Seeing the World through an Expert's Eyes: Context-Aware Display as a Training Companion

Marc T. Tomlinson, Michael Howe, and Bradley C. Love

Department of Psychology, University of Texas at Austin, Austin Texas
{marctomlinson,Michael.howe,Brad_love}@mail.utexas.edu

Abstract. Responsive Adaptive Display Anticipates Requests (RADAR) is a domain general system that learns to highlight an individual's preferred information displays, given the current context. Previous studies with human subjects in a video game environment demonstrate that RADAR is an effective cognitive aid. RADAR increases situation awareness and reduces cognitive load by anticipating and providing task relevant information. Additionally, because RADAR's fit to a user's behavior encapsulates the user's situation-driven information preferences, RADAR also excels as a descriptive and predictive assessment tool. Here, we focus RADAR as a training aid. We test the hypothesis that novices can benefit from training under a RADAR model derived from an expert's behavioral patterns. The results indicate that novices exposed to an expert's information preferences through RADAR rapidly learn to conform to the expert's preferences.

1 Introduction

When boarding an airplane, a furtive look into the cockpit reveals a vast array of dials, displays, and controls. The expert pilot can make sense of this array of options and can appreciate when each instrument is relevant to operating the aircraft. For example, expert pilots know which gauges are relevant to different phases of flight. In this article, we discuss a context-aware approach to information display named Responsive Adaptive Display Anticipates Requests (RADAR). RADAR learns to highlight the situation-relevant information by observing the user.

We discuss how RADAR can be used to analyze and describe individual differences in information needs, as well as present evidence that RADAR can be used to allow novices to see the world through the eyes of an expert. When training under an expert's RADAR model, we find that novices' information use patterns converge to those of the expert from whom the model was derived.

Related work has attempted to predict user information needs by correctly attributing intentions, beliefs, and goals to the user. Plan recognition models tend to subscribe to the Belief-Desires-Intention framework [1]. This line of work relies on knowledge-based approaches for user modeling and encoding insights from domain-specific experts [2]. These approaches can involve identifying a user's subgoals through task-analysis [3]. Once a user's beliefs, intentions, and goals are understood, a display can be adapted appropriately [2].

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 668–677, 2009.

Alternatively, instead of focusing on identifying the internal state of the user, some approaches rely on input from domain experts. For example, human experts can label episodes that can serve as training instances for machine learning models that prioritize display elements [4]. Alternatively, input from human experts can be used to build expert systems or Bayesian models to prioritize displays [5]. This approach relies on extensive input from human experts, and the ability of those experts to introspect on the reasons for their performance.

Our approach diverges from the aforementioned work. Rather than prescribe which information source a user should prioritize, RADAR highlights the information a user would select if the user searched through all possible options. This approach may be preferable in domains where it is unclear what is normative. Unlike work in plan recognition, we sidestep the problem of ascribing and ascertaining the user's internal mental state. Instead, RADAR learns to directly predict a user's desired display from contextual (i.e., situational) features (see Figure 1).

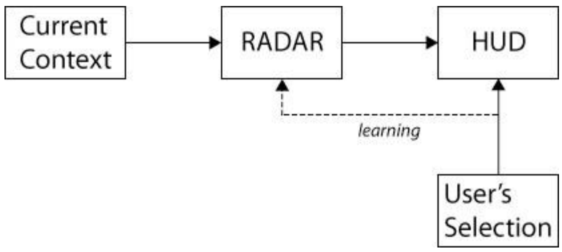

Fig. 1. RADAR takes as input the current context (e.g. Recent game history) and outputs its preferred display to the HUD. The user (e.g., the game player) can override RADAR's choice. Such corrections serve as learning signals to RADAR and increase the likelihood that RADAR will select the user's preferred display in similar situations in the future. Over time, RADAR approximates the information preferences of a specific user, allowing the user to offload the task of selecting the relevant information source (i.e. display) from numerous options.

