
$= \frac{\alpha^\alpha}{(1+\alpha)^{\alpha+1}}$ (by Lemma 21)
$= \frac{1}{1+\alpha}(1-\frac{1}{1+\alpha})^\alpha$
since $x(1-x)^\alpha$ achieves its maximum at $x = \frac{1}{1+\alpha}$, and $x = \frac{1}{1+\alpha}$ is the only root for the equation
$x(1-x)^\alpha = \frac{1}{1+\alpha}(1-\frac{1}{1+\alpha})^\alpha$
Therefore,
$\lim_{r\to\infty} A(r) = \lim_{r\to\infty} B(r) = \frac{1}{1+\alpha}$. □

Proof of Theorem 1 $c(r)$ is increasing where $r > 0$ $c'(r) = \frac{d}{dr}\int_{A(r)}^{B(r)}(\frac{(x^r(1-x)^{\alpha r}}{\int_0^1 y^r(1-y)^{\alpha r}dy} - 1)dx$

$$
\begin{aligned}
&= B'(r)(\frac{B^r(r)(1-B(r))^{\alpha r}}{\int_0^1(y^r(1-y)^{\alpha r}dy} - 1)\\
&\quad -A'(r)(\frac{A^r(r)(1-A(r))^{\alpha r}}{\int_0^1(y^r(1-y)^{\alpha r}dy} - 1)\\
&\quad +\int_{A(r)}^{B(r)}\frac{d}{dr}(\frac{(x^r(1-x)^{\alpha r}}{\int_0^1 y^r(1-y)^{\alpha r}dy} - 1)dx\\
&= \int_{A(r)}^{B(r)}\frac{d}{dr}\frac{(x^r(1-x)^{\alpha r}}{\int_0^1 y^r(1-y)^{\alpha r}dy}dx\\
&= \frac{1}{d^2}(\int_{A(r)}^{B(r)}\frac{d}{dr}(x^r(1-x)^{\alpha r})\int_0^1 y^r(1-y)^{\alpha r}dy\\
&\quad -\int_{A(r)}^{B(r)}(x^r(1-x)^{\alpha r})\frac{d}{dr}\int_0^1 y^r(1-y)^{\alpha r}dy)\\
&= \frac{1}{d^2}(\int_{A(r)}^{B(r)}(x^r(1-x)^{\alpha r})\ln(x(1-x)^\alpha)\int_0^1 y^r(1-y)^{\alpha r}dy\\
&\quad -\int_{A(r)}^{B(r)}(x^r(1-x)^{\alpha r})\int_0^1 y^r(1-y)^{\alpha r}\ln(y(1-y)^\alpha)dy)\\
&= \frac{1}{d^2}\int_0^1\int_{A(r)}^{B(r)} x^r(1-x)^{\alpha r}y^r(1-y)^{\alpha r}\ln\frac{x(1-x)^\alpha}{y(1-y)^\alpha}dxdy
\end{aligned}
$$

where $d = \int_0^1 y^r(1-y)^{\alpha r}dy$ According to Lemma 18 $x^r(1-x)^{\alpha r} > y^r(1-y)^{\alpha r}$ when $x \in [A(r), B(r)]$ and $y \in (0, A(r)] \cup [B(r), 1)$ so we have
$\int_0^{A(r)}\int_{A(r)}^{B(r)} x^r(1-x)^{\alpha r}y^r(1-y)^{\alpha r}\ln\frac{x(1-x)^\alpha}{y(1-y)^\alpha}dxdy > 0$
and
$\int_{B(r)}^1\int_{A(r)}^{B(r)} x^r(1-x)^{\alpha r}y^r(1-y)^{\alpha r}\ln\frac{x(1-x)^\alpha}{y(1-y)^\alpha}dxdy > 0$
since

$\int_{A(r)}^{B(r)} \int_{A(r)}^{B(r)} x^r(1-x)^{\alpha r} y^r(1-y)^{\alpha r} \ln \frac{x(1-x)^\alpha}{y(1-y)^\alpha} dxdy = 0$
we have $c'(r) > 0$, so $c(r)$ is increasing when $r > 0$. □

Lemma 23. *Define* $L(r) = \frac{1}{\int_0^1 f(x,r)dx} \int_0^{A(r)} f(x,r)dx$ *and* $R(r) = \frac{1}{\int_0^1 f(x,r)dx} \int_{B(r)}^1 f(x,r)dx$. *Where*
$f(x,r) = x^r(1-x)^{\alpha r}$ *Then*
$\lim_{r\to\infty} L(r) = 0$ *and* $\lim_{r\to\infty} R(r) = 0$

