UTILITY EVALUATION OF MODELS

ABSTRACT

In this paper, we present three case studies of utility evaluations of underlying models in software systems: a user-model, technical and social models both singly and in combination, and a research-based model for user identification. Each of the three cases used a different approach to evaluating the model and each had challenges to overcome in designing and implementing the evaluation. We describe the methods we used and challenges faced in designing the evaluation procedures, summarize the lessons learned, enumerate considerations for those undertaking such evaluations, and present directions for future work.

Categories and Subject Descriptors

H.5.2 [**Information Systems**]: Information Interfaces and Presentations – *evaluation/methodology, user-centered design.*

General Terms

Design, Human Factors.

Keywords

Utility evaluation; abstract models; evaluation design.

1. INTRODUCTION

The human-computer interaction community has made us all aware of the importance of evaluating different aspects of new systems to ensure that we develop systems that are both usable and useful. The visualization and visual analytics communities are still struggling with user evaluations and with just cause. Evaluation of software systems involving visualizations is significantly different from the evaluation of a more typical software application designed for general use. Munzner [10] suggested that evaluation of visualization software needs to be accomplished at four levels: the problem level, the abstraction level, the encoding and interaction level, and the algorithmic level. Each of these levels takes a different type of evaluation technique and different metrics. In particular, Munzner notes that the abstraction levels need to be evaluated with real users doing real work in order to obtain information on the utility of the system.

In this paper, we focus on evaluations at the abstraction level. Munzner describes this level as mapping domain problems into operations feasible in the computer science terminology. We believe that underlying models incorporated into the software should also be included at this level and that they should be evaluated by the targeted user population. We disagree with Munzner on the timing of these evaluations. We believe that these models should be evaluated as early as possible, and hence, evaluation methodologies that substitute for "real users doing real work" are needed.

In this paper, we present three evaluations of very different types of models conducted at different phases of development using different evaluation techniques. The descriptions illustrate these conclusions:

- The earlier the evaluation occurs, the more creativity is needed to develop an appropriate evaluation methodology.
- Metrics for evaluation are established based on the objectives of the model.
- Data collection and analysis for the metrics need to be factored into the design and implementation of the evaluation procedure.
- Pilot tests are essential to ensure that the evaluation implementation will yield reliable data.

Currently most evaluations of models focus on verification and validation [9]. Evaluations of utility from the user's perspective are rare in the literature and hence, most likely from practice as well. One example of an example of an evaluation of utility was done after an extensive study of the work of automotive engineers [16]. This study showed that the visual analytics software provided the engineers with novel views of the data and facilitated finding more issues with less effort than previously. We present our work in this domain to encourage others to pursue early evaluations of underlying models.

1.1 Model Evaluation

Models are abstractions used to explain a system's behavior, forecast or predict events, or aid decision-making [9]. While there are many combinations and different types of models, the three described in this paper illustrate the basic types of models. Today, models are used for many purposes, including describing large computer systems (system models), modeling the ecology, modeling people's understanding and thinking processes (cognitive models), modeling individuals' risk of disease, and

BELIV'14, November 10, 2014, City, Paris, France.
Copyright 2014 ACM 978-1-4503-3209-5/14/11…$15.00
http://dx.doi.org/10.1145/2669557.2669562

modeling users' understanding of computer interfaces (mental models). As models are abstractions, they are imperfect to various degrees of user satisfaction [9]. The issue is to determine if a given model is useful by a user population for a given task. We term this type of evaluation as "utility evaluation."

In the examples described in this paper, we work with underlying models: a user-model (explanatory), technical and social models both singly and in combination (prediction), and a research-based model that suggests user pathways for investigating identity (decision making). It should be noted that although the evaluation of underlying models is not restricted to visual analytics environments, the presence of such models adds yet another level of complexity to the evaluation of visual analytics software.

2. DESCRIPTION OF THE PROJECTS AND MODELS

2.1 Novel Intelligence from Massive Data

The Novel Intelligence from Massive Data (NIMD) project [11] was funded by the Advanced Research and Development Agency (ARDA)[1] starting in 2002 and had a number of components. One was a user-modeling component that was designed to better understand what information was useful to any given analyst and to use that knowledge to help the analyst obtain more relevant search results. The NIMD program was unique in that it was designed with user evaluation in mind. Early in the program, NIMD funded the design and development of the Glass Box [3], which captured the computer interactions analysts made as they went about doing analysis. This data capture software was used by a number of analysts hired to perform analysis at an unclassified level. Thus, we had a baseline that was useful for comparisons once various NIMD components were testable.

