Reputation-based Trust Evaluations through Diversity

Anthony Etuk, Timothy J. Norman, and Murat Şensoy

Department of Computing Science, University of Aberdeen, Scotland, UK
{aaetuk,t.j.norman,m.sensoy}@abdn.ac.uk

Abstract. Trust and reputation are significant components in open dynamic systems for making informed and reliable decisions. State-of-the-art trust models that exploit reputational evidence generally rely on reports from as many sources as possible. Situations exist, however, where seeking evidence from all possible sources is unrealistic. This is particularly the case in resource-constrained environments where querying information sources is costly, for instance in terms of time and bandwidth. This paper describes an approach that exploits *diversity* among information sources in order to select a small number of candidates to query for reputational evidence. We demonstrate that reliable decisions can be reached using evidence from small groups of individuals. We show that our approach is robust in contexts of variable trust in reputational sources and to a degree of deception.

Categories and Subject Descriptors: I.2.11 [Distributed Artificial Intelligence]: Multi-Agent Systems

General Terms: Experimentation, Performance

Keywords: Diversity, Reputation, Trust

1 Introduction

Reputation-based trust is a critical mechanism in large, open, and dynamic systems, where agents must interact with one another in order to achieve their goals. Agents operating in such environments often rely on *indirect* experience acquired from their peers in order to make informed decisions, especially when *direct* experience on a subject is lacking or insufficient [10]. Whereas state-of-the-art trust models exploiting this kind of evidence generally rely on reports from as many sources as possible, in the physical world capturing and distributing evidence can be costly. For instance, in distributed environments such as peer-to-peer networks, sensor networks, and pervasive computing, each participant is responsible for collecting and combining evidence from others due to lack of central authority or repository. A major constraint in such systems is bandwidth, motivating the need to minimise the number of messages exchanged in order to arrive at a decision. As a result, reputation assessments are often based on a subset of evidence, usually from the agent's neighbourhood [3], an approach which in itself can be problematic, as it does not make use of all the information available, and therefore is prone to biases and deception.

Motivated by this problem, we present an approach for minimising the costs associated with making effective trust assessments, while still remaining robust to biases and deception. In particular, we exploit *diversity* among information sources to intelligently sample from the *crowd* of reputation sources. Our notion of diversity is inspired by the work of Surowiecki et al. [8]. Their work highlights some interesting criteria for effective decisions in large groups of individuals. Diversity ensures that the experiences of different agents based on their private information is exploited in a decision process. As agents most often operate in social contexts where interaction with others is inevitable, so much could go to influence their behaviour. Agent behaviour could be conditioned by organisation, profession, or proximity to other agents in a network. While Surowiecki et al. consider such effects detrimental to a decision process because of the possibilities of collusive behaviour and subjectivity of opinions, we see great potential. For example, where logical subgroups exist in the agent population as informed by a feature-behaviour correlation, we can exploit this to limit the number of agents queried for evidence. In the sections that follow, we demonstrate how this concept can be employed in a manner that leads to positive outcomes.

The rest of this paper is organised as follows: Section 2 highlights the different components of the diversity model. An evaluation of the approach is presented in Section 3, and in Section 4 we conclude with a discussion and avenues for future research.

2 The Diversity Model

The Diversity Model (DM) enables an agent to arrive at reliable decisions using evidence from small groups of individuals. The model employs trust and machine learning techniques in order to build models of information sources from which potential candidates may be sampled for evidence.

We consider an evaluator x, who wishes to evaluate the truth of a proposition ρ, and has access to a set of information sources A, where individual information sources or agents are denoted $x, y, ... \in A$. The notation x will be used in the course of this paper to denote an agent acting in the capacity of an evaluator, while the notation y will be used to represent an agent regarded as an information source. In a more general sense, an agent is regarded as a tuple $\langle ID, F, \mathcal{R} \rangle$, where ID denotes a unique identifier, F is a set of features, and $\mathcal{R}$ is a set of past reports.

