POMDPs provide a natural model for sequential decision making under uncertainty. The main advantage that the POMDP scheme brings to the seller selection problem is that it enables optimal trade-off of the expected benefit and cost of obtaining more information, aiming to maximize the total utility of the buyer.

Given $I$ advisors that can be queried about the quality of $J$ sellers, each SALE POMDP agent can be described in terms of states, actions, observations and rewards as follows.

**States.** A state contains the quality levels of each seller ($high, low$), each advisor ($trustworthy, adversarial, random$) and the status of the transaction with the seller ($not\_started$, $satisfactory$, $unsatisfactory$, $gave\_up$, $finished$).

**Actions.** The model knows the following types of actions: 1) $seller\_query_{ij}(SQ_{ij})$, ask advisor $i$ about quality of seller $j$; 2) $advisor\_query_{ii'}$ ($AQ_{ii'}$), ask advisor $i$ about quality of advisor $i'$; 3) $buy_j$, buy from seller $j$; 3) $do\_not\_buy$ ($DNB$), decide not to buy from any seller in the market.

**Transitions.** We assume that when taking a query action, the state does not change. When taking $buy_j$ and $DNB$ actions, the state will always transition to a terminal state, i.e., $buy_j$ actions may result in a successful ($sat$ = $satisfactory$) or unsuccessful ($sat$ = $unsatisfactory$) transaction and $DNB$ will result in $sat = gave\_up$. Transition probabilities to terminal states give the *definition* of quality levels. Generally, chances of transition to $satisfactory$ should be high on buying from 'high quality' sellers.

**Rewards.** There is small cost for the ask actions. A reward is associated with a successful transaction, otherwise a penalty is levied. There is a penalty for taking $DNB$ action, when in fact there is a seller of high quality, otherwise there is a reward for this action. Once the terminal states are reached, no further rewards are given.

**Observations.** When a *query* action is performed, the agent will receive an observation based on the set of discriminated quality levels. After $SQ_{ij}$ action, the agent receives an observation $o \in \{good, bad\}$, corresponding to the quality of seller $j$. After $AQ_{ii'}$ action, it gets an observation $o \in \{trustworthy, untrustworthy\}$, corresponding to the quality of advisor $i'$. On transition to a terminal state, it receives the observation $ended$.

**Observation Function.** It specifies the likelihood of receiving an observation given the current state and the action that led to this state. There is no a priori correct way to specify the observation probabilities. Similar to the transition probabilities for buy action, probabilities for the observation function *define* the meaning of different trust levels. In general, the idea is that trustworthy advisors give more accurate and consistent answers than untrustworthy ones.

**Initial State Distribution.** We assume a uniform belief over the quality levels, but a different initial belief can also be obtained as a result of previous interactions. We will also assume an infinite horizon problem.

The SALE POMDP model works by improving its beliefs over the quality levels of sellers and advisors by querying advisors about the quality of sellers/other advisors in the system, until it is sure that it has identified a seller with sufficient quality. If $b(s)$ specifies the probability of a state $s$ (for all $s$), we can derive $b'$ an updated belief after taking action $a$ and receiving observation $o$ using Bayes' rule,

$$b'(s') = \frac{\Pr(s', o|b, a)}{\Pr(o|b, a)} = \frac{\Pr(o|a, s')}{\Pr(o|b, a)} \sum_{s} Pr(s'|s, a)b(s) \quad (1)$$

Also, the belief updates are performed such that they correlate the state factors in meaningful ways, e.g., observing *good* after $seller\_query_{ij}$ will give more weights to states where the seller is $high$ quality and the advisor is $trustworthy$, and less weights to states where the seller is $low$ quality. We also represent the SALE POMDP in factored form to improve its scalability and use symbolic Perseus [6] as the POMDP solver for the experiments. Extensive evaluation on the ART testbed demonstrates that SALE POMDP balances the cost of obtaining and benefit of more information more effectively, leading to more earnings, than traditional trust models. Experiments also show that it is more robust to deceptive advisors than a previous POMDP based approach.

# 3. FUTURE WORK

For the future work, we plan to extend our current trust models, improving their robustness and enhancing their applicability to different real-world scenarios. More specifically, the current biclustering based approach works well, if the users assign equal importance to the various evaluation criteria. However, in real-world, users may have different subjectivity for evaluation. In this case, we will extend our approach by considering the importance of different criteria. Also, in real-world, users may not provide ratings to all the criteria, every time. We will deal with such scenarios, by using correlation information between criteria, to predict the ratings for the missing criteria. Also, we plan to improve the complexity of the biclustering approach using approximation strategies, e.g., random initialization of bicluster members.

Our current SALE POMDP model optimally selects sellers by modeling seller and advisor trustworthiness on a single-criterion. We will extend the model to a multi-criteria scenario, where a seller is selected based on its trustworthiness on a number of criteria. We will also include more detailed advisor models (e.g., differentiating its trustworthiness in providing opinions about sellers and other advisors) to improve the robustness of the approach.

We also consider to apply the SALE POMDP model to the routing problem in wireless sensor networks. Here, the SALE POMDP agent will select a suitable (trustworthy) sensor node to route packets. The neighboring sensor nodes will be assigned beliefs based on their ability to route packet data (based on multiple-criteria). The SALE POMDP agent will work by improving its beliefs over the quality levels of its neighbors by querying information until it has identified a suitable neighboring sensor to route packets. The optimal policy suggested by the SALE POMDP model will balance the expected benefit of obtaining more information about the sensor nodes against the cost (in terms of energy consumption) of obtaining this information.

# 4. REFERENCES


[1] Irissappane, Athirai A., Jiang, Siwei, Zhang, Jie: A testbed to evaluate the robustness of reputation systems in e-marketplaces. In AAMAS, 2014.

[2] Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision support systems, 43(2), 618–644, 2007.

[3] Irissappane, Athirai A., Jiang, Siwei, Zhang, Jie: A framework to choose trust models for different e-marketplace environments. In IJCAI, 2013.

[4] Reece, S., Rogers, A., Roberts, S., Jennings, N.R.: Rumours and reputation: Evaluating multi-dimensional trust within a decentralised reputation system. In AAMAS, 2007.

[5] Irissappane, Athirai A., Jiang, Siwei, Zhang, Jie: A biclustering-based approach to filter dishonest advisors in multi-criteria e-marketplaces. In AAMAS, 2014.

[6] Irissappane, Athirai A., Oliehoek, Frans A., Zhang, Jie: A POMDP based approach to optimally select sellers in electronic marketplaces. In AAMAS, 2014.