Furthermore, RADAR emphasizes the benefits of continuous learning by the display, as opposed to preprogrammed interfaces [6]. Adopting a learning approach to an adaptive display has a number of positive consequences, including the ability to take into account individual differences across users [7]. Another positive consequence is that minimal input from subject matter experts is required to build a system. Like other context-aware applications that adopt a keyhole approach [8,9], our approach infers a user's preferences without interfering with or directly querying the user [10]. Interfaces that highlight recently selected menu items follow a similar logic [11], though our approach is more open ended in terms of possible predictors and learnable relationships from predictors to display preferences.

Whereas previous work with RADAR [12], which we review below, has evaluated RADAR as a cognitive aid and assessment tool, the current experiment evaluates RADAR's promise as a training companion. The current experiment asks whether RADAR can speed the novice to expert transition by exposing novices to the display preferences of an expert (i.e., train under an expert's RADAR model).

Our approach has some potential benefits. In some domains, knowledge can be directly elicited from experts and simple instruction can boost novice performance to expert levels [13]. However, in many domains, an expert's knowledge is not accessible by self-report [14,15]. In practice, training methods for novices that rely on both direct instruction and pattern recognition methods work best [16]. Indeed, having novices view expert solutions is more effective than even providing corrective feedback [17,18]. These findings suggest that there is more to expertise than what an expert can report verbally from introspection. This conclusion is not surprising given that human learning is subserved by multiple learning systems, only some of which are accessible to introspection and verbal report [19].

Our training goal is to make novices conform to the information preferences of experts in order to improve task performance. In service of this goal, standard verbal instructions, coupled with RADAR, provide users with training opportunities that can engage both verbal and non-verbal learning systems. A RADAR system trained based on an expert's performance data is a potentially powerful training tool for novices. Such a tool might allow a novice to become sensitive to the information preferences of an expert while performing the relevant task. Importantly, such a training system does not require eliciting explicit knowledge from an expert and can impart expert knowledge that is not readily verbalized. We present an experiment that investigates how novices trained with an expert RADAR system perform compared to those trained under a control model.

1.1 RADAR's Operation

RADAR is designed to operate in task environments in which the user must select which display among numerous displays to monitor. For example, we evaluate RADAR in an arcade game environment in which players select which of eight possible displays to show on a Head-Up Display (HUD). RADAR takes as input the current context (e.g., recent game history) encoded as a feature vector and outputs to the HUD the display it predicts the user wishes to view (See Figure 1). The user is free to override RADAR's choice. RADAR learns from the user's acceptance or rejection of its display choices and over time converges to selecting the displays the user desires. Alternatively, RADAR can observe and learn to mimic a user's display preferences offline.

RADAR employs a two-stage stochastic decision process at every time step. In the first stage, RADAR estimates the probability that a user will update the HUD given the current context. When the sampled probability from the first stage results in a display update, RADAR proceeds to the second stage (otherwise the current display remains unchanged). In the second stage, RADAR estimates the probability distribution for the next display choice given the current context, and samples this probability distribution to select the next display.

The motivation for the two-stage approach is both computational and psychological. Separating display prediction into two stages improves RADAR's ability to predict display transitions. The same display currently desired is highly likely to be desired in 250 ms. This constancy would dominate learning if both stages were combined. The second stage's focus on display transitions allows for improved estimation of these relatively rare, but critical, events.

Psychologically, the first stage corresponds to identifying key events in a continuous (unsegmented) environment, whereas the second stage corresponds to predicting event transitions. To make an analogy to speech perception, people segment the continuous speech stream into words (akin to RADAR's first stage) in the absence of reliable acoustical gaps between words [20]. Akin to RADAR's second stage, people anticipate which word (i.e., event) is likely to follow given the preceding words [21].

One view is that event segmentation serves an adaptive function by integrating information over the recent past to improve predictions about the near future (see [22], for a review). In support of this view, individuals who are better able to segment ongoing activity into events display enhanced memory [23]. People's judgments of event boundaries are reliable [24] and tend to show high agreement with others [25]. For example, two people watching a person make a peanut butter and jelly sandwich will tend to agree on the steps involved. These two people will also both segment off surprising or unexpected events, like the sandwich maker dropping the sandwich on the floor.