A report $\mathcal{R}$, is an opinion about a subject ρ, provided by an agent y, to an evaluator x, in response to a query. An agent y, records its perceived opinion about ρ as $\mathcal{R}_{y,\rho}$, and reports $\mathcal{R}_{y \to x,\rho}$ when queried by x. The variable t denotes a time step associated with a report from y, such that $\mathcal{R}^{t}_{y,\rho}$ represents a report at time t. Consequently, $\mathcal{R}^{t:t+k}_{y \to x,\rho}$ denotes the set of reports received by x from y between the interval t and $t + k$.

Let F denote a finite set of features, such that $f_1, f_2, ..., f_d \in F$. We define a feature as an observable attribute of an agent, e.g., an agent's organisation or its location. An evaluator x, has a view on the relative importance of features

represented by the vector $\langle w_1^x, w_2^x, ..., w_d^x \rangle$, where w_i^x is x's view of the importance of feature f_i, and $w_i^x :\rightarrow [0,1]$. Subsequently, an evaluator uses this metric to compute the similarity between different agents. Similarity between agents is, therefore, a measure of the *distance* between their features as informed by the vector of weights on the features. We employ the *weighted Euclidean distance* to compute similarity between any two agents y_1 and y_2 as:

$$D^{\mathcal{F}}_{y_1,y_2} = \|y_1 - y_2\| \cdot \mathcal{F} = \sqrt{\sum_{k=1}^{|F|} w_k \left(f_{k_{y_1}} - f_{k_{y_2}}\right)^2}. \quad (1)$$

2.1 Group

Let G denote a stratification on A, such that $G = \{G_i, G_j, ..., G_m\}$ where $G_i \bigcap G_j = \emptyset$, if $i \neq j$. We define a group G_i as a collection of homogenous agents, such that $\{x : x \in G_i, G_i \in G\} = A$. Groups are formed subjectively by an agent who attempts to disambiguate what metrics lead to a better stratification of information sources. The group formation process is discussed in Section 2.3. However, the aim of an agent in partitioning the population, is to provide a suitable generalisation of information sources using different distinguishing characteristics. Subsequently, an agent exploits this model to limit the number of sources queried for evidence, and to protect itself against deception. Agents are partitioned into groups based on how similar they are to one another, as specified by a similarity metric. We denote by $G_{i(y)}$, an agent y's membership of a group G_i. An agent x, maintains two parameters $\delta^x_{G_i^F}$ and $\delta^x_{G_i^B}$, which denote the feature-based similarity and the behaviour-based similarity of a group respectively.

The feature similarity $\delta^x_{G_i^F}$ of a group is the degree of similarity of members of the group given their features. This parameter is measured by computing the average weighted distance between pairs of agents in the group as follows:

$$\delta^x_{G_i^F} = 1 - \tfrac{2}{n(n-1)} \sum_{y_p, y_q \in G_i} D^{\mathcal{F}}_{y_p,y_q} \text{ where } p < q, \text{ and } n = |G_i|. \quad (2)$$

An evaluator learns over time the importance of different features while computing similarity. Consider for instance, the following feature set $\langle$*age, profession, location*$\rangle$ describing agents in a population. An agent may assign different weights to different features while measuring similarity. For example, an agent could measure similarity using *age*, or *location*, a combination of *age* and *location*, etc. Although different feature subsets may define different subgroups in the population, not all feature subsets might be distinguishing enough for identifying relevant subgroups in the population. In an example scenario, an agent wishing to evaluate the reliability of a delivery company, may learn informing subgroups in the population of reputation providers, by partitioning agents based on their location for instance, rather than their age or profession. The fact that some locations may be easily accessible (e.g. metropolitan areas), than others (e.g. rural areas), might impact on the satisfaction level of agents obtaining services from

the company, and potentially reveal a relationship between the feature *location* and the ratings obtained from agents.

