

MULTINOMIAL AGENT'S TRUST MODELING USING ENTROPY OF THE DIRICHLET DISTRIBUTION

Mohammad Anisi¹ and Morteza Analoui²

¹School of Computer Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran

m.anisi@ece.ut.ac.ir

²School of Computer Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran

analoui@iust.ac.ir

ABSTRACT

Nowadays one of the most important challenges for integrated systems such as social networks is the evaluation of trust for agents which are interacting with each other in the environment. It plays an important role that the trust has been evaluated from the agent's experiences. In this paper, we propose a new mathematical approach based on the entropy of Dirichlet distribution, to model the agent's trust to another agent, based on the past observation from him. The use of Dirichlet distribution for trust model, allows us to evaluate the trust in multinomial cases. Also our approach considers the uncertainty and conflict of agent's behavior for underlying trust model.

KEYWORDS

Agent, Trust, Uncertainty, Entropy, Dirichlet distribution

1. INTRODUCTION

In simple words, a key intuition about trust is that reflects the trusting party's belief which the trusted party will support its plans [2]. For example, Alice trusts Bob as a mechanic. This means that Alice relies on Bob to repair his car and he believes that Bob will do his job perfectly. It is obvious that rationally Ali's belief resulted from experiences and observations established in interaction with Bob.

An agent keeps his experiences of interaction with the other one as a set of observation quality factors. For an example, Agent B has served agent A perfectly 3 times and imperfectly 5 times. Obviously, observation quality spaces is not limited to perfect and imperfect—that is *good* and *bad*. It can be modeled in more category such as *good*, *average* and *bad*.

Another important issue in observations analysis is the existence of conflict among them. Suppose this: agent A has acted perfectly 4 times and also imperfectly 4 times, where, agent B has acted perfectly 7 times and imperfectly 1 time. It is obvious that agent A has more conflicting actions in comparison with B and consequently, one may trust him with less uncertainty.

Agents trust modelling is the subject of many researches. [12] has proposed a trust evaluation method based on trust uncertainty measure, but the shortcoming is that there is no relation between agents experiences to the evaluated trust. [8] has introduced an approach to evaluate trust based on agents experiences and uncertainty. Following him, [9] has modelled observations using multinomial space. However, none has mentioned the conflicts between observations and they both have just assumed uncertainty as a function of total observations

count. [10] and [11] have modelled agents trust based on their observations and a probability model for uncertainty of observations. Although this method for trust modelling has considered conflicts between observations, but it models observations in a binomial case which consists of *good* and *bad* as observation quality and has no solution for multinomial space.

On the other hand, [16] has a method to model trust in wireless sensor networks using discrete case entropy, but its approach has two main weaknesses, first his trust model has been accept just simple two dimensions, the rate of good observations and the rate of bad observations, and second he just has been used from entropy in his model to deal with evidence's conflict, and there is no way to model the strength of evidence.

Therefore, this research tries to model trust, considering the following properties: **1)** Agents trust is a result of their observations in interaction with each other. **2)** Observation qualities are not limited to *good* and *bad*, so there is no limitation for observation space. **3)** Uncertainty is modeled in agents trust. **4)** The proposed uncertainty measure depends on conflicts between observations and considers them.

This paper is organized as follows: In section 2, first Dirichlet distribution is described as a basis for the introduced method of agents trust modelling. In section 3, the entropy notion is described, and furthermore, by the use of the Dirichlet distribution entropy, a method is proposed to model the uncertainty in agents' behaviour. Additionally, method's properties and also results are analyzed in this section. Section 4 describes the mapping of observations space to trust space and finally, the last section contains concluding remarks on the proposed method.

2. MODELLING AGENT'S BELIEF

In this section we introduce the preliminaries for modelling agent's belief to another agent using the Dirichlet distribution.

2.1. The Dirichlet Distribution

The multinomial distribution is a generalization of the binomial for the situation in which each trial results in one and only one of several categories, as opposed to just two, as in the case of the binomial experiment. Let $y = (y_1, \dots, y_k)$ where y_i is the number of independent trails that results in category $i = 1, 2, \dots, k$. Then the likelihood function of this experiment is $f(y|p)$, such that

$$f(y|p) \propto \prod_{i=1}^k p_i^{y_i} \quad (1)$$

where p_i is the probability that a given trails results in one of k category, $i = 1, 2, \dots, k$. The parameter space of vector P is $P = \{p_i | p_i \geq 0; i = 1, 2, \dots, k; \sum_{j=1}^k p_j = 1\}$ and the vector of observation y satisfies, $y = \{y_i | \sum_{i=1}^k y_i = N\}$ where N equals to the total number of observations.