2.2 Technosocial Predictive Analysis Initiative

The Pacific Northwest National Laboratory (PNNL) sponsored the Technosocial Predictive Analysis Initiative (TPAI) in 2008-2009. This work blended technical models (weather, power and energy consumption, etc.) with social models (terrorists, cultural models, etc.) to enhance predictions of future states. The models operated under a gaming environment in which users interacted with the models to understand how changes would affect decisions. The user evaluation for this initiative was quite complex and required four different levels of evaluation: models, knowledge encapsulation, a gaming environment, and an evaluation of the final system. Due to funding issues, the actual evaluations were never completed, but in Section 3.2, we discuss the proposed evaluation design. [12,14].

2.3 SuperIdentity

The SuperIdentity (SID) project is a joint project with PNNL and six universities in the United Kingdom: Bath, Dundee, Kent, Leicester, Oxford, and Southampton [1]. One aim of the project is to provide intelligence and law-enforcement services with a greatly enhanced ability to identify, and attribute information to, individuals and groups in both natural and cyber domains. SID deviates from existing approaches in that the work incorporates contributions from an expansive spectrum of scientific domains, including biometric, psychological, behavioral, and online indicators of identity, enabling a broader set of identity measures to be considered than ever before. As a way of bringing all the various technical contributions together, researchers at Oxford developed a model of inferences; that is, the SID model contains the interplay between facets of an individual or group identity [7]. An enriched identity can be created by taking a set of known elements of identity and inferring new, previously unknown elements of identity from these known elements. For example, if an individual's hand length is known, then an individual's gender can be determined (with some degree of certainty) as men's hands are typically bigger than women's [8]. The model contains a large number of these inferences and hence, pathways can often be found from a given known attribute to a desired attribute of identity. A typical task might be to find a person's real name given the person's username on a social network site. SID offers ways for law enforcement and intelligence analysts to use a broad spectrum of information in identity. It should be noted that no profiles have been created using the SID model; the model is just a model and contains no inferred data.

3. DESIGNING UTILITY EVALUATION FOR MODELS

The three major steps in designing a utility evaluation are determining the metrics to use, finding the appropriate end users, and implementing the evaluation, including deciding which materials to give to users and how to collect and analyze the data.

3.1 Determining the Metrics

In doing utility evaluation of the models, the first step is to clarify the objective of the model and determine the appropriate metrics to use. Unlike typical usability evaluations, there are no standard metrics such as efficiency, effectiveness, and user satisfaction. In the NIMD case, we want to determine if the modeling of the user used in the search routine provides more relevant information than not modeling the user. In the TPAI situation, we want to understand if the parameters available to the user to modify the model are sufficient to appropriately change the model's behavior and to provide the user sufficient information to understand different possibilities that can occur. We were also interested in determining the quality of the end users' predictions contrasted to the face-to-face exercises that are currently used for investigating complex situations. In the SID evaluation, we want to determine if users will use more attributes from the cyber, psychological, and biometric domains to investigate identity than they currently do. As we implement the evaluation methodology, we need to ensure that we collect the necessary data to derive qualitative and quantitative measures to understand if the objective has been achieved.

3.2 Determining the End Users

For each evaluation, it is important to determine the most appropriate users. NIMD was designed for intelligence analysts. As getting time from working intelligence analysts is extremely difficult, we were fortunate to find a number of naval reservists who had been analysts and who were given time to participate in our evaluations.

For TPAI, we started by working with academics who were experienced in the technical domains that were being modeled in the initial versions of the project. We intended to find academics for the social models as well. We planned to use both academics and analysts (or surrogates) for analyzing the combined models.

As with the TPAI evaluation work, several types of end users are necessary to evaluate the SID model. Currently, some evaluation work has been conducted with general end users. In evaluations with the general public, we were interested in seeing if the model was useful to them in showing what information could be inferred

[1]

by postings on social networks [5]. An evaluation was carried out in the United Kingdom with a group of teenagers to see what pathways they thought were interesting and also what pathways they found surprising or did not believe were possible [6].