The behaviour similarity of a group $\delta^x_{G_i^B}$, is a subjective measure of the likelihood of agents in the group behaving in a similar manner. In the context of this paper, behaviour is represented by the report of agents. It is important to emphasise here that behaviour similarity does not capture a semblance of the agents based on their level of trustworthiness (e.g. honest, deceitful), rather it is a measure of the consistency of agents in giving similar reports (be it honest or deceptive ones), in response to the same query. Although agents belonging to the same group may be regarded as having the same level of trustworthiness as depicted in Section 2.2, in our model this condition alone does not satisfy the criteria for grouping agents. It is possible for *dissimilar* agents to have similar level of trustworthiness (e.g. agents from different but highly reputable organisations). In order to effectively exploit diversity in the system, our model requires agents in a group to be similar both in feature and behaviour. To compute the $\delta^x_{G_i^B}$ of a group, a *report matrix* is constructed as illustrated in Figure 1. The

	t_1	t_2	t_3	t_4
y_1	1	5	1	4
y_2	4	1	5	1
y_3	1	5	2	4

Fig. 1. Report matrix for similarity calculation

matrix captures the rating provided by different agents in a given sampling interval. A sampling interval is the time frame for which reports from different agents are considered, and is the same for all the agents. In Figure 1 for example, the sampling interval considered is $t_1 : t_4$ (i.e. t_1, t_2, t_3,t_4), and the reports from the agents could be represented as $\mathcal{R}^{t_1:t_4}_{y_i \to x,\rho}$, $i = 1, .., 3$. Also, agents y_1 and y_3 with report vectors $\langle 1, 5, 1, 4\rangle$ and $\langle 1, 5, 2, 4\rangle$ respectively, may be considered much more similar to each other than agents y_1 and y_2 with report vectors $\langle 1, 5, 1, 4\rangle$ and $\langle 4, 1, 5, 1\rangle$ respectively. Details for the computation of this measure is given in Equation 3 and Equation 4.

$$D^{\mathcal{R}}_{y_1,y_2} = \frac{1}{h}\sqrt{\sum_{t\in h}\left(\mathcal{R}^t_{y_1,\rho} - \mathcal{R}^t_{y_2,\rho}\right)^2}, \tag{3}$$

where $h \in H$ represents the number of past reports taken into consideration. Following Equation 3, $\delta^x_{G_i^B}$ of a group G_i can be computed as:

$$\delta^x_{G_i^B} = 1 - \tfrac{2}{n(n-1)} \sum_{y_p,y_q\in G_i} D^{\mathcal{R}}_{y_p,y_q} \text{ where } p < q, \text{ and } n = |G_i|. \tag{4}$$

2.2 Group Trust and Subjective Logic

An evaluator depending on evidence from third party sources faces the risk of misleading reports from these sources. Not all agents may act in a benevolent

manner or even possess a required level of expertise to report on a subject. Sometimes information sources may exaggerate perceived opinion, or offer testimonies that are outrightly false. Finding ways to reduce the influence of misleading reports from third-party sources is a fundamental problem in reputation systems [10]. One way of mitigating against this problem, is maintaining a reputation of the information sources, and using this to determine the weight given to their reports [9].

Subjective Logic (SL) [1] is a belief calculus which allows agents to express opinions as degrees of belief, disbelief, and uncertainty about propositions. Binary propositions, such as agent y, is trustworthy concerning ρ, lead to opinions which are equivalent to a beta distribution. SL contains operations to represent consensus, recommendation, and ordering, which are useful tools for evidence aggregation. We adopt SL to represent trust because it provides an intuitive way to represent the belief an entity has in another, and a way to aggregate evidence to support such belief. An evaluator x's opinion about an agent y, reporting correctly on ρ is represented in Subjective Logic as a tuple:

$$\begin{aligned} \omega^x_{y:\rho} \quad &= \quad \langle b^x_{y:\rho}, d^x_{y:\rho}, u^x_{y:\rho}, a^x_{y:\rho} \rangle \\ &\text{where } b^x_{y:\rho} + d^x_{y:\rho} + u^x_{y:\rho} = 1, \text{ and } b^x_{y:\rho}, d^x_{y:\rho}, u^x_{y:\rho}, a^x_{y:\rho} \in [0,1]. \end{aligned} \tag{5}$$