Proof: We only need to show that $\lim_{r\to\infty} L(r) = 0$. Since $\lim_{r\to\infty} R(r) = 0$ can be proved similarly. The idea is to show that $L(r)$ is the remainder of the Taylor expansion of $(A+1-A)^{\alpha r+r}$

$$\int_0^A x^r(1-x)^{\alpha r}dx$$
$$= \int_0^A x^r d(\tfrac{-1}{\alpha r+1}(1-x)^{\alpha r+1})$$
$$= \tfrac{-1}{\alpha r+1} x^r(1-x)^{\alpha r+1}|_0^A + \tfrac{r}{\alpha r+1} \int_0^A x^{r-1}(1-x)^{\alpha r+1}dx$$
$$= \tfrac{r}{\alpha r+1} \int_0^A x^{r-1}(1-x)^{\alpha r+1}dx - \tfrac{1}{\alpha r+1} A^r(1-A)^{\alpha r+1}$$
$$= \cdots$$
$$= \tfrac{1}{\alpha r+r+1} \prod_{i=1}^{r} \tfrac{i}{\alpha r+i}(1-(1-A)^{\alpha r+r+1})$$
$$- \sum_{i=1}^{r} \prod_{j=i}^{r} \tfrac{j}{\alpha r+r+1-j} \tfrac{A^i}{i}(1-A)^{\alpha r+r+1-i}$$

So

$$L(r) = \tfrac{1}{\int_0^1 x^r(1-x)^{\alpha r}dx} \int_0^A x^r(1-x)^{\alpha r}dx$$
$$= (\alpha r+r+1) \prod_{i=1}^{r} \tfrac{\alpha r+r+1-i}{i} \int_0^A x^r(1-x)^{\alpha r}dx$$
$$= 1-(1-A)^{\alpha r+r+1}$$
$$-(\alpha r+r+1) \sum_{i=1}^{r} \begin{pmatrix} \alpha r+r \\ i-1 \end{pmatrix} \tfrac{A^i}{i}(1-A)^{\alpha r+r+1-i}$$
$$= (\alpha r+r+1)(\int_0^A (x+1-A)^{\alpha r+r}dx$$
$$- \sum_{i=1}^{r} \int_0^A \begin{pmatrix} \alpha r+r \\ i-1 \end{pmatrix} x^{i-1}(1-A)^{\alpha r+r+1-i}dx)$$

where $\begin{pmatrix} \alpha r+r \\ k \end{pmatrix} = \prod_{i=1}^{k} \frac{\alpha r+r+1-i}{i}$ for any positive integer k. Since

$$(x+1-A)^{\alpha r+r} = \sum_{i=0}^{\infty} \begin{pmatrix} \alpha r+r \\ i \end{pmatrix} x^i (1-A)^{\alpha r+r-i}$$

so we have

$$L(r) = (\alpha r+r+1) \sum_{i=r}^{\infty} \int_0^A \begin{pmatrix} \alpha r+r \\ i \end{pmatrix} x^i (1-A)^{\alpha r+r-i} dx$$

$$= (\alpha r+r+1) \sum_{i=r}^{\infty} \begin{pmatrix} \alpha r+r \\ i \end{pmatrix} \frac{A^{i+1}}{i+1} (1-A)^{\alpha r+r-i}$$

$$\leq \frac{\alpha r+r+1}{r} A \sum_{i=r}^{\infty} \begin{pmatrix} \alpha r+r \\ i \end{pmatrix} A^i (1-A)^{\alpha r+r-i}$$

$$= \frac{\alpha r+r+1}{r} A((A+1-A)^{\alpha r+r} - \sum_{i=0}^{r-1} \begin{pmatrix} \alpha r+r \\ i \end{pmatrix} A^i (1-A)^{\alpha r+r-i})$$

since

$\sum_{i=0}^{r-1} \begin{pmatrix} \alpha r+r \\ i \end{pmatrix} A^i (1-A)^{\alpha r+r-i}$ is the Taylor expansion of $(A+1-A)^{\alpha r+r} = 1$, so

$$\lim_{r\to\infty} 1 - \sum_{i=0}^{r-1} \begin{pmatrix} \alpha r+r \\ i \end{pmatrix} A^i (1-A)^{\alpha r+r-i} = 0$$

and by Lemma 22 $\lim_{r\to\infty} \frac{\alpha r+r+1}{r} A = 1$. Therefore,

$\lim_{r\to\infty} L(r) = 0$ and similarly $\lim_{r\to\infty} R(r) = 0$. □

Lemma 24. $\lim_{r\to\infty} c(r) = 1$

Proof: Let $f = \frac{x^r(1-x)^{\alpha r}}{\int_0^1 x^r(1-x)^{\alpha r} dx}$. Then we have

$c(x) = \int_0^1 f(x)dx - L(x) - R(x) - (B-A)$

since $\int_0^1 f(x)dx = 1$, $\lim_{r\to\infty} B - A = 0$ (by Lemma 22) and $\lim_{r\to\infty} L(r) = \lim_{r\to\infty} R(r) = 0$ (by Lemma 23). So

$\lim_{r\to\infty} c(r) = 1$

Lemma 25. $\lim_{r\to 0} c(r) = 0$.