In this paper, we focus on the design of an evaluation for law enforcement officers and intelligence analysts using the SID model. While this evaluation has not yet been conducted, we have conducted three pilot studies. Based on the lessons learned from those, we are currently redesigning the actual evaluation. These changes will be discussed in the next section.

3.3 Implementing the Evaluation

The last and most difficult step is to implement the evaluation. This process involves assessing what materials to provide to the users, determining how to collect the measures, and ensuring that the model is appropriately presented and explained to the users.

Because it is important to do the evaluation as early as possible, users will not be able to actually perform a task using the software. Possible techniques include paper-based tasks or a Wizard of Oz system [4].

In a paper-based system (now often replaced with PowerPoint slides), user interactions need to be anticipated and the result shown either on paper or a slide. The evaluator must keep track of the various choices and show the correct result to the user based on the choice made. Another paper-based approach is to show the user a single image of the model and to ask questions to determine the user's understanding of the model. It is necessary to ensure that the presentation doesn't affect the evaluation of the model.

In a typical Wizard of Oz evaluation, the user thinks she is interacting with the system, but one of the evaluators plays the part of the system. The evaluator must have a script describing how to react to each anticipated user interaction so that all users have the same treatment.

Collecting the data for analysis is also an issue. Of course, we can give users a questionnaire after the session is over to ask about their impressions, but it is often desirable to collect more quantitative and less subjective information as well. Having users describe their understanding of the model and then coding their descriptions for accuracy is one possibility. If the models have various parameters that can be adjusted, then we can ask specific questions about what would happen if the parameters were set to specific values. If the user is asked to think aloud as she selects various options and is shown various results, then coding these responses in some agreed-upon classification scheme is necessary. Multiple coders need to code them independently, compare their results, and work out any instances they disagree on. Coding from listening to audio is time-consuming as coders may need to listen to some sections multiple times.

For any evaluation, it is necessary to conduct several pilot tests to ensure that the evaluation procedure is workable. In the following examples, we will describe what we did and describe some of the problems we encountered.

3.3.1 NIMD User Model Evaluation

The NIMD model experiment was used to determine if the user model produced more relevant documents than searching with no user model. As the model was already implemented, we were able to perform searches using it, and hence, we could do a comparison. To implement this evaluation, we used the Interests, Preferences and Context model (IPC) [13] embedded in an information retrieval system and a traditional information retrieval system based on keyword retrieval. Three former professional analysts participated in the evaluation. They used information that had been collected and distributed by the Center for Nonproliferation Studies (CNS, Sept. 2001 distribution) [2]. The two systems were run side-by-side, and analysts were not aware of which system had an embedded user model. Analysts were given 10 scripted queries to use to ensure that we were able to make valid comparisons. After each query returned results, the analysts were asked to examine the top 10 documents returned from each system and rank them as relevant or not relevant. Not only did the system with the embedded user model return more relevant documents then the traditional information retrieval system but also the analysts all used different approaches in their search strategies. The documents returned were customized to fit each analyst's search style.

While this evaluation went extremely smoothly and the evaluation returned positive results, it would have been better to have done the evaluation earlier. However, an early evaluation would have been more difficult to design and implement. As this user model was a piece of a much larger project, the funders considered evaluating this piece separately and before the integration process as a positive step.

3.3.2 TPAI Model Evaluation

As mentioned earlier, this evaluation was not conducted due to budgetary constraints. However, we did spend considerable time planning how to evaluate the models. The system had two types of models: technical and social. Our goal was to evaluate each type of model separately, and then to evaluate the two types of models together. A summative evaluation would be done when the models were incorporated into the gaming environment.

Analytic exercises are commonly used to analyze world situations, resulting in a report back to the agency who commissioned the exercise. Experts in the appropriate areas are asked to participate in face-to-face exercises lasting anywhere from 4 hours to 4 days. These exercises are expensive to develop and to run because of their complexity and the amount of time needed from various experts. In addition, the lessons learned in the exercise have to be captured and conveyed to those responsible for the decisions. As the TPAI system would be a replacement for these face-to-face exercises, the hypotheses were that the TPAI system would compare favorably to the face-to-face systems in the following ways:

- Necessitate fewer experts
- Allow individual analysts to appropriately use the blended models without necessarily having expertise in both social and technical domains.
- Consume less time
- Allow analysts to make fewer assumptions about modeled systems
- Help analysts produce the same or better level of understanding
- Help analysts produce the same or better insights
- Help analysts produce a debrief of the same or better quality
- Provide references and simulation data for supporting report recommendations.