In the above representation, $b^x_{y:\rho}, d^x_{y:\rho}, u^x_{y:\rho}, a^x_{y:\rho}$ represent the degrees of belief, disbelief, uncertainty, and the base rate (*a priori* probability in the absence of evidence) respectively. Opinions are formed on the basis of positive and negative evidence. The variables $r^x_{y:\rho}$ and $s^x_{y:\rho}$, represent the number of positive and negative past observations of x about y respectively, and can be used by x to obtain an opinion about y as follows:

$$b^x_{y:\rho} = \frac{r^x_{y:\rho}}{r^x_{y:\rho} + s^x_{y:\rho} + 2}, d^x_{y:\rho} = \frac{s^x_{y:\rho}}{r^x_{y:\rho} + s^x_{y:\rho} + 2}, b^x_{y:\rho} = \frac{2}{r^x_{y:\rho} + s^x_{y:\rho} + 2} \tag{6}$$

An opinion's probability expectation value computed using Equation 6, can be used by x as a measure of y's trustworthiness with respect to ρ.

$$\tau^x_{y:\rho} = b^x_{y:\rho} + a^x_{y:\rho} \times u^x_{y:\rho} = \frac{r^x_{y:\rho} + a^x_{y:\rho} + 2}{r^x_{y:\rho} + s^x_{y:\rho} + 2} \tag{7}$$

The base rate parameter $a^x_{y:\rho}$, also known as the relative atomicity, represents *a priori* degree of trust x has about y giving accurate report about ρ before any evidence has been received. The parameter determines how uncertainty shall contribute to the computed expectation value. The default value of $a^x_{y:\rho}$ is 0.5 [2], which means that before any positive or negative evidence has been received, both outcomes are considered equally likely.

The trust value for a group G_i is based on past interactions with members of the group, and computed as a function of the trust of the individual members encountered from the group.

$$\tau^x_{G_i:\rho} = \frac{\sum_{y \in G_i} \tau^x_{y:\rho}}{|G_i|}. \tag{8}$$

2.3 Learning Diversity

We define Diversity as a function $\Delta : 2^A \rightarrow G$ that maps the feature set and past reports (behaviour) of agents to a set of groups. We take as a working assumption, that there may be some correlation between the features of an agent and its behaviour. Where this exists, we could exploit information from observable features of agents, as well as evidence from their past behaviour to build a model of diversity. Diversity learning may be carried out in two stages: the first stage involves an attempt at disambiguating what metrics lead to a better stratification of the population of agents. The best metric in our estimation is one that produces the highest feature-behaviour correlation, such that the probability of agents in the same group giving similar reports is maximized. We refer to this correlation as *group behaviour*. In the second stage, the learned metric is employed to partition agents into semi-homogenous subgroups. We focus here on the process of group formation by assuming a learned metric.[1] We employ a clustering mechanism that incorporates a feature threshold fT, and a behaviour threshold bT in order to control the formation of clusters. There are various clustering techniques that can be used for this purpose. In this work, we employ the hierarchical clustering [7] as an algorithm of choice because it is well-known, and allows us to cluster into a set of groups the cardinality of which we do not know in advance. The clustering process is illustrated in Algorithm 1.