The probability distributions associated with both stages (event segmentation and event prediction) are estimated by simple buffer networks [26]. Buffer networks represent time spatially as a series of slots, each containing the context (e.g., game situation) at a recent time slice, encoded as a feature vector. The buffer allows both ongoing events and events from the recent past to influence display prediction. Despite their simplicity, buffer networks have been shown to account for a surprising number of findings in human sequential learning [27]. At each time step, weights from the buffer are increased from activated features to the display option shown in the HUD, whereas weights to the other display options are decreased. Over time, this simple error correction learning process approximates a user's information preferences. RADAR's weights can be used to assess individual differences and user performance. Details of RADAR's implementation are discussed elsewhere [12].

1.2 Previous Work

Previous experiments with RADAR have shown that it is an effective cognitive aid [12]. RADAR model trained from the aggregated data of several domain experts have been shown to be better at highlighting important information, than control models which only display information using the same base rates as the experts. Furthermore, when users are assisted in making display choices by an individually tailored RADAR model, their performance is better than when they are solely responsible for controlling the display.

RADAR has also demonstrated its usefulness as an assessment tool. By comparing model fits between expert and novice players, RADAR revels that there are significant differences in the pattern of information usage between the two groups. Furthermore, a novice player's success in the game is predicted by how well an expert's RADAR model fits their display choices.

As previously discussed, RADAR's first stage is hypothesized to be akin to scene segmentation. The first stage of the model learns to predict when a user chooses to update the display. The first stage is independent of the second stage which chooses the successor display. As discussed above, cognitive load and change in the environment are greatest at event boundaries (the very times one would want RADAR to update the display). Results from subjects playing our video game without RADAR support the notion that RADAR's first stage is akin to event segmentation. For an example expert subject, Figure 2 shows the mean number of feature changes in the environment over a ten second window before and after display channel changes. Figure 2 suggest that change is greatest at display updates, as they are at event boundaries. This consequence of people's interactions with the environment may explain why RADAR is effective as a cognitive aid. Interestingly, there is a lag between the change in features and the actual time of the channel change. We believe this lag arises because people are slow to respond to the changing event due to concurrent demands in the video game task.

While experts show individual differences in which channels they choose to view at any given moment, they have a remarkable level of agreement on when they should change display channels. This is reflected by assessing the fit of models created under one individual to the actual data provided by another individual. We see that the first stage of a given expert fits all experts almost as well as it fits the individual that created the model. In contrast, the second stage shows a marked decrease in fit to other experts compared to the individual's own data. Individual differences arise in information preferences (stage 2), but not in event segmentation (stage 1).

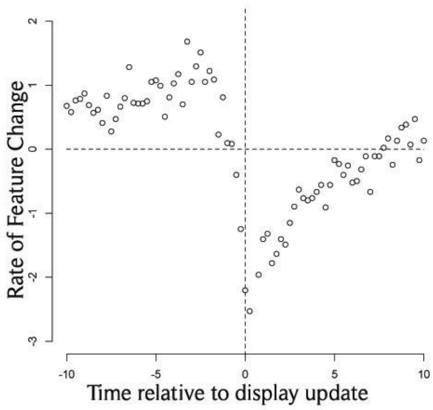

Fig. 2. Feature change (a proxy for change in the environment) is plotted in z-scores.Time on the horizontal axis (in seconds) is relative to display updates (negative is prior to update, positive is post update). The plot indicates that feature change is greatest prior to a display change. These results support the notion that display updates are akin to event boundaries.

2 Training Novices with Expert Displays

We explore the possibility that novices can learn to sample the environment like experts following training under an expert's RADAR model. An expert's RADAR model captures the expert's situational information preferences. Thus, a novice training under an expert's model can potentially benefit from the expert's perspective. Potential advantages of this approach include exposing novices to expert knowledge that is not readily verbalizable and providing expert insight in the context of performing the relevant task. To test this hypothesis, we had novices play in the tank environment with display choices determined by either an expert or control RADAR model. Subjects alternated between having displays provided to them (either by an expert or control model) and choosing displays manually. We compare manual information selections for these two groups of novice subjects.