The plan for evaluating the technical models was to work with the subject matter experts who built the technical models in each domain and devise a sketch of the model along with inputs and outputs. These evaluations would be done by other subject matter experts using a given scenario and would result in expert reviews of the models and the allowed parameter inputs and outputs. The same procedure was to be followed in the social models, again working with experts in that area. The questions to be addressed in each evaluation were whether the models captured the essential parameters and whether changing these parameters gave sufficient information to explore the decision space.

The combined models were to be evaluated using two experts for each evaluation, one expert in the technical subject matter and another in the social subject matter. A second version of this evaluation would be with analysts who, while not experts in the technical or social models, were quite knowledgeable. The questions for the integrated models were more complex than evaluating the models alone.

Questions included:

- What impact did the specific social features as blended with the technical features have on usefulness?
- Did the features of the social and technical models blend well? Were there unexpected direct or side effects? Were results overly biased toward one model over the other? If so, could this be adjusted by the user?
- Were there any features of the social or technical models that were impacted by the blending?

Unfortunately, these evaluations were not carried out. We were eager to determine if our descriptions and diagrams of the models along with the input parameters and outputs were clear enough to give the analysts sufficient understanding. Our intent was to modify these descriptions and diagrams based on the expert reviews to use in the actual evaluations.

3.3.3 SuperIdentity Evaluation

Although this evaluation has not yet been conducted, we have carried out three pilot studies. We describe what was done in the pilot evaluations, the problems we found, and our redesign for the actual evaluation.

The Identity Map (Figure 1), developed by PNNL and Oxford is a visualization of the model allowing the user to explore its capabilities in more depth. It has been extremely useful in helping develop scenarios for our evaluations, as we can easily see if paths contain some of the novel identification attributes. We conducted a number of interviews at the beginning of the project to understand the different types of investigations into identity by law enforcement officers and intelligence analysts.