In the first stage of the clustering, each agent is regarded as belonging to a separate cluster, and the two clusters with the shortest 2-norm (Euclidean) normalised feature distance are then merged to form a new cluster. In the second stage, the merging of clusters continues as in the first stage, until either all the agents are assigned to a single cluster, or the $\delta^x_{G_i^F}$ of a potential group exceeds the feature threshold fT. At each stage of clustering, the expected behaviour $\delta^x_{G_i^B}$, of a potential group, is validated against the bT threshold based on available evidence, to ensure the behaviour threshold is not exceeded.

Our clustering approach has some interesting characteristics. It imposes restriction on group membership for *outlier* agents. An outlier agent has features matching that of a particular group, but with a non-conforming behaviour to the group. Our model regards such agents as belonging to singleton groups pending evidence suggesting otherwise. In line with this, unknown agents start off in singleton groups even though their features may be matched to any of the existing groups, until there is sufficient evidence supporting their group membership.

2.4 Sampling and Evidence Aggregation

The DM model offers rich context from which an aggregation set may be derived. In the general case, an aggregation set is made of candidates randomly selected from different groups, from which evidence may be drawn in order to form an opinion. However, depending on the specific requirements of a task, richer contexts could be explored using the learned model of diversity. For instance, the cost and risk assessments of a potential transaction [4], may serve to inform the

[1] Relaxing this assumption is left to a future work.

Algorithm 1 Hierarchical clustering algorithm for group formation using feature and behaviour criteria.

Require: A set of agents A
Require: A feature and behaviour based similarity thresholds, fT,bT
Require: A feature and behaviour similarity functions, δ_{G^F},δ_{G^B}
allocate each agent in A as a single cluster
let C be the set of clusters
continue $\Leftarrow$ **true**
while *continue* **do**
  continue $\Leftarrow$ **false**
  for all $X, Y \in C$ **do**
    compute the between-cluster similarity $\delta_{G^F_{(X,Y)}}$, such that $\delta_{G^B_{(X,Y)}} < bT$
  end for
  if $fT < \delta_{G^F_{(X,Y)}}$ **then**
    $Z \Leftarrow X \cup Y$, where $\delta_{G^F_{(X,Y)}}$ is the minimum
    remove $X$and Y from C
    $C \Leftarrow C \cup Z$
    continue $\Leftarrow$ **true**
  end if
end while
return C

sampling process. Members in a group comprising of trustworthy agents may be favoured, for example, over agents in less trustworthy groups in a high-risk transaction. Also, in situations where the cost associated with sampling from specific groups of agents (e.g. groups of experts) exceeds a budget, groups of less knowledgeable agents may be considered, who in combination may provide a sufficiently similar service. We consider the random selection of one representative candidate from each of the groups to form an aggregation set. Provided the likelihood of agents in each of the groups behaving in a similar manner is relatively high, then evidence from the set may be considered a sufficient representation of the entire population. We do not suggest this to be the only approach for sampling, but only that it demonstrates one possible realisation of our model, which we have used in our evaluation. Other sophisticated sampling techniques may be explored to meet specific requirements.

We denote by S the aggregation set comprising of candidates sampled from groups in G. $G_{i,last}$ is the number of agents in G_i. Further, we define $G_i(l)$ as the index of the lth element in G_i, for $l = 1, ..., last$. A candidate $y \in G_i$ is selected to be added to S, by choosing a random integer $z \in [1, last]$.

The DM model does not limit the chances of unknown agents being sampled by simply assigning them to an existing group having members with similar features. Every unknown agent, as already mentioned is regarded as belonging to its individual group, until there is sufficient evidence to classify it differently. This approach prevents a stereotypical treatment for such agents with regards to group membership, but gives each agent in this category a fair chance of being *heard*. There are benefits to this: in the first instance, a benevolent agent sharing similar features with a group of malicious agents will not be automatically labelled malicious, when there is no concrete evidence suggesting such. Also, a group of benevolent agents will not risk the abuse of its reputation by

malevolent agents who, for example, may be *masquerading* by presenting similar features [6].