2.1 Methods

RADAR Training Models
Subjects trained under various RADAR models. These models were built from fitting three subjects from a previous study. In the previous study, subjects played for 11 hours, controlling the display manually for the entire period. We created an expert model for each of three subjects based on the last three hours of play. Rather than use all the features available in the game, we determined the features that subjects actually entertained. This was done by evaluating subsets of all possible features using cross validation [28]. In cross validation, including features that are not psychologically real decreases performance on the data held out to test for generalization. These fits provided our three *expert* models. We then created a set of control models. The control models were specified to choose the channel that its corresponding expert model is least likely to choose. The control models also change the channel when the expert model is least likely to change, but importantly maintains the same rate of changes over time. The first stage of the control models were also decoupled from the environment, so that channel changes would not be indicative of the underlying event structure.

Design and Procedure
Thirty students were recruited from the University of Texas at Austin and were paid and given class credit for participation. The subjects played in the tank environment for three 1.5 hour sessions over a one-week period. Subjects were randomly assigned to either the expert or control condition. Subjects in the expert condition were randomly assigned to train under one of the three expert models, whereas subjects in the control condition were randomly assigned to one of the three control models. Participants in both conditions alternated between five-minute blocks of manually controlling the display and having their RADAR control the display. Which RADAR model controlled the display is the only difference in procedure across subjects.

2.2 Results

Fit of Subjects' Manual Play Data by their RADAR Model used in Training
One question is whether subjects conform to the RADAR model that they trained under. Here, we assess the probability that a subject's RADAR model correctly predicts the subject's display choices under manual play. Expert condition subjects' display choices were more accurately predicted (.25 vs. .13) by their RADAR model than were control subjects, $\mathbf{F}(1,18) = 36.10$, $\mathbf{p} < .001$. There was also a main effect of session (i.e., improvement over time), $\mathbf{F}(1,18)=4.61$, $\mathbf{p}<.05$. Importantly, there was an interaction of these two factors, such that expert condition subjects came to conform more to their model over session, $\mathbf{F}(1,18) = 8.70$, $\mathbf{p} < .01$. The left panel of Figure 3 shows that the interaction is driven by gains made by subjects in the expert condition.

Evaluating Novices' Progression Toward Expert-Like Performance

The previous analysis demonstrates that subjects come to conform to their display model, particularly subjects in the expert condition. One key question is the degree to which people come to behave like experts. To answer this question, we used all three expert models to predict each subject's display choices and averaged the fit of the three models to get a measure of expert-like behavior. There is a main effect of convergence over session, **F**(1,18) = 13.20, **p** < .01, although there is no main effect for training condition. Importantly, there is an interaction such that subjects in the expert condition become more expert-like over sessions than do subjects in the control condition, **F**(1,18) = 4.96, **p** < .05. In fact, subjects in the control condition show no significant difference in expert fit (.24 vs. .25) between the first and last session, **t** < 1, whereas subjects in the expert condition improve (.23 vs. .27) significantly, **t**(15)=3.20, **p** < .01. These results suggest that subjects in the expert condition become more expert-like in their information selections, whereas subjects in the control condition did not. Mere experience on task does not appear to guarantee the emergence of expert-like behavior in terms of display choice.

Display Updating as a Function of Training-Mode

The previous analysis focused on display choice, RADAR's second stage. One question is whether differences between expert and novice condition subjects exist in display update (i.e., when to change the display), RADAR's first stage. Analyses indicate a main effect for converging to the average expert fit over session, **F**(1,18) = 13.01, **p** < .01, but no effects of training condition were observed. The change patterns of subjects in both the expert and control conditions were fit equally well by the expert models (see the right panel of Figure 3). This result is highly suggestive that event segmentation, in contrast to display choice, is something that is learned by experience on task and is not facilitated by training under an expert model. The lack of an interaction between session and training condition also agrees with previous work that finds that different expert models' fist stages have higher inter-agreement than do their the second stages.

