Observing Web Users: Conjecturing and Refutation on Partial Evidence

Miguel-Ángel Sicilia
DEI Laboratory, Carlos III University
Avda. Universidad, 30 – 28911 Leganés (Madrid), Spain
msicilia@inf.uc3m.es

Abstract

Personalized hypermedia and Web systems are confronted with the challenge of inferring complex user traits like knowledge or preferences from very basic data like the 'clickstream' or ordinal–scale ratings. In consequence, the resulting user models are only approximations that must be subject to continuous revision. Nonetheless, knowledge revision procedures are rarely made explicit in existing adaptive systems and models. In this paper, we sketch a framework for user modeling structured around revision and refutation of provisional conjectures drawn from basic data. This model can be used as a reference framework for the evaluation of the adequacy of the inferences carried out by existing adaptive hypermedia systems. Additionally, a number of existing adaptive systems is reviewed according to the core concepts of this model. It is also argued that Possibility Theory can be used to generalize different forms of uncertainty that are not precisely justified in existing applications.

1. Introduction

Adaptive or personalized Web applications [4] face a big challenge in the process of building *user models* from the interaction of (registered or anonymous) users with browser–based interfaces. In spite of the fact that some regularities have been found in Web usage [13], the inference of characteristics of Web users from the *clickstream* — i.e. from the navigation history — poses significant epistemological problems. These problems are amplified by the inherent difficulty of modeling abstract psychological traits like preferences, knowledge or attitudes, that conform the focus of the majority of existing personalized systems [5].

For example, derivative or secondary assumptions in `AVANTI` [7] come in some cases from weak inductive rules like "if the user requests more than once detailed information on the history of some churches, he/she can be assumed to be interested in churches", in `SETA` [1] user preferences are updated by using bayesian networks — assuming that interactions are driven by probability —, and in some educational systems, page visits are used to make the assumption that students has learned the concepts described on them (for example, this unappropriate inference was fixed in the second version of `ELM-ART` [22]). In all these cases, models of highly abstract traits are built from very simple empirical facts, and theories or hypothesis for that form of conjectures in many cases are not explicitly justified or at least remain unreferenced. In addition, the focus of most research in the area is on monotonic inference, neglecting the importance of refutation of previous conjectures, i.e. if a basic fact somewhat contradicts the current model, how should the model be updated?. And more importantly, do adaptive Web systems engage in a continuous process of *inquiry* to test the validity of their current user model?. An example of such a system is `NewsDude` [2], but revision and refutation are not considered an essential part of the mainstream architecture of adaptive systems.

This paper describes a preliminary model for the described problem, inspired in Popper's philosophical contributions about the growth of knowledge through rational means [14]. We argue that adaptive systems somewhat embody theories of user interaction that provide the ground for inferences, and that theories should be justified and confronted to others as part of their evaluation method. The main point of our model is the explicit focus on means of refutation and the provisionality and inherent uncertainty of user models. Related work includes the model of *conceivable situations* described in [12], based on *doxastic* logics.

The rest of this paper is structured as follows. In Section 2, the principal elements of the model and their relationships are described, along with their applicability as adaptive system assessment instruments. In Section 3, a number of existing adaptive systems are described in terms of the model, making explicit their revision and refutation procedures. Finally, conclusions and some future research directions are provided in Section 5.

2. A (Rational) Abstract Model for Web Users

Our approach is based in a reference model for user model acquisition from the *clickstream* and other basic–data gathering procedures, which is intended to make explicit the assumptions taken in user modelling actions, so that both the design and evaluation of adaptive Web systems can be better informed about the appropriateness of concrete techniques. In this section, the main elements of the model are described, along with the rationale of its usefulness as an evaluation tool.

2.1. Main Elements of the Model

In our model, basic facts about the users are described as *factual propositions* in a set F that are related to an specific interaction with the hypermedia structure. This entails that a notation for them requires a hypermedia meta-model. The MAZE abstract model [18] — or any other one with similar expressive power — can be used for that purpose.

For example, if we have the sets C, N, L and U of contents, nodes, links and users respectively, visits to a node can be denoted as $visit(n, u, t, l, sessionId)$ $n \in N, u \in U$, $l \in L$ — where t is a time-stamp, l denotes the link that guided the user to the node, and $sessionId$ identifies the user's session —, and user ratings for a given content can be formulated, for example, in the form $rate(c, u, t, sessionId, value)$, where value belong to a given ordered rating scale.

Taking the set F as raw data, the system is able to make conjectures about a specific user or a group. It is also required a number of refutation procedures (possibly consisting on series of more basic ones) in a set P that are able of retracting or making dubious (some of) the conjectures in a set