The so-called *Dirichlet* distribution is the conjugate family—the family of distributions in *Bayesian* analysis that the prior and posterior distributions have the same type—of priors for the multinomial distribution. The probability density function of Dirichlet distribution is such that

$$Dirichlet(p; \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^k p_i^{\alpha_i - 1} \quad (2)$$

where $p_i \geq 0$; $\sum_{i=1}^k p_i = 1$ and $\alpha_i > 0$. The parameter α_i can be interpreted as *prior observation counts* for the events of category i . The normalization constant $B(\alpha)$ is $B(\alpha) = \prod_{i=1}^k \Gamma(\alpha_i) / \Gamma(\sum_{i=1}^k \alpha_i)$ that is the multinomial generalization for *Beta* function [1]. Using α_i as the prior in the above multinomial experiment, yields a Dirichlet posterior with parameters $y_i + \alpha_i$. So the probability density function of posterior distribution is

$$Dirichlet(p; y + \alpha) = \frac{1}{B(y + \alpha)} \prod_{i=1}^k p_i^{y_i + \alpha_i - 1}. \quad (3)$$

We can use natural non-informative prior by taking $\alpha_i = 1$ for all categories i [3]. The expected value of each probability p_i in the Dirichlet distribution is also known as $E(p_i) = \alpha_i / \sum_{j=1}^k \alpha_j$ that is the relative frequency for occurrences of events in the category i . The expected value can be interpreted as the chance of occurring event of category i in the next trail [13]. So if there is an experiment with observation parameter x , the posterior distribution with non-informative prior has been modelled by $Dirichlet(p; x + 1)$ and the chance of occurring event of category i at next trail, becomes

$$E(p_i) = \frac{x_i + 1}{k + \sum_{i=1}^k x_i} \quad (4)$$

In the case of Beta distribution, that we have only two categories and events of each category has been occurred x_1 and x_2 times respectively, the expected value for p_1 becomes $x_1 + 1 / x_1 + x_2 + 2$, that leads the same result as *Laplace's rule of succession* in the course of treating the sunrise problem[7]. So for contract, we will use non-informative prior for Dirichlet distribution throughout this paper.

2.2. Agents Belief

In this section we introduce a way to model the belief of agent to the quality of another one, based on the quality of past experiences had been took place between them. Let see some examples.

Example 1- Suppose there are two agents A and B interacting with each other. Agent A had 5 positive and 2 negative observations from Agent B in the past experiences. In this situation, the probability that B act positively in next interaction with A , according to (4), is equal to the expected value of p_1 in Dirichlet distribution with posterior $Dirichlet(p; [5 + 1, 2 + 1])$, so it results $E(p_1) = 6/9$. Similarly the probability that B act negatively with respect to A is $E(p_2) = 3/9$.

There is no limitation on the number of qualities for experiences that take place between agents and because of the capability of Dirichlet distribution; these qualities can be extended to multiple ones. Let see the example below.

Example 2- Suppose the quality of experiences in example 1 is divided to three parts, good, average and bad. And there are 4, 1 and 2 observations for good, average and bad experiences respectively. So the probability that Agent B act in positive manner with A will be calculated

from the expected value of p_1 for $Dirichlet(p; [4 + 1, 1 + 1, 2 + 1])$ that is $5/10$. Likewise the probability of being average and bad will be $2/10$ and $3/10$ respectively.

With these preliminaries, we define the belief of agent A on the quality of agent B based on the past observations of agent B as follows.

Definition 1- The belief of agent A with respect to agent B and from the past observation of this agent is

$$Belief_B^A(x) = \{b|b_i = E(p_i)\} \quad (5)$$

where $x = \{x_1, \dots, x_k\}$ is the observation vector of agent A from B and $E(p_i)$ is the expected value for each parameter of Dirichlet distribution $Dirichlet(p; x + 1)$. With this definition, the belief of agent A from B for example 1 and 2 becomes $[2/3, 1/3]$ and $[5/10, 2/10, 3/10]$ respectively.

