Supporting Intelligence Analysts with a Trust-based Question-Answering System

Maaike de Boer[*], Peter-Paul van Maanen[*] and Gerard Vreeswijk[†]
[*]Department of Cognitive Systems Engineering, TNO Human Factors
P.O. BOX 23, 3769 ZG Soesterberg, The Netherlands
Email: maaike.deboer@tno.nl, peter-paul.vanmaanen@tno.nl

[‡]Department of Information and Computing Sciences, University of Utrecht
Princetonplein 5, 3584CC Utrecht, The Netherlands
Email: gv@uu.nl

***Abstract*—Intelligence analysts have to work in highly demanding circumstances. This causes mistakes with severe consequences, which is the reason that support systems for intelligence analysts have been developed. The support system proposed in this paper assists humans by offering support that improves their performance, without reducing them in their freedom. This is done with a trust-based question answering system (T-QAS). An important part of T-QAS are trust models which keep track of trust in each of the agents gathering information. Using these trust models, the system can support the intelligence analyst by: 1) helping to decide which agents are trusted enough to receive questions, 2) providing information about the reliability of each of the sources used, and 3) advising in making decisions based on information from possibly unreliable sources. An implementation of last two capabilities of T-QAS is evaluated in an experiment in which participants perform a decision making task with information from possibly unreliable sources. Results show that the proposed T-QAS support indeed helps participants to improve their performance. We therefore expect that future intelligence analyst support systems can benefit from the inclusion of T-QAS.**

I. Introduction

Intelligence analysts supervise, coordinate and participate in the analysis, processing and distribution of intelligence in defense organizations [1]. Intelligence analysts have to gather and provide timely and relevant information to decision makers. If crucial information is missed or not provided timely, friendly fires or wrong decisions can have severe consequences. An example is an attack by American helicopters which killed twenty-three civilians [2]. In this example a team of operators of drones failed to pass crucial information about the makeup of a crowd. Reports that indicated that the group included children were missed due to a massive amount of data and time constraints.

The biggest challenges for intelligence analysts are the uncertain, ambiguous and complex information and the amount of information. In order to deal with uncertain information due to unreliable sources, trust models can be used. Trust models indicate the estimated reliability of the information from a source. Lower trust in certain sources indicates less reliable information and higher trust more reliable information.

In this paper, we propose that a support system T-QAS that uses the above mentioned trust models can support users.

II. Theoretical Background

As intelligence analysts have to work in highly demanding circumstances, it has been subject to many studies, proposing solutions to challenges, such as JIGSAW [3] and AICMS [4] for ambiguous questions, [5] and [6] for the cumbersome data gathering process and [7] and [8] for analyzing data.

One of the bigger challenges nowadays is the incapability of analysts to cope with (large amounts of) uncertain, ambiguous and complex information. Also several solutions to support analysts in dealing with this have been proposed [9], [10], [11] and [12]. However, it remains unclear what the added value of these proposed systems is. At the same time, in [13] it is even claimed that dealing with high uncertainty, ambiguity and complexity of information will likely remain beyond the capabilities of software tools for some time.

The possibility of information overload also is a challenge for intelligence analysts [14]. According to [9] approximately 10,000 messages per hour are received and only 15,000 messages can be scanned a day [10]. [15] states that heavy information load affects performance, measured as accuracy or speed, negatively. With too much information people may have difficulties with identifying relevant information or relationships between details and the overall perspective.

Not only intelligence analysts deal with (large amounts of) uncertain, ambiguous and complex information, but personnel from other organizations and common people have to deal with it. For example, in searching information using search engines, participating in online social media, and reading one's e-mail. Search engines such as Google, Yahoo!, Bing and Ask use word matching and ranking techniques to select relevant information. Question-answering websites such as Answerbag, Answers, ChaCha and Quora use reputation and votes to display the most probable answers to a question.