An evaluator x, combines reports from agents in the aggregation set $S \subseteq A$, in order to arrive at an opinion about ρ. Every agent s, in the aggregation set, has its report weighted by the subjective trust value τ, assigned its group $G_{i(s)}$ by x. We use an *a priori* trust for unknown agents which is often set at 0.5 in literature [2]. The combined evidence is computed as:

$$E^x_\rho = \frac{\sum_{s \in S} \mathcal{R}_{y,\rho} \times \tau^x_{G_{i(s)}:\rho}}{\sum_{s \in S} \tau^x_{G_{i(s)}:\rho}}. \tag{9}$$

3 Evaluation

In this section we describe experiments conducted to evaluate (in simulation) the performance of the Diversity Model. The aim of the experiments is to study the effect of group behaviour and deception on an aggregation result, and how these mechanisms may be exploited to limit the number of agents queried for evidence. We describe the methods adopted in the experiments, and present our results and discussion. The factors taken into account in the evaluation are: the predictive accuracy of the model to some ground truth, and the proportion of agents in the population sampled for evidence. We compare our approach to other approaches such as sampling the entire population of agents, randomly sampling a number of agents, and sampling based on the trustworthiness of the agents (in this instance we compare the performance of our model to the trust computation engine used in Beta Reputation System [5]).

3.1 Experimental Setup

Our experiments are based on a simulation testbed which models a logical network of agents as defined by their features. The environment consists of 100 sources and one evaluator. The evaluator relies on evidence obtained from the sources to evaluate a subject of interest. Our network is connected, with *undirected* edges from each node to its neighbours. The network fitness is based on the distance between features of agents, such that nodes that are highly similar gravitate towards each other. For simplicity, we assume that similarity is defined on the same feature dimension for all agents. Agents possess incomplete knowledge, and therefore, report with some amount of uncertainty. We simulate this phenomenon by drawing each agent's report from a Gaussian distribution $\mathrm{N}(0, 1)$. However, agents closer to each other report in a similar manner. To simulate this we have each agent broadcast its report at each sampling phase to all its one-hop neighbours. Each node maintains a buffer of reports received from its neighbours. At each sampling phase, a node reports following $\mathrm{N}(0, 1)$. However, if there are corresponding reports in a node's buffer for the same sampling phase,

a node alters its report to reflect a conformity to reports of its neighbours, with only a slight deviation. In this way, we define the underlying logical network we wish our evaluator to identify and exploit. The experimental parameters are listed in Table 1.

Table 1. Experimental Parameters

Parameter	Value	Parameter	Value
#Information sources	100	# evaluators	1
feature threshold (fT)	0.8	behaviour threshold (bT)	0.9
degree of nodes	8	report distribution	$N(0, 1)$

3.2 Experiments and Results

We consider different experimental conditions to analyse the effects of group behaviour and malicious agents on the aggregation result. We indicate the number of agents used by our model in each case to arrive at a decision. Each evaluation condition was initialised with random models of the information sources. 100 runs were conducted in 10 rounds for each case and the mean of the runs reported.

Effect of Group Behaviour In this experiment, we analyse the effect of the increasing rate of group behaviour in the population in the predictive accuracy of the evaluator agent. In real world scenario, subgroups may arise, for instance, as a result of agents having similar expertise, being constrained by organisational policies, or by a coordinated act of collusion by different agents. This may be regarded as a kind of deception, since agents exhibiting some of these traits may not be reporting their perception objectively. We linearly increase the percentage of conformity to group behaviour from 0% to 100%. In each case, agents that conform to group behaviour are selected randomly from the population. Figure 2 shows the effect of group behaviour on the aggregated result. The deviation from the ground truth is reported in each case. Although each agent reported with some uncertainty, given incomplete knowledge on a subject the evaluator took advantage of the large number of agents queried, and was able to make better predictions because the noise in the aggregated reports cancelled out, leaving only reliable reports. However, when the rate of group behaviour increased, as expected, the reports also became skewed in favour of opinions held by different subgroups, leading to lower accuracy in predictions. An evaluator in such circumstance may no longer benefit from sampling large number of agents, as each new report may only be a repetition of an already sampled opinion.