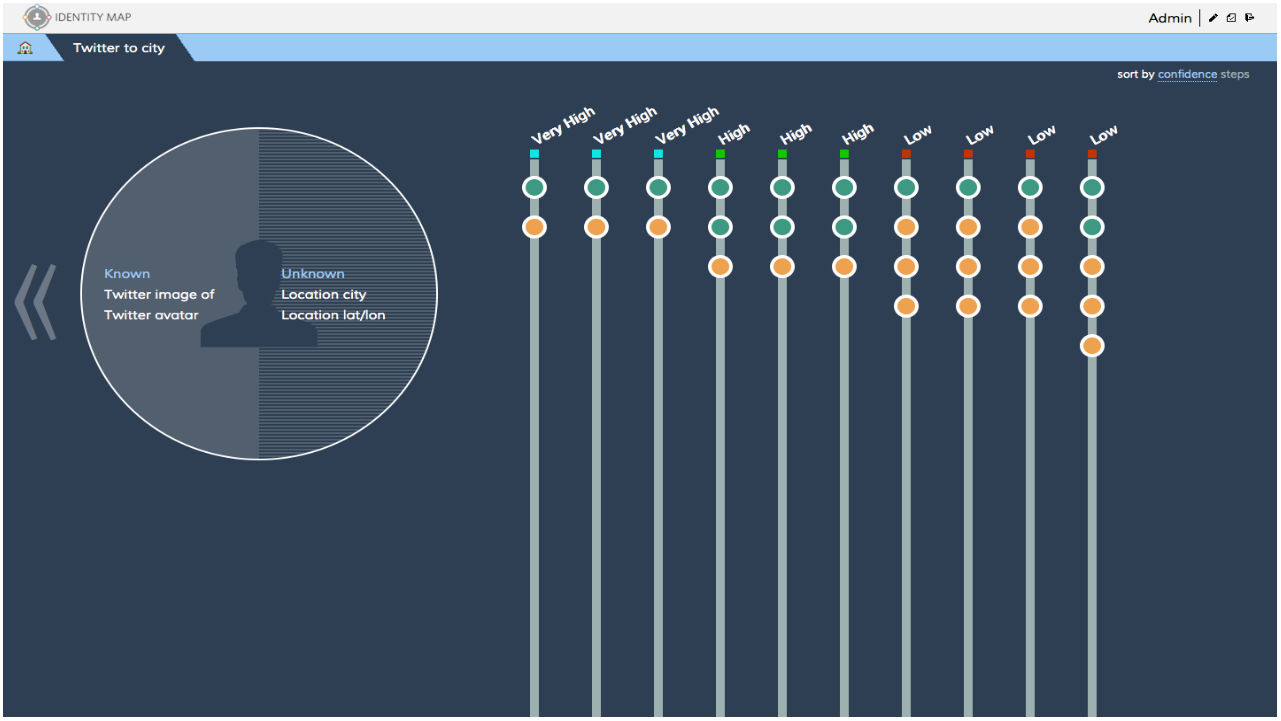

Figure 1. A visualization from Identity Map showing pathways and confidence levels from Twitter images and avatar (known information) to location city and latitude/longitude presence (unknown information).

Using the Identity Map, we can use various known and unknown attributes identified in our early interviews and find where personality traits and other new biometrics work would be helpful in the identification task [15].

A future version of the Identity Map is expected to become the user interface for the model. This visualization presents the user with various pathways that would result in the desired end-user attributes given what is already known. The user could then select the pathway based on the accessibility of information and/or the confidence of the desired results.

The first two pilots were run using a diagram of the "critical path" (e.g., Figure 2). The critical path shows all the possible pathways from the set of known attributes to desired attributes. The diagram in Figure 2 is shown to convey the complexity of the model presented and is not expected to be readable.

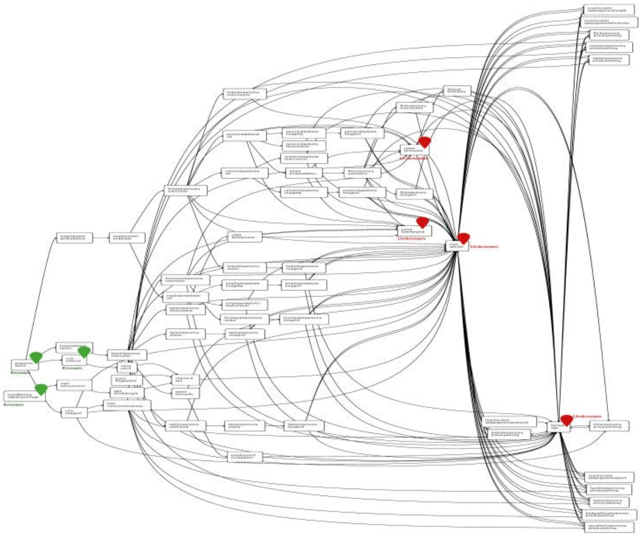

Figure 2. The critical path used in our initial pilot study.

We had two different scenarios: one for intelligence analysts and one for law enforcement officers. These scenarios were developed based on interviews we had conducted earlier with intelligence analysts and law enforcement personnel and were augmented to provide new identity attributes such as swipes on touch screens, gait analysis, blog analysis, and avatars used in chat rooms.

For the pilot study in the law enforcement area, we selected a scenario about a stolen credit card: "A stolen credit card was used at a local business. The police have the latitude/longitude coordinates of the business, an image of the signature used, and footage on a CCTV from the time when the card was used. What they are interested in is the real name of the person, the age and gender to make a positive identification." The critical path in Figure 2 is for this scenario.

We asked participants to use this scenario and then to explore the various pathways and nodes in the model and to think about which ones they would normally use and which new ones they would consider based on model suggestions. We were interested in determining if there were types of information that would not be acceptable. For example, in some environments there may be a perception that the biometric or psychological domains might not be as acceptable as the biographical domain (e.g., factual, personal information) and cyber information (e.g. user names, e-mails, blog contents). Participants were asked to color code the links and nodes to indicate whether they felt the inference was acceptable and whether the resulting information was useful. Not surprisingly, participants found it extremely difficult to look at the critical path and all the various pathways and color code them. Most importantly, it was not clear to our participants that the model was not a database; it contained no information. While the model suggests what type of information could be used to get to desired attributes via the pathways, any tangible information would have to be from sources to which the agency had access.