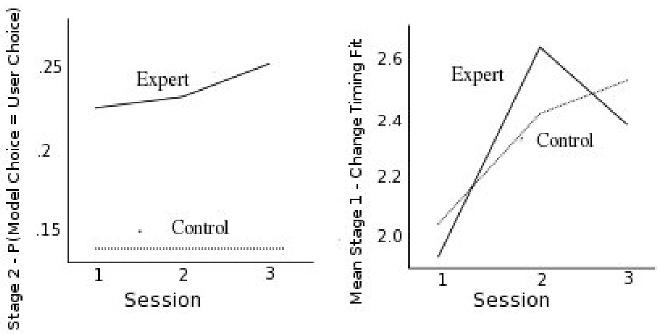

Fig. 3. The left panel shows that subjects' RADAR models better predict their display choices in the expert condition and this advantage grows with training. The right panel shows that expert and control condition subjects make display updates in roughly the same fashion. The average fit of the three expert models yields similar results for both conditions with regards to display update.

3 Discussion

Advances in information technology make large quantities of information available to human decision makers. In this deluge of information, finding and selecting the relevant piece of information imposes a burden on the user. This burden is particularly onerous for novices within complex, dynamic environments. RADAR is a domain-general system that learns to approximate the information search processes of an individual user.

RADAR contains two stages. The first stage is akin to event segmentation and determines when to update the display. The second stage determines, given a display update, which display to select. Previous work demonstrates that RADAR improves user performance [12]. Here, we report results that indicate that subjects who train under an expert's RADAR model learn to choose displays consistent with the second stage of expert RADAR models.

The same result did not hold for display update, embodied in RADAR's first stage. In the case of display update, subjects trained under expert and control RADAR models both converged to expert-like updates over time. This result supports previous research [12, 24, 25] demonstrating reliability and agreement among people perceptions of event boundaries. Mere task experience appears sufficient to identify basic events in a novel domain, although the same is not true of determining proper display choice.

The above results should not be taken to indicate that subjects are slaves to the model they trained under and the task environment. While subjects did converge to the display model they were trained under, subjects in the expert condition appeared to generalize their knowledge broadly. These subjects showed increased convergence over time to the second stage of all three expert models. In other words, exposure to one expert's view of the world encouraged more general expert-like behavior, rather than behavior that was only closely coupled to the particular training model. Control subjects did not show this systematic improvement in fit to all expert models. While control subjects might display idiosyncratic behaviors that agree with one expert model, they did not learn behaviors that were consistent across experts.

Overall, our results suggest that related training methods should prove successful in expediting the transition from novice to expert-levels of performance. Using an expert's RADAR model to train novices sidesteps several thorny issues. RADAR's fit of an expert quantifies the expert's action patterns (avoiding the limitations and effort involved in self-report) and provides a means to communicate this expertise to a novice in a task-situated manner.

There is a lot more research to be done before such training methods can be perfected. Although not reported above, expert RADAR models differed greatly in how well they fit each subject. One expert model fit particularly well, achieving the best fit for 18 of the 30 subjects. Interestingly, the model of the highest performing expert fit the subjects in our study the worst, with only 5 subject being well fit by it. One important challenge is determining which expert model is most beneficial for each novice at each stage in training. One possibility is that novices will vary in terms of which expert model is best. Hopefully, RADAR's formal approach will allow for best practices to be determined.

Acknowledgments. This research was supported in part by AFOSR Grant FA9550-07-1-0178 and NSF CAREER Grant #0349101 to B.C.L.