3. UNCERTAINTY MODELLING IN AGENT BEHAVIOR

There are a lot of cases that derived belief will be equal for those ones because the belief measure has been affected just by the rate of observations at each category. Therefore we are looking for a measure with the ability of differentiating such cases. So our concentration in this paper is to model uncertainty of agent's behavior that truly models these cases. For this purpose, we introduce a method to model uncertainty based on information theoretic entropy of Dirichlet distribution based on the belief model proposed in section 2.

3.1. Entropy

Entropy as a scientific notion was introduced first in thermodynamics by Clausius in 1850. Several years later, it was given a probabilistic interpretation in the context of statistical mechanics (Boltzmann 1877). In 1948 Shannon established the connection between entropy and typical sequences. This led to the solution of a number of basic problems in coding and data transmission and finally became the initial basics for the information theory [4]. Shannon entropy calculate the number of bits required to obtain a secure transmission in communication but there are different type of interpretation for the concept of entropy that depend critically on the nature of the context of the underlying problem [5].

Generally the entropy of a discrete random variable X with possible value $\{x_1, \dots, x_n\}$ is $H(X) = E(I(X))$ where E is the expected value and I is the information content of X . $I(X)$ is itself a random variable, so if $p(X)$ denotes the probability mass function of X , then the entropy can explicitly be written as follows.

$$H(X) = - \sum_{i=1}^n p(x_i) \times \ln(p(x_i)) \quad (6)$$

This is the so-called Shannon entropy. There is also a well defined extension for entropy of continuous random variable X with probability density function $f(X)$, that represents the maximum information obtainable within the duration of an experiment that is modelled by this random variable [6]. The extension is

$$H(X) = -E(\ln f(X)) = - \int_X f(x) \times \ln(f(x)) dx \quad (7)$$

In 6 and 7 when the argument of \log is zero, then its result equals to zero.

3.2. Uncertainty in Agent's Behavior

We mentioned in section 2 that the agent's belief to the behavior of another one according to his experiences, can be modeled by expected value of parameters of the Dirichlet distribution and then we saw in previous section that the information obtainable from a random experiment can be achieved by the entropy of the random variable that models the experiment. So if we look at the observation from an agent with the viewpoint of a random experiment, thus we can calculate the information that we have from the agent until now. So we define the information of an agent from another agent as follows.

Definition 2- Information of agent A from agent B with respect to the last observation from agent B is

$$Info_B^A(x) = H(Dirichlet(p; x + 1)) \quad (8)$$

where x is the observation vector from agent B and H is the Shannon entropy calculated by (7). It can be mathematically proved that entropy of Dirichlet distribution with parameter α is equivalent to $\ln(B(\alpha)) + \sum_1^k ((\alpha_i - 1) \times (\psi(\alpha_i) - \psi(\alpha_o)))$ where $\alpha_o = \sum_1^k \alpha_i$, $\psi(\alpha) = \frac{\partial}{\partial \alpha} \ln(\Gamma(\alpha))$ and $B(\alpha)$ is the multinomial case for Beta function [1]. So the $Info_B^A(x)$ can be rewritten as follows.

$$Info_B^A(x) = \ln(B(x + 1)) + \sum_{i=1}^k x_i \times (\psi(x_i + 1) - \psi(x_o + k)) \quad (9)$$

Where x_o is equivalent to $x_1 + \dots + x_k$. Figure 1 shows $Info_B^A(x)$ where x contains two elements x_1 and x_2 , representing the number of successes and failures of agent B respectively and when these elements vary between 0 and 50. As we see in figure 1, the information of agent A from agent B , generally increases when the total number of observation increased. This is intuitively a good feature for $Info_B^A(x)$ that we will explain later in section 3.3.

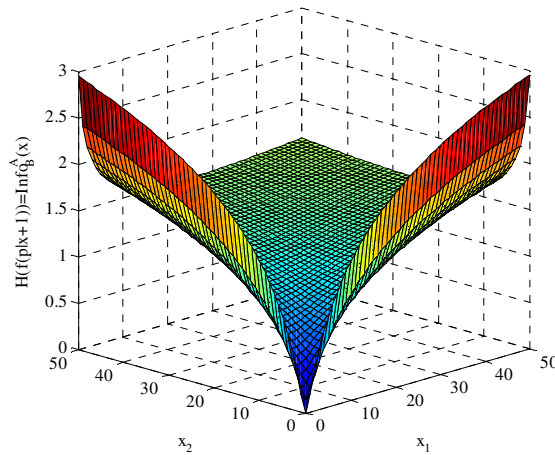


Figure 1, Entropy of Dirichlet distribution for $0 \leq x_1, x_2 \leq 50$.