Proof: We only give a sketch of the proof. Let $f(x) \leq M$ when $r < 1$. For $\forall \epsilon > 0$, let $\delta = \frac{\epsilon}{2(M+1)}$, since $\frac{x^r(1-x)^{\alpha r}}{\int_0^1 x^r(1-x)^{\alpha r} dx}$ approaches to 1 uniformly in the interval $[\delta, 1-\delta]$, when $r \to 0$. So $\exists\ r_0 > 0$ such that,

$|f(x)-1|<\epsilon$ when $r<r_0, x\in[\delta,1-\delta]$. So when $r<r_0$,
$c(r)=\frac{1}{2}\int_0^1|f(x)-1|dx$
$=\frac{1}{2}(\int_0^\delta|f(x)-1|dx+\int_\delta^{1-\delta}|f(x)-1|dx\int_{1-\delta}^1|f(x)-1|dx)$
$<\frac{1}{2}((M+1)\delta+\epsilon+(M+1)\delta)=\epsilon$. Hence we have $\lim_{r\to 0}c(r)=0$. □

Theorem 5 The concatenation operator $\otimes$ is associative.

Suppose $M_1=(b_1,d_1,u_1)$, $M_2=(b_2,d_2,u_2)$, $M_3=(b_3,d_3,u_3)$, $M_1\times(M_2\times M_3)=(b,d,u)$, and $(M_1\times M_2)\times M_3=(b',d',u')$, Then
$b=b_1(b_2b_3)=b_1b_2b_3$ and
$b'=(b_1b_2)b_3=b_1b_2b_3$, so we have
$b=b'$
$d=b_1(b_2d_3)=b_1b_2d_3$ and $d'=(b_1b_2)d_3=b_1b_2b_3$, so we have $d=d'$, and
$u=1-b-d=1-b'-d'=u'$
Which completes the proof.

Proof of Theorem 6

Aggregation ($\oplus$) is associative. Since the belief space is equivalent to the evidence space. So the aggregation operator in the belief space is equivalent to the plus operator in the evidence space by our definition, and the plus operator is associative, so the aggregation operator in the belief space is also associative.

Proof of Theorem 8

Aggregation ($\oplus$) is commutative.

Since the belief space is equivalent to the evidence space. Since the aggregation operator in the belief space is equivalent to the plus operator in the evidence space by our definition, and the plus operator is commutative, so the aggregation operator in the belief space is also commutative.

Proof of Theorem 10

We need to prove that

$$M_1\otimes(M_2\oplus M_3)=(M_1\otimes M_2)\oplus(M_1\otimes M_3)$$

Assume the equivalent events corresponding to M_2 and M_3 are $\langle,r_2,s_2\rangle$ and $\langle,r_3,s_3\rangle$ respectively. Let $M_1=(b_1,d_1,u_1)$. Then the equivalent event corresponding to $M_1\otimes(M_2\oplus M_3)$ is $(b_1(r_2+r_3),b_1(s_2+s_3)$ and Then the equivalent event corresponding to $(M_1\otimes M_2)\oplus(M_1\otimes M_3)$ is $(b_1(r_2),b_1(s_2)+(b_1(r_3),b_1(s_3)=(b_1(r_2+r_3),b_1(s_2+s_3)$ Which completes the proof.

Proof of Theorem 14 $c(t)$ is increasing where $t > 0$

Let $O = \{o_1, o_2, \cdots, o_m\}$ be the *Frame of Discernment* and $\{b|b : 2^O \mapsto [0,1]\}$ where $b(\emptyset) = 0, \sum_{A \subseteq O} b(A) = 1$ be a belief.

$$\begin{array}{l} b_i = \sum_{A, A \subset O, A \neq \emptyset, o_i \in A} \frac{b(A)}{|A|} \log_m \frac{m}{|A|}, 1 \leq i \leq m \\ c(t) = \frac{1}{2} \int_\Omega |\frac{\prod_{i=1}^{m-1} x_i^{b_i t} (1 - \sum_{i=1}^{m-1} x_i)^{b_m t}}{\int_\Omega \prod_{i=1}^{m-1} x_i^{b_i t} (1 - \sum_{i=1}^{m-1} x_i)^{b_m t} dx} - (m-1)!| dx \end{array} \tag{.3}$$