To summarize, it is clear that intelligence analysts have a difficult job. They will greatly benefit from support systems if they are able to tackle one or more of the challenges mentioned. But, as was stated by [13], one of the bigger challenges, to deal with high uncertainty, ambiguity and complexity of information, is expected to be a too difficult endeavor for the near future. At the same time question-answering systems, such as those using reputation, have proven to be helpful in providing users the tools to better determine trustworthy information from larger amounts of possibly uncertain data. In this study we therefore propose a *trust-based question-answering system (T-QAS)* to support intelligence analysts.

III. Support Model

A support model is proposed that can be used to support a *question agent* (s.a. intelligence analysts) in retrieving answers from possibly unreliable *answer agents*. This process of retrieving answers on questions is called the *question-answering process* and a system that regulates this process is called a *question-answering system (QAS)*. The application of the support model (*Trust-based QAS* or *T-QAS*) is expected to help the question agent to cope with (large amounts of) uncertain, ambiguous and complex information in question-answering systems.

In Fig. 1 the support model used for this study (T-QAS), which regulates the question-answering process between the question agent and answer agents, is shown. Within T-QAS three processes are available: *Question Delivery*, *Trust Management* and *Selection of Answer(s)*. The next sections explain the function and processes of these three processes.

A. Question Delivery

The process *Question Delivery* is provided with a *smart question* from the question agent. A smart question is a question with addition of meta-information, such as location, expiration date, security level and domain of the question as proposed by [6]. The smart question together with key words and answer-category is sent to the processes *Trust Management* and *Selection of Answer(s)*. From the process *Trust Management*, information about answer agents such as location, availability and trust is sent. With this information appropriate answer agents are selected. A delivery list of these appropriate answer agents is sent to the process *Trust Management* and these answer agents receive the question.

B. Trust Management

In the process *Trust Management* information about agents, questions and received answers is stored in a database. This information is fed by the delivery list and formatted question from the process *Question Delivery*, agent information from the answer agents and feedback from the process *Selection of Answer(s)*. The trust in each of the agents is managed by independent trust models. These independent trust models estimate the *trust value* of an agent through the following formula (partially adopted from [16]):

$$T_i(t) = T_i(t-1) \cdot \lambda_i + (E_i(t) \cdot (1 - \lambda_i)) \tag{1}$$

where $T_i(t-1)$ is trust in agent i at time $t-1$, λ_i is a decay parameter, $E_i(t)$ is the experience for agent i at time t.

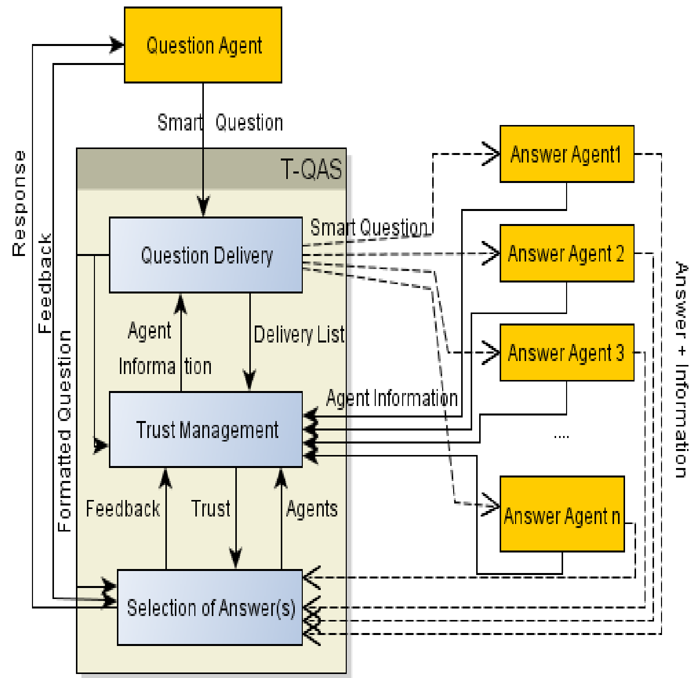

Fig. 1. Proposed general model

In the present study, trust values are based on experiences of T-QAS with each answer agent. The value of an experience is based on feedback about the correctness of the answer of the agent with the following formula:

$$E_i(t) = ((r-1)/(m-1)) \cdot -1 + 1 \tag{2}$$

where r is the rank number of the answer of agent i, m is amount of answers.