It is intuitively justifiable that when we have more information from someone or something, we have more certainty to make decision about him. With this point of view, we can relate certainty in making decision to the information obtained about a subject until now. Thus we can say $C_B^A(x) \propto Info_B^A(x)$ where $C_B^A(x)$ is the certainty of agent A in relationship with agent B . Certainty in the literature of trust modelling usually has a real value between 0 and 1 [8,9,10,11]. But it is obvious from (9), $Info_B^A$ ranges over 0 and $+\infty$. So we have to rescale $Info_B^A$ to define a metric for C_B^A . Thus we define certainty as follows.

Definition 3- The certainty of agent A for making decision about agent B with the past observation from agent B is equal to $C_B^A(x) = g(Info_B^A(x))$, where $g(v)$ is a monotonically increasing function from $[0, +\infty)$ to $[0,1]$ that acts as a scaling function and x is the observation vector of agent A from agent B .

There are so many functions satisfy the condition of def. 3 for $g(v)$ and it is obvious that the result of C_B^A for different g will differ. But it is worth mentioning that g just acts as a scaling function, and it does not affect the final decision of agent. Please look at this example,

Example 3- Let $Info_B^A(x) = 1.5$ and $Info_C^A(x') = 2$ and also there are $g(v) = 1 - e^{-v}$ and $g'(v) = 1 - e^{-\frac{v}{2}}$. It is clear that both g and g' satisfy conditions of def. 3. Therefore when g is the scaling function, we have $C_B^A(x) = 0.777$ and $C_C^A(x') = 0.918$, and for g' we have $C_B^A(x) = 0.528$ and $C_C^A(x') = 0.714$. Finally, it is obvious that for both g and g' , agent A has more certainty on agent C rather than B . So there is no difference to choose either of g or g' . So we choose $g(v) = 1 - e^{-v}$, then certainty in def. 3 equals to $1 - e^{-Info_B^A(x)}$.

Since certainty and uncertainty are complement concepts, so we can define uncertainty as follows.

Definition 4- The uncertainty of agent A for making decision about agent B with the past observation from agent B is equal to

$$U_B^A(x) = 1 - C_B^A(x) = e^{-Info_B^A(x)} \quad (10)$$

Figure 2 shows $U_B^A(x)$ when x has two element x_1 and x_2 , and these elements vary from 0 to 50. One can see in fig. 2, when there is no observation from agent B , uncertainty has the maximum value 1 and when observations are increased, the uncertainty approaches to 0. This approach is faster when all observations are likely placed in one of two categories. We will discuss this behavior in section 3.3.

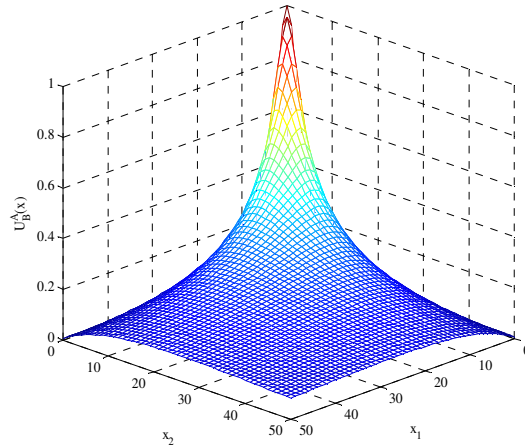


Figure 2, $U_B^A(x)$ when $0 \leq x_1, x_2 \leq 50$.

3.3. Properties of Uncertainty

There are two conditions under which the behavior of the proposed uncertainty measure must be analyzed. Firstly, when the observations from an agent for each quality, are growing with the same rate and secondly, different conditions under which the total observations do not vary while the observations of each quality vary.