In Figure 2, the performance of the DM model is compared against other approaches. The metric we are specifically interested in is the performance of our Diversity Model compared to other approaches. We considered an approach based on sampling all agents in the population, which we refer to as the *baseline*. Also considered are models based on the trustworthiness of agents, and

the random selection of agents. The same number of agents as that sampled by the DM model was employed to select agents when using the trust model and the random selection respectively. The trust model involved sampling the most trustworthy agents in the population. Our first observation is that all the models begin very well when there were no group behaviour (at 0%). The predictions made were closer to the ground truth, with the baseline model slightly outperforming the other approaches. Also nearly as many agents as the baseline approach were queried by the DM model. This may be regarded as the worst case, where no compelling evidence could be established for the formation of informative groups. However, when evidence of group behaviour in the agent population emerged, our model was able to exploit this to reduce the number of agents queried, while still making better predictions. The performance of the trust model was worse off, undoubtedly caused by the uncertainty in the reports of the agents in each sampling phase. Specifically, when there were no expert or malicious agents in the system, the trust model was unable to learn any useful pattern in the reliability of agents, and therefore consistently made poor choices. The approach based on random selection of sources is an uninformed selection strategy, which leans much on chance. This, as observed is likely to perform poorly in environments where there are defined patterns of behaviour among agents. This observation is encouraging, as it demonstrates the efficacy of the DM model in guiding decision making rather than relying on chance.

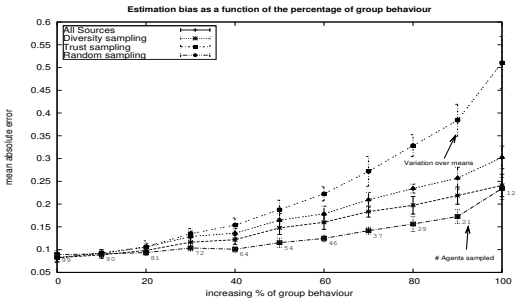

Fig. 2. Effect of group behaviour in aggregated result

Effect of Malicious Sources Until now, we have discussed the scenario in which agents reported objectively based on incomplete knowledge. However, in real life settings, agents may not always behave benignly. There may be incentives for agents to lie, leading to distorted reports aimed at subverting the system. In this section we consider an attempt by malicious agents to systematically distort the aggregation result, by reporting a value different from their observation. Our goal was to determine robustness of our model with varying degrees of deception. As before, we compare the performance of the Diversity Model against the baseline approach, random selection, and trust filtering. The baseline, as in previous case, involved sampling all the agents in the population.