With this formula the experience with the agent(s) who gave the worst answer is 0 and the experience with the agent(s) who gave the best answer is 1. If one answer is given, the experience is 1. The influence of past experiences is varied with decay parameter λ.

C. Selection of Answer(s)

In the process *Selection of Answer(s)* the answers (with information such as location and time of answering) received from the answer agents are analyzed in order to form a response to the question. The answers are tiled based on their similarity, in this case two answers are similar if and only if the strings of characters of the answers are identical. From each of the agents that provided an answer, trust is retrieved from the process *Trust Management* by sending the *ID* of the agents to this process. By using trust in each of the agents, answers can be ranked. The estimated value of an answer is calculated as follows:

$$EV_a = 1 - \prod_{i=1}^{n} (1 - T_i(t)) \tag{3}$$

where n is the amount of agents who gave answer a and $T_i(t)$ is trust in agent i at time t.

In the above formula a combination of ranking by *confidence* and ranking by *popularity* is used. Ranking by confidence takes trust in each of the answer agents into account. Ranking by popularity takes the amount of agents that gave a similar answer into account.

By using the estimated values of each answer, the answers are ranked. The ranked answers are then sent to the question agent as a response to the question. The question agent gives feedback about the response such as "confirmed" or "rejected", which is received by the process *Selection of Answer(s)* and sent to the process *Trust Management*.

IV. Method

The processes *Trust Management* and *Selection of Answer(s)* of T-QAS are evaluated in a controlled lab experiment.

A. Participants

Thirty-six participants aged between eighteen and thirty-five ($M = 24.2$, $SD = 4.3$; 15 male, 21 female) with a higher education level participated in the experiment as paid volunteers. No special training or military experience was required. Participants were not dyslectic, had no concentration problems and no RSI.

B. Task

Just like real intelligence analysts, participants had to form an accurate intelligence "picture" of the situation in a fictive scenario which took place in an area called "Lowland". However, the task had to be simplified, because participants had no experience with analyzing intelligence. Within the task first part of the process, from posing a smart question to tiling the answers in the process *Selection of Answer(s)*, is simulated.

The task was set up as a quiz in which participants had to rank possible answers, given by one or more answer agents, to a question according to their estimated soundness. For each question context information was provided to give background information related to the question. No further analysis or implications of this ranking had to be made by the participants.

After ranking the answers, feedback about the correct ranking was given. The feedback, in its turn, could result in an idea of the participant's task performance, but could also result in experiences about the reliability of the different answer agents. These experiences helped in the calibration of trust in each of the answer agents (both for the participant and T-QAS). Because feedback was given immediately after the ranking, the quiz can be seen as an accelerated version of the task of real intelligence analysts, where analysis and feedback on performance are alternated much less quickly.

For data gathering purposes, after each question participants also had to indicate their trust in each of the answer agents explicitly. Both ranking and indication of trust had to be done within thirty seconds in order to simulate high demanding circumstances.

C. Conditions

The experiment employed a repeated measures, within-subjects setup, with four conditions: *human alone (H)*, *team human-system 1 (T1)*, *team human-system 2 (T2)* and *system alone (S)*. In the H-condition the participant had to perform the task alone and in the S-condition an implementation of T-QAS had to perform the task alone. The S-condition was done off-line (without participants). H and S were baseline conditions. In the T1-condition T-QAS provided advice about the trust in each of the sources. In the T2-condition T-QAS provided advice about the trust in each of the sources and advice about the ranking. In all conditions but S, participants had to fill in the trust ratings and ranking themselves.