For the first case, assume there are three different qualities for observations, good, normal and bad. In this situation, the probability of observing each quality remains identical when new observations take place. Therefore the belief does not vary among the time. For an example let observation vector of agent A from B in two different time, is $[1, 5, 7]$ and then $[2, 8, 11]$. So the belief according to (5) in both cases, is $[0.125, 0.375, 0.5]$. But it is obvious that in the second case, agent A has more certainty to make decision about agent B . We can calculate the uncertainty according to (10) for first and second case is equal to 0.133 and 0.095 respectively. Figure 3 shows this property in a sample situation in the case of two quality categories, which verifies our claim. Theorem 1 generalizes this behavior.

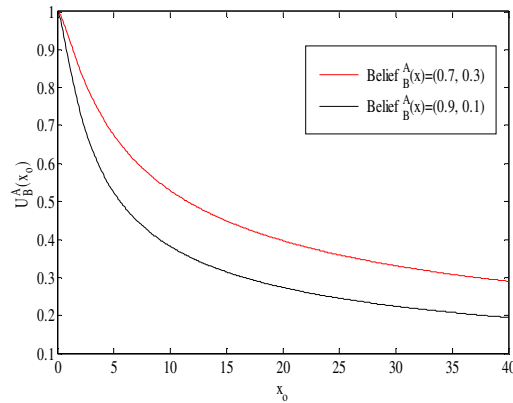


Figure 3, $U_B^A(x)$ for identical belief while total observations are increasing

Theorem 1- If $Belief_B^A(x)$ remains constant for different x , as x_o increases, $U_B^A(x)$ will be decreased.

Proof sketch- While $Belief_B^A(x)$ is constant, $U_B^A(x)$ will be a function of $x_o = \sum_1^k x_i$. So It can be rewritten as $U_B^A(x_o)$. Thus it is enough for the proof to show that $\frac{d}{dx_o}\{U_B^A(x_o)\} < 0$ or accordingly $\frac{d}{dx_o}\{C_B^A(x_o)\} \geq 0$. Appendix contains the complete proof. \square

In the second case, consider the situation which the total observation is constant and observation of events at each category varies. To explain more, we have to introduce a way of clarifying this situation. So according to [14] we define conflict measure as follows.

Definition 5- For $b_i \in Belief_B^A(x)$ the conflict measure is equivalent to $Conflict_B^A(x) = \frac{H(Belief_B^A(x))}{\ln k} = \frac{-\sum_1^k b_i \times \ln b_i}{\ln k}$ where H is equal to the discrete entropy for the element of belief vector according to (6).

Conflict measure according to [15] is equivalent to measure of fuzziness for fuzzy sets or according to [14] equals to specificity measure in evidence theory. According to the definition of discrete entropy, when all b_i have the same value, the entropy is at its maximum value which equals to $\ln k$ and when the differences between all b_i is maximum, the entropy and therefore conflict is at its minimum value which is 0 [4]. So we can conclude that conflict measure has a value in the range of $[0, 1]$. Figure 4 shows the behavior of proposed uncertainty measure with respect to conflict when total observation is 40 and there are 2 quality categories. We can see that when $Conflict_B^A$ increases, U_B^A increases.

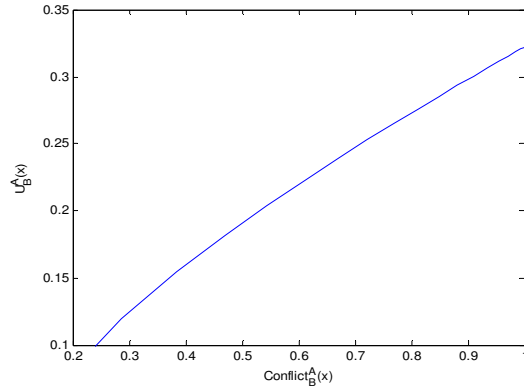


Figure 4, $U_B^A(x)$ behavior against $Conflict_B^A(x)$

Table 1 is the comparison of uncertainty measure in our approach against the approaches introduced in [8, 9] and [10, 11] while total observation is 4. We can see from the result, our approach and [10, 11] care the conflict existed in observations but [8, 9] does not care about it. However, our approach has dominant on [10, 11], because of the ability of treating multinomial cases.