Based on the complexity of the two first pilot evaluations, we revised the pilot study and tested it out on a third participant. We explained the model, noting that no data was contained in it and that in actual use the data would be supplied by the agency using the model. We attempted to simplify the critical path by showing the possible pathways iteratively, slowly building up the critical path. Moreover, we used a simpler scenario, putatively closer to the participants' experience. We also included the rationale for being able to go from one attribute to another. The scenario used was that a suspicious online article gets the attention of law enforcement officials. The first place this article surfaces is from a link posted on Twitter. The Twitter username of this individual and the other users who were affiliated with the account were collected. The Twitter host was not able to share any additional information. An investigator wishes to understand who this person is (and quickly). In particular, she would like to know the person's area of expertise, age, gender, location, and ideology.

Figures 3, 4 and 5 show samples of the pathways we gave to the participant. At the very end, the participant saw an image of a section of the actual model to ensure that the scope of the model was understood. These changes are all aimed at slowly introducing the complexity of an abstract model by allowing the participant to engage and understand the model by anchoring it to a scenario relevant to each participant's experience.

We learned from our first pilot tests that the model diagrams were too complex for participants to easily comment on. Also, it was difficult for participants to understand that no data was associated with the model; the model suggested pathways but it would be up to users to put in the data they were able to access. Explanations for the models must be clearly communicated so that the participants understand exactly what the input and outputs will be. As most users are accustomed to working with complete software packages, care must be taken to explain that these models are just a component that would most likely go into a piece of finished software.

In the revised evaluation, the selected scenario did not use any of the more novel identity attributes and inferences, such as touchscreen gesture biometrics and hand-vein imagery. Therefore, it was viewed by the participant as what he currently does—i.e., the model didn't necessarily add value for the use-case. This will be rectified in the next iteration of the evaluation procedure. We will identify a scenario that stresses innovative inferences based on the project's research—specifically, identity information and information about the relationships between cyber behavior and personality traits that can be inferred from touchscreen swipe gestures. These novel attributes and inferences have been identified as potentially powerful research that is encapsulated within the model.

4. CONSIDERATIONS IN EVALUATING MODELS

We believe that evaluation of underlying models is extremely useful and should be done as early as possible. The complete steps in designing and implementing a user evaluation of a model are:

2. Identify the time in the research/development cycle to conduct the evaluation. This should be based on trade-offs for time and resources to do the evaluation versus cost of re-coding or re-designing the model.
3. Identify the users needed as participants in the evaluation and where these users can be obtained.
4. Design and implement an evaluation technique.
5. Conduct one or two pilot studies to determine if the evaluation technique will work.
6. Revise as needed and conduct another pilot study.
7. Conduct the evaluations.
8. Analyze the data.

There are a number of challenges in these steps. While identifying the objectives and metrics is a reasonable step, identifying the time to conduct the evaluation is often dictated by other constraints in the projects. As few utility evaluations of models in software have been done or at least documented, identifying an evaluation technique to use requires some creativity and potentially several pilot studies to refine. Another issue is how many users should be used in the evaluation. Three users were sufficient in the NIMD case, but if the model had not worked well, we might have needed more users to determine specific issues. We plan to use five to seven users in the evaluation for the SID project for each type of user (law enforcement and intelligence analyst) in the United States and we will duplicate this in the United Kingdom. Currently, we plan to use the same portion of the model (same critical path) for the two user types but with different scenarios.

We also discovered a number of issues that should be considered when designing and implementing the evaluation. These include:

- Make sure that the metrics selected are sufficient to determine if the model is achieving the desired objective.
- Test the presentation of the model to the user to ensure that it is accurate and that the presentation will not distract users. This is particularly important if an interactive means is being used to view portions of the model.
- If a scenario is used (as in the SID evaluation), make sure that it is realistic for the user population and uses portions of the model that are important to test.
- Make sure that data collection is feasible for the user and reasonable for later analysis. This is particularly true if the user is asked to mark up something or write something down.
- As far as analysis, make sure data from all users can be accumulated and compared. If verbal responses are collected independent coders are needed.

Table 1 contains a summary of the three model evaluations described in this paper. While it is not feasible to generalize from such sparse data, our experience should help others considering designing evaluations of underlying models early in the software design process.