1. We can simplify the notations. Define $f(x) = \prod_{i=1}^{m-1} x_i^{b_i} (1 - \sum_{i=1}^{m-1} x_i)^{b_m}$ Then $c(t) = \frac{1}{2} \int_\Omega |\frac{f(x)^t}{\int_\Omega f(y)^t dy} - (m-1)!| dx$
2. Eliminate the absolute sign. There exists a region Ω_t such that $\Omega_t \subset \Omega$, $\frac{f(x)^t}{\int_\Omega f(x)^t dx} > (m-1)!$ when $x \in \Omega_t$, $\frac{f(x)^t}{\int_\Omega f(x)^t dx} < (m-1)!$ when $x \in \Omega \backslash \Omega_t$, and $\frac{f(x)^t}{\int_\Omega f(x)^t dx} = (m-1)!$ when $x \in \partial\Omega_t$ where $\partial\Omega_t$ is the boundary of Ω_t. So
   $c(t) = \int_{\Omega_t} \frac{f(x)^t}{\int_\Omega f(y)^t dy} - (m-1)! dx$
3. Differentiate $c(t)$.
   $\frac{d}{dt} \int_{\Omega_t} (\frac{f(x)^t}{\int_\Omega f(y)^t dy} - (m-1)!) dx$
   $= \int_{\partial\Omega_t} (\frac{f(x)^t}{\int_\Omega f(y)^t dy} - (m-1)!) g(x,t) ds$
   $+ \int_{\Omega_t} \frac{d}{dt} (\frac{f(x)^t}{\int_\Omega f(y)^t dy} - (m-1)!) dx$
   The first term is the integral on the boundary of Ω_t which is denoted by $\partial\Omega_t$, while ds is the variable defined on the boundary. $g(x,t)$ is a function which is related with the deformation of Ω_t. The first term is 0 since $\frac{f(x)^t}{\int_\Omega f(y)^t dy} - (m-1)! = 0$ when $x \in \partial\Omega_t$.
   So $c'(t) = \int_{\Omega_t} \frac{d}{dt} \frac{f(x)^t}{\int_\Omega f(y)^t dy} dx$
   $= \int\limits_{\Omega_t} \frac{f(x)^t \ln f(x) \int\limits_\Omega f(y)^t dy - f(x)^t \int\limits_\Omega \ln f(y) f(y)^t dy}{(\int\limits_\Omega f(y)^t dy)^2} dx$
   $= \frac{1}{d^2} \int\limits_{\Omega_t} \int\limits_\Omega f(x)^t f(y)^t \ln \frac{f(x)}{f(y)} dy dx$
   where $d = \int_\Omega f(y)^t dy$. We divide region Ω by Ω_t and $\Omega \backslash \Omega_t$, so the integral above will be the sum of two terms, $c'(t) = A_1 + A_2$.

$$A_1 = \frac{1}{d^2} \int\limits_{\Omega_t} \int\limits_{\Omega_t} f(x)^t f(y)^t \ln \frac{f(x)}{f(y)} dy dx \tag{.4}$$

$$A_2 = \frac{1}{d^2} \int\limits_{\Omega_t} \int\limits_{\Omega \setminus \Omega_t} f(x)^t f(y)^t \ln \frac{f(x)}{f(y)} dydx \qquad (.5)$$

If we exchange x and y , we have $A_1 = \frac{1}{d^2} \int\limits_{\Omega_t} \int\limits_{\Omega_t} f(y)^t f(x)^t \ln \frac{f(y)}{f(x)} dxdy = -A_1$
so we have $A_1 = 0$.
For the second term, by definition of $A(t)$, we have $\frac{f(x)}{d} > (m-1)!$ and $\frac{f(y)}{d} < (m-1)!$, then $\ln \frac{f(x)}{f(y)>0}$, and $A_2 > 0$. So we have $c'(t) > 0$.

Proof of Theorem 13 $c(t) \in (0,1), t > 0$. We only show $c(t) < 0$ since $c(t) > 0$ is trivial.
$c(t) = \frac{1}{2} \int_\Omega |\frac{f(x)^t}{\int_\Omega f(y)^t dy} - (m-1)!| dx$
$< \frac{1}{2} \int_\Omega (\frac{f(x)^t}{\int_\Omega f(y)^t dy} + (m-1)!) dx = 1.$
So $c(t) < 1$ which completes the proof.