Table 1, Uncertainty comparison for different approaches

[good, bad]	[0,4]	[1,3]	[2,2]	[3,1]	[4,0]
[10,11]	0.46	0.65	0.71	0.65	0.46
[8,9]	0.2	0.2	0.2	0.2	0.2
Our Approach	0.44	0.69	0.76	0.69	0.44

4. TRUST SPACE

In previous sections, as [8, 9], we consider experiences of an agent in relationship with another agent as an observation vector which elements are the observation count for each quality. Now we are going toward the final target which is defining our trust report from agent's experiences. And then define a mapping from experience space to trust space.

Experience space can be formally defined as $S_E = \{x | x_i \geq 0 ; x_o = \sum_1^k x_i ; x_o > 0\}$ where k is the number of different quality categories. According to [9] a trust report contains the agent's belief to another with the uncertainty of its behavior. So we can say a trust report has $k + 1$ element, first k elements represent the belief and the last element act as uncertainty. Thus the trust space contains all vectors satisfying (11).

$$S_T = \left\{ (b_1, \dots, b_k, u) \mid b_i \geq 0 ; u \geq 0 ; u + \sum_{i=1}^k b_i = 1 \right\} \quad (11)$$

Now we can define a mapping from sample experience to trust which means we can evaluate agent's trust for a given sample experience.

Definition 6- Trust of agent A in relation with agent B , based on the past experiences x , is $T_B^A(x) = (b'_1, \dots, b'_k, u)$ where $b'_i = b_i \times C_B^A(x)$, $b_i \in Belief_B^A(x)$ and $u = U_B^A(x) = 1 - C_B^A(x)$.

It can be easily proved that trust in definition 6, satisfies all conditions of (11). While the consisting components of the trust in Def. 6—which are uncertainty and belief—this trust model in contrast of [8, 9] considers the conflict of experiences and in opposite of [10, 11] considers the multinomial cases for experiences.

5. CONCLUSION

In this paper we proposed a model to evaluate agents trust by the use of Dirichlet distribution. Our model uses the experiences of agent in relationship with each other, for this evaluation. The ability of Dirichlet distribution to model multinomial random experiments provides a multinomial trust modelling basis which is the major weakness for the other approaches of this literature. Additionally, the use of information theoretic entropy notion, as an information measure, provides a good rational and mathematical fundament for our model which is its advantage over other proposed models in trust modelling context.

Also we introduce a measure to show conflict in observations. Although this is not a new concept is computer science, but we are the first one who use it to model conflict in observations in the context of trust modelling. We model uncertainty in agent's behaviour with the use of entropy of Dirichlet distribution which has a good behaviour against total number of observations and the conflict of observations.

REFERENCES

- [1] Olver F. W., Lozier D. W., Boisvert R. F. and Clark C. W., *NIST Handbook of Mathematical Functions*, :Cambridge University Press, 2010.
- [2] Castelfranchi C. and Falcone R., "Principles of trust for MAS: cognitive anatomy, social importance, and quantification," in Proc. of the 3rd Int. Conf. on Multi Agent Systems, 1998, pp. 72--79.

- [3] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin, Bayesian Data Analysis, 2nd Edition.: Chapman and Hall, 2003.
- [4] Papoulis A., Probability, Random Variables and Stochastic Processes. McGraw-Hill Companies, 1991.
- [5] Shannon C. E., "A mathematical theory of communication," Bell system technical journal, vol. 27, 1948.
- [6] Smith J. D., "Some observations on the concepts of information theoretic entropy and randomness," Entropy and Randomness, vol. 3, pp. 1--11, 2001.
- [7] Zabell S. L., "The rule of succession," Erkenntnis, vol. 31, no. 2, pp. 283-321, 1989.
- [8] Jøsang A. and Ismail R., "The Beta Reputation System," in Proc. of the 15th Bled Electronic Commerce Conference, :Citeseer, 2002, vol. 160, pp. 324–337.
- [9] Jøsang A. and Haller J., "Dirichlet Reputation Systems," in Proc.s of International Conference on Availability, Reliability and Security, :IEEE Comp. Soci., 2007, pp. 112--119.
- [10] Wang Y. and Singh M. P., "Formal trust model for multiagent systems," in Proc. of the 20th international joint conference on Artificial intelligence, 2007, pp. 1551--1556.
- [11] Wang Y. and Singh M. P., "Evidence-based trust: A mathematical model geared for multiagent systems," ACM Trans. Auton. Adapt. Syst., vol. 5, no. 4, pp. 14:1--14:28, November 2010.
- [12] Carbone M., Nielsen M. and Sassone V., "A Formal Model for Trust in Dynamic Networks," in Proc. of International Conference on Software Engineering and Formal Methods, 2003, pp. 54--63.
- [13] Honkela A., "Nonlinear Switching State-Space Model," in Procs of Informaatiotekniikka, 2001, pp. 56--73.
- [14] Yaeger R. R., "Entropy and Specificity in a Mathematical Theory of Evidence," Classic Works of the Dempster-Shafer Theory of Belief Functions, vol. 219, Springer Berlin / Heidelberg, 2008.
- [15] Wang Z. and Klir G. J., Fuzzy measures Theory, :Kluwer Academic Publishers Norwell, MA, USA, 1993.
- [16] Hongjun D., Zhiping J. and Xiaona D., "An Entropy-based Trust Modeling and Evaluation for Wireless Sensor Networks," in Proceedings of the 2008 International Conference on Embedded Software and Systems, IEEE Computer Society, 2008, pp. 27-34