D. Hypotheses

Several hypotheses about the conditions are expected to be accepted. First, task performance of the human-system teams will be higher than both task performance of the human alone and of the system alone. Because the support model has no mechanism for natural language understanding, it is expected that a human is better at using the context information for a decision. The system, on the other hand, is hypothesized better in objectively estimating trust values, because of the process *Trust Management*. A cooperation of both human and system is therefore expected to improve performance:

Hypothesis 1: Task performance in T1 and T2 is higher than task performance in H.

Hypothesis 2: Task performance in T1 and T2 is higher than task performance in S.

An important factor determining the previous two hypotheses is performance in H. Good performance in H may result in a lower improvement of performance in T1 and T2, because T-QAS cannot add as much value to good performers compared to poor performers. Poor performers may benefit more from T-QAS, because they are not able to perform the task well themselves.

Hypothesis 3: The difference in task performance between H and both T1 and T2 is higher for poor task performers than for good task performers.

V. Results

A. Hypothesis 1 and 2

Repeated Measures ANOVA showed a statistically significant effect on task performance ($F(3, 105) = 7.252$; $p < .0005$). Post hoc tests using Bonferroni correction revealed a significant difference ($p < .0005$) between task performance in the T1-condition ($.287 \pm .592$) and task performance in the S-condition ($-.200 \pm .000$) and a significant difference ($p < .0005$) between task performance in the T2-condition ($.172 \pm .477$) and task performance in the S-condition ($-.200 \pm .000$). No other significant differences were found (see Fig. 2). Hence Hyp. 1 cannot be accepted. Hyp. 2, however, is accepted, because task performance in both the T1- and T2-conditions is significantly higher than in the S-condition. Adding the "human in the loop" thus indeed improve performance. Trust estimation significantly

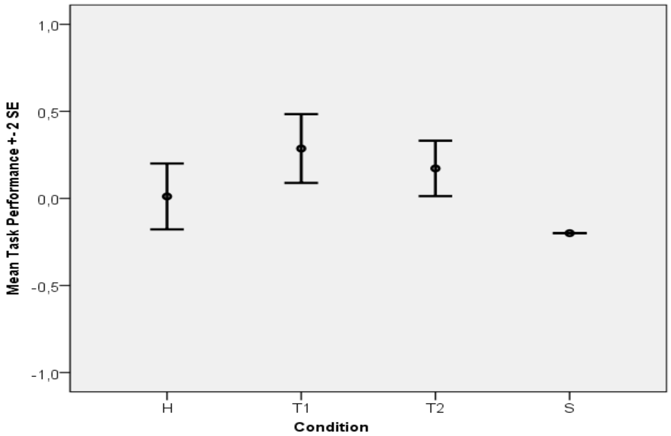

Fig. 2. Task Performance

improves ($p = .017$) with advice (T1-condition; $.220 \pm .999$) as compared to without advice ($-.253 \pm 1.081$). Adding the system to the human thus also improve performance, but the difference in task performance between the H-condition and T1- and T2-conditions is not significant.

B. Hypothesis 3

The results regarding good ($n = 16$) and poor performers ($n = 20$) indicate that poor task performers perform significantly better in both the T1- ($p = .021$; $.231 \pm .635$) and T2-condition ($p = .007$; $.106 \pm .454$) as compared to the H-condition ($-.363 \pm .323$) and good task performers do not perform significantly better in the T-conditions compared to the H-condition. Hyp. 3 can be accepted. Furthermore, Hyp. 1 is accepted for poor performers, because poor performers perform significantly better with support from T-QAS compared to without support, as mentioned before. On the other hand, good performers perform not significantly worse with support from T-QAS as compared to performing alone, even if T-QAS performs significantly worse than the good performers ($p < .0005$). Support has thus no negative effect on performance.

VI. Discussion, Conclusion and Future Work

The support system proposed in this paper assists humans by offering support that improves their performance. This is done with a trust-based question answering system (T-QAS). An important part of T-QAS are trust models which keep track of trust in each of the agents gathering information. Using these trust models, the system can support the intelligence analyst by: 1) helping to decide which agents are trusted enough to receive questions, 2) providing information about the reliability of each of the sources used, and 3) advising in making decisions based on information from possibly unreliable sources. An implementation of last two capabilities of T-QAS is evaluated in an experiment in which participants perform a decision making task with information from possibly unreliable sources.