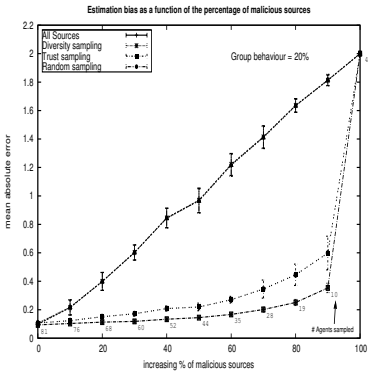

Fig. 3. Effect of malicious sources with 20% group behaviour

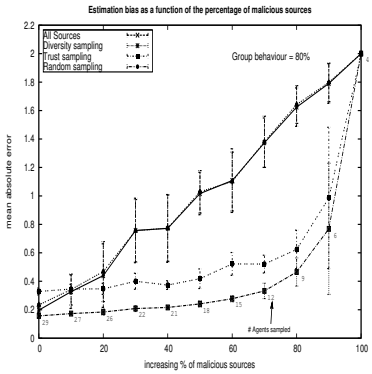

Fig. 4. Effect of malicious sources with 80% group behaviour

In the experiment, malicious agents report with a distribution that is different from normal agents. Deceptive reports were drawn from a Gaussian distribution $N(2, 0.01)$, with an attempt at distorting the aggregation result. However normal agents continued to report according to $N(0, 1)$. We gradually increased the percentage of deception in the system from 0 to 100, and observed the effect when group behaviour was kept constant at 20% and 80%, respectively. In each case, performance of the four approaches was considered, and the number of agents queried at each instance was also recorded. The number of agents queried by the diversity model in each case is captured in the result against the diversity sampling curve, as shown in Figure 3 and Figure 4.

An interesting observation, given this set of experiments is the notable improvement in the predictive accuracy of the trust model. Although still outperformed by the Diversity Model, the trust approach performed significantly better than the the baseline and random approaches. The trust model, in this instance, was able to learn from its experience when the activity of malicious agents became obvious and stable in the system. The effect of deception is all the more highlighted and amplified in the system when considered in parallel with group behaviour. An analogy of this could be drawn to agents in a social network, where agents may be influenced based on the kind of social group they belong to, or *who* they listen to. This kind of phenomenon is referred to as rumour spreading in the literature [10]. As observed, the DM model could still make better predictions because it was able to adjust the weights for different types of sources.

4 Conclusion and Future Work

We have presented a framework for selecting sources of reputational evidence, in a way that guarantees reliable decisions from small groups of individuals. The approach presented in this work is oriented towards resource-constrained envi-

ronments where querying of information sources is costly. However, the proposed approach could also be extended to other environments to facilitate selection of interaction partners, and to guard against deception, especially the more coordinated attempts of collusion. Where hidden networks defining group behaviour exist in the population, our model is able to exploit this in order to limit the number of information sources sampled while still remaining robust to deception. Where a naïve approach of evidence aggregation would perform poorly under these conditions, our model shows positive results that outperforms classical trust approach.

This work exploits features and perceived behaviour of agents in order to cluster them into groups. We intend to exploit richer information contexts such as domain ontology and provenance in future research in order to form better clusters. For simplicity, this work assumes that malicious agents behave in a consistent manner. We hope to incorporate more complex deception models in order to evaluate robustness of our model in other scenarios. Although we considered a very simple sampling mechanism for the selection of information sources, we intend to incorporate richer sampling techniques, aimed at satisfying different information needs of an application.

Acknowledgements. This research was supported by the Petroleum Technology Development Fund (PTDF) Nigeria, Overseas Scholarship Scheme (OSS).

References

1. A. Jøsang. Artificial reasoning with subjective logic. In *Proceedings of the Second Australian Workshop on Commonsense Reasoning*, 1997.
2. A. Jøsang, R. Hayward, and S. Pope. Trust network analysis with subjective logic. *Proceedings of the 29th Australasian Computer Science Conference*, 48:85–94, 2006.
3. A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.
4. A. Jøsang and S.L. Presti. Analysing the relationship between risk and trust. *Trust Management*, pages 135–145, 2004.
5. A. Jøsang and R.Ismail. The beta reputation system. In *Proceedings of the 15th Bled Electronic Commerce Conference*, pages 17–19, 2002.
6. C. Karlof and D. Wagner. Secure routing in wireless sensor networks: Attacks and countermeasures. *Ad hoc networks*, 1(2-3):293–315, 2003.
7. O.D. Richard, E.H. Peter, and G.S. David. Pattern classification. *A Wiley-Interscience*, pages 373–378, 2001.
8. J. Surowiecki and M.P. Silverman. The wisdom of crowds. *American Journal of Physics*, 75:190, 2007.
9. W.T.L. Teacy, J. Patel, N.R. Jennings, and M. Luck. TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.
10. B. Yu and M.P. Singh. Detecting deception in reputation management. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 73–80, 2003.