References

1. McTear, M.F.: User modeling for adaptive computer systems: a survey of recent developments. Artifcial Intelligence Review 7, 157–184 (1993)
2. Goodman, B.A., Litman, D.J.: On the interaction between plan recognition and in-telligent interfaces. UMUAI 2, 83–115 (1992)
3. Yi, W., Ballard, D.: Behavior recognition in human object interactions with a task model. In: AVSS, pp. 64–64 (2006)
4. St. John, M., Smallman, H.S., Manes, D.I.: Assisted focus: Heuristic automation for guiding users' attention toward critical information. In: Proc. of the 1st Inter. Conf. on Augmented Cognition (2005)
5. Horvitz, E., Barry, M.: Display of information for time-critical decision making. In: Proc. of Conf. on Uncertainty in AI, pp. 296–305 (1995)
6. Mantyjarvi, J., Seppanen, T.: Adapting applications in mobile terminals using fuzzy context information. In: Mobile HCI, pp. 95–107 (2002)
7. Schneider-Hufschmidt, M., Kuhme, T., Malinowski, U.: Adaptive User Interfaces: Principles and Practice. North-Holland, Amsterdam (1993)
8. Albrecht, D.W., Zukerman, I., Nicholson, A.E.: Bayesian models for keyhole plan recognition in an adventure game. User Modeling and User-Adapted Interaction 8(1-2), 5–47 (1998)
9. Teevan, J., Dumais, S.T., Horvitz, E.: Personalizing search via automated analysis of interests and activities. In: Proceedings of SIGIR, pp. 449–456 (2005)
10. Jones, G.J.F., Brown, P.J.: Context-Aware Retrieval for Ubiquitous Computing Environments. Springer, Heidelberg (2004)
11. Gajos, K.Z., Everitt, K., Tan, D.S., Czerwinski, M., Weld, D.S.: Predictability and accuracy in adaptive user interfaces. In: Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1271–1274 (2008)
12. Love, B.C., Jones, M., Tomlinson, M.T., Howe, M.: Learning to predict informa-tion needs: Context-aware learning to predict information needs: Contex-aware display as a cognitive aid and an assessment tool. In: Proceedings of the 27th annual SIGCHI conference (2009)
13. Biederman, I., Shifrar, M.: Sexing day-old chicks: A case study and expert sys-tems analysis of a difficult perceptual learning task. Journal of Experimental Psychology: Learning, Memory, & Cognition 13, 640–645 (1987)
14. Feldon, D.F.: The implications of research on expertise for curriculum and pedagogy. Educational Psychology Review 19, 91–110 (2007)
15. Hoffman, R.R., Crandall, B., Shadbolt, N.: Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. Human Factors 40, 254–276 (1998)
16. Ark, T.K., Brooks, L.R., Eva, K.W.: The benefits of flexibility: The pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. Medical Education 41, 281–287 (2007)
17. Kluger, A., DeNisi, A.: Effects of feedback intervention on performance: An his-torical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin 119, 254–284 (1996)

18. Gonzalez: Decision support for real-time, dynamic decision-making tasks. Organizational Behavior and Human Decision Processes 96, 142–154 (2005)
19. Ashby, F., Maddox, W.: Human category learning. Annual Review of Psychology 56, 149–178 (2005)
20. Saffran, J.R.: Statistical language learning: Mechanisms and constraints. Current Directions in Psychological Science 12, 110–114 (2003)
21. McRae, K., Spivey-Knowlton, M.J., Tanenhaus, M.K.: Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language 38, 283–312 (1998)
22. Kurby, C.A., Zacks, J.M.: Segmentation in the perception and memory of events. Trends in Cognitive Sciences 12, 72–79 (2008)
23. Zacks, J., Speer, N., Vettel, J., Jacoby, L.: Event understanding and memory in healthy aging and dementia of the Alzheimer type. Psychology & Aging 21, 466–482 (2008)
24. Speer, N., Swallow, K., Zacks, K.: Activation of human motion processing areas during event perception. Cognitive, A_ective & Behavioral Neuroscience 3, 335–345 (2003)
25. Newton, D.: Foundations of attribution: The perception of ongoing behavior. In: Harvey, J. (ed.) New Directions in Attribution Research, Erlbaum, Mahwah (1976)
26. Cleeremans, A.: Mechanisms of implicit learning: Connectionist models of sequence processing. MIT Press, Cambridge (1993)
27. Gureckis, T., Love, B.C.: Transformational vs. statistical complexity in sequential learning. Cognitive Science (in press)