Experimental evaluation proved that T-QAS improved task performance (poor performers) and estimation of the reliability of the sources. Task performance of good performers was not improved, but also not decreased, even if T-QAS performs significantly worse than the good performers. T-QAS has thus no negative effect on performance. In order to further improve performance, performers have to be trained in when to use T-QAS. Poor performers can use T-QAS in more cases than good performers, so good performers have to know in which case T-QAS has to be used. Furthermore, T-QAS can be improved in order to get performance on the level of good human performers. This could be done with using more semantic information.

Humans in cooperation with T-QAS also have an improved performance as compared to the task performance of the system (T-QAS) alone. Indeed this shows that both humans and the system are benefited when they cooperate with the other.

References

[1] M. Powell, "Intelligence analyst," December 2012. [Online]. Available: http://www.prospects.ac.uk/intelligence_analyst_job_description.htm

[2] T. Shanker and M. Richtel, "In new military, data overload can be deadly," January 16 2011, new York Times.

[3] M. B. Cook and H. S. Smallman, "Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 5, pp. 745–754, 2008.

[4] P. Gonsalves and R. Cunningham, "Automated ISR Collection Management System," in *Proc. Fusion 2001 Conference*, 2001.

[5] S. McGrath, D. Chacón, and K. Whitebread, "Intelligent mobile agents in the military domain," in *Fourth International Conference on Autonomous Agents*, 2000.

[6] J. van Diggelen, M. Grootjen, E. Ubink, M. van Zomeren, and N. Smets, "Content-based design and implementation of ambient intelligence applications," 2012.

[7] Q. Chen and Q.-n. Deng, "Cloud computing and its key techniques," *Journal of Computer Applications*, vol. 29, no. 9, p. 2565, 2009.

[8] A. Bahrami, J. Yuan, P. R. Smart, and N. R. Shadbolt, "Context aware information retrieval for enhanced situation awareness," in *Military Communications Conference, 2007. MILCOM 2007. IEEE.* IEEE, 2007, pp. 1–6.

[9] G. M. Powell and B. Broome, "Fusion-based knowledge for the objective force," DTIC Document, Tech. Rep., 2002.

[10] M. P. Jenkins and A. M. Bisantz, "Identification of human-interaction touch points for intelligence analysis information fusion systems," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on.* IEEE, 2011, pp. 1–8.

[11] E. Patterson, D. Woods, D. Tinapple, E. Roth, J. Finley, and G. Kuperman, "Aiding the intelligence analyst in situations of data overload: From problem definition to design concept exploration," *Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report, ERGO-CSEL*, 2001.

[12] E. A. Newcomb and R. J. Hammell II, "Examining the Effects of the Value of Information on Intelligence Analyst Performance," in *Proceedings of the Conference on Information Systems Applied Research ISSN*, vol. 2167, 2012, p. 1508.

[13] R. V. Badalamente and F. L. Greitzer, "Top ten needs for intelligence analysis tool development," in *proceedings of the 2005 international conference on intelligence analysis*, 2005.

[14] W. Matthews, "Data surge and automated analysis: The latest isr challenge," January 2012, a Briefing from GBC: Industry Insights. [Online]. Available: http://www.emc.com/collateral/analyst-reports/the-latest-isr-challenge-insights-gbc.pdf

[15] M. J. Eppler and J. Mengis, "A framework for information overload research in organizations," *Insights from organization science, accounting, marketing, MIS, and related disciplines. paper*, vol. 1, p. 2003, 2003.

[16] P.-P. van Maanen, "Adaptive support for human-computer teams: Exploring the use of cognitive models of trust and attention," Ph.D. dissertation, Vrije Universiteit Amsterdam, 2010, 72–99.