APPENDIX

Theorem 1- If $Belief_B^A(x)$ remains constant for different x , as x_o increases , $U_B^A(x)$ will be decreased.

Proof- While $Belief_B^A(x)$ is constant, all elements of belief vector— b_i —are constants, and we can say $U_B^A(x)$ is a function of $x_o = \sum_1^k x_i$ while $\frac{x_{i+1}}{x_o+k} = b_i$. So uncertainty can be rewritten as $U_B^A(x_o)$. Thus it is enough for the proof to show that $\frac{d}{dx_o}\{U_B^A(x_o)\} < 0$ or accordingly $\frac{d}{dx_o}\{C_B^A(x_o)\} \geq 0$. Since $C_B^A(x_o)$ is scaled by our scale function $g(v)$ that is a monotonically

increasing function of $Info_B^A(x_o)$, it is enough to show that $Info_B^A(x_o)$ is a monotonically increasing function. So we have to prove that

$$\frac{d}{dx_o} \{Info_B^A(x_o)\} \geq 0$$

For simplicity we can substitute $x_i + 1$ with α_i and rewrite the $Info_B^A$ as $\ln(B(\alpha)) - \sum_1^k (\alpha_i - 1) \times (\psi(\alpha_i) - \psi(\alpha_o))$ where $\frac{\alpha_i}{\alpha_o} = b_i$; $b_i \geq 0$ and $\alpha_i \geq 1$. Since all b_i are known for us, we can say

$$Info_B^A(\alpha) = \ln\left(\frac{\prod_1^k \Gamma(b_i \times \alpha_o)}{\Gamma(\alpha_o)}\right) + \sum_{i=1}^k (b_i \times \alpha_o - 1) \times (\psi(b_i \times \alpha_o) - \psi(\alpha_o))$$

so $Info_B^A(\alpha)$ is just a function of just α_o . While $\frac{d}{dx} \ln(\Gamma(x)) = \psi(x)$, for the first derivative we have

$$\begin{aligned} \frac{d}{d\alpha_o} \{Info_B^A(\alpha_o)\} &= \sum_{i=1}^k b_i \times \psi(b_i \times \alpha_o) - \psi(\alpha_o) - \sum_{i=1}^k b_i \times \psi(b_i \times \alpha_o) + \psi(\alpha_o) \times \sum_{i=1}^k b_i \\ &+ \sum_{i=1}^k (b_i \times \alpha_o - 1) \times (b_i \times \psi'(b_i \times \alpha_o) - \psi'(\alpha_o)) \end{aligned}$$

Since $\sum_1^k b_i = 1$, four first clauses will be removed and we have

$$\frac{d}{d\alpha_o} \{Info_B^A(\alpha_o)\} = \sum_{i=1}^k (b_i \times \alpha_o - 1) \times (b_i \times \psi'(b_i \times \alpha_o) - \psi'(\alpha_o))$$

where ψ' is the first derivation of ψ . Since $\alpha_i \geq 1$, the first clause of multiplication in summation is positive and from [1] the second clause is also positive. So $\frac{d}{d\alpha_o} \{Info_B^A(\alpha_o)\}$ is positive and theorem is proved. \square