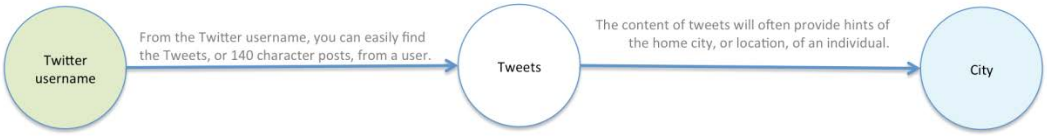
Twitter username
From the Twitter username, you can easily find the Tweets, or 140 character posts, from a user.
Tweets
The content of tweets will often provide hints of the home city, or location, of an individual.
City

Figure 3. The initial pathway showing going from Twitter username to city.

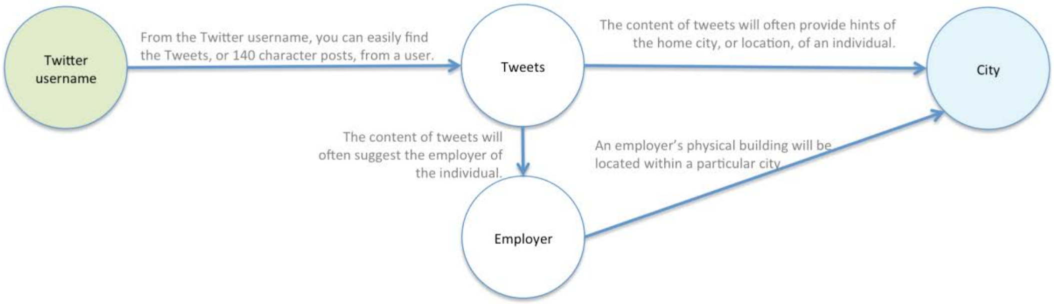
Twitter username
From the Twitter username, you can easily find the Tweets, or 140 character posts, from a user.
Tweets
The content of tweets will often provide hints of the home city, or location, of an individual.
City
The content of tweets will often suggest the employer of the individual.
An employer's physical building will be located within a particular city.
Employer

Figure 4. Another pathway showing the tweets that might contain information about the employer.

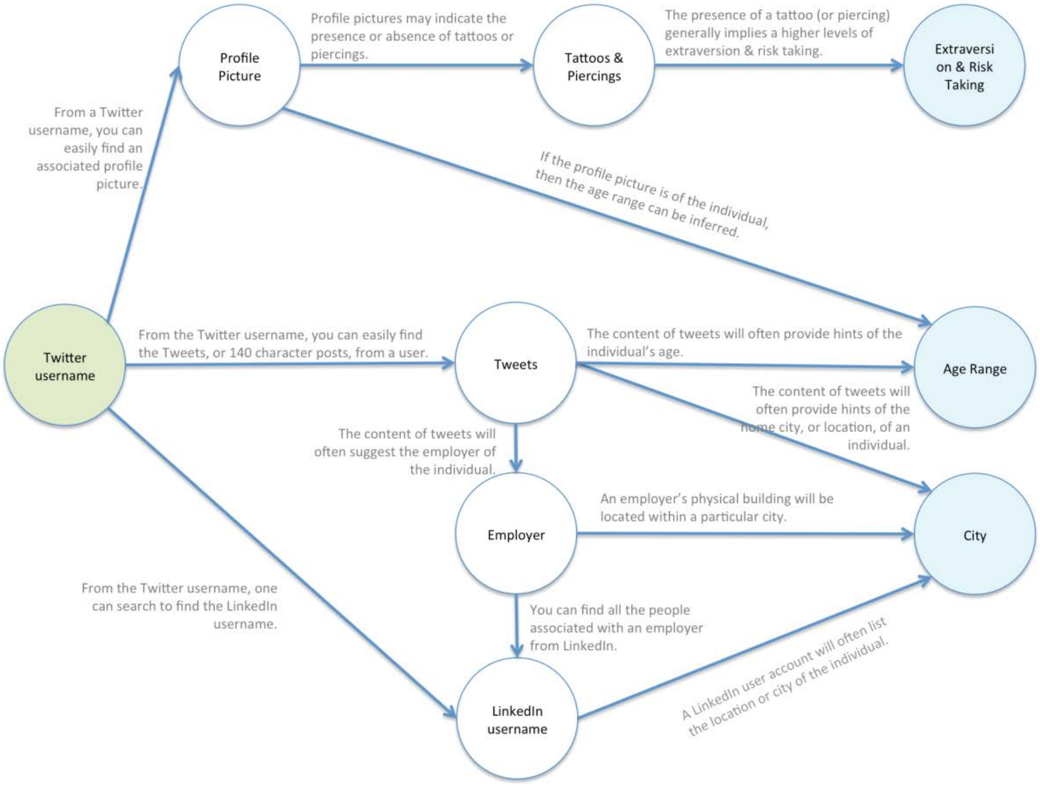
Profile Picture
Profile pictures may indicate the presence or absence of tattoos or piercings.
Tattoos & Piercings
The presence of a tattoo (or piercing) generally implies a higher levels of extraversion & risk taking.
Extraversion & Risk Taking
From a Twitter username, you can easily find an associated profile picture.
If the profile picture is of the individual, then the age range can be inferred.
Twitter username
From the Twitter username, you can easily find the Tweets, or 140 character posts, from a user.
Tweets
The content of tweets will often provide hints of the individual's age.
Age Range
The content of tweets will often provide hints of the home city, or location, of an individual.
The content of tweets will often suggest the employer of the individual.
Employer
An employer's physical building will be located within a particular city.
City
From the Twitter username, one can search to find the LinkedIn username.
You can find all the people associated with an employer from LinkedIn.
LinkedIn username
A LinkedIn user account will often list the location or city of the individual.

Figure 5. The final display of pathways including personality traits inferred from the Twitter profile picture.

Table 1. Summary of the model evaluations described in this paper

Project name	NIMD	TPAI	Super Identity
Type of model	Explanatory (user)	Predictive (Decision when models integrated into system)	Decision
Evaluation stage	Late – coding completed	Early – pre code	Early – model under construction
Users	Naval reservists substituting for intelligence analysts	Academic experts/analysts	Law enforcement officials and intelligence analysts
Evaluation results	Yes	No – never conducted	Pilot tests only
Metrics used	Number of relevant documents compared to no user model	Understanding of models and impact of changes using parameters Understanding of interaction of technical and social models	Pathways that would be used and are novel; attributes that would be used and are novel
Design notes	Used two systems side by side. User did not know which was which. Each user completed 10 queries we assigned. Counted the number of relevant documents from each system.	Paper-based “picture” of model with parameters. Users would be experts in technical models and social models; also would be run with analysts who were skilled in technical/social areas.	Scenarios given based on early user interviews as to the type of identity information that was known and what was unknown. Paper-based evaluation; user to mark pathways/attributes as to whether they were useful and novel.
Comments	Would have liked to have done the evaluation earlier but that would have been more complex to design and implement.	The actual collection of the data was not completely addressed.	Needed to present model in stages rather than all at once. Scenarios need to include novel user attributes not currently used.

5. CONCLUSIONS

We are encouraged by the evaluations we have designed and in some cases carried out. However, designing and implementing these models is difficult. We would like to encourage more work in this area. Specifically, more work is needed to find different methods to evaluate underlying models as early as possible. We are also interested in looking at various types of models—user models, system models, technical or social models, and prediction models among others—to determine if there are different treatments needed for evaluation. The development of a classification scheme for the various models and their objectives and associated metrics would be useful; for example, it would help others implementing these evaluations to use previous work as templates for evaluation or to see where new methods for evaluation are needed. In addition, such issues as the number of users needed for the different types of evaluations and the appropriate time for evaluations will only come to light as a body of literature on this work becomes available. As we conduct more evaluations in this area, we will be able to determine how useful they are and to evolve this area of evaluation as usability testing of user interfaces has evolved over the years.

6. ACKNOWLEDGMENTS

SuperIdentity work performed at the Pacific Northwest National Laboratory (PNNL) was supported by the U.S. Department of Homeland Security. PNNL is managed for the U.S. Department of Energy by Battelle under Contract DE-AC05-76RL01830. The work performed at Oxford and Bath was performed under the Engineering and Physical Sciences Research Council (EPSRC) grant EP/J004995/1.

The Technosocial Predictive Analytics work was an internal PNNL project supported by the U.S. Department of Energy under Laboratory Directed Research and Development funds.

The Novel Intelligence from Masssive Data evaluation work referenced here was performed at the National Institute of Standards and Technology by one of the authors and was funded by the Advanced Research and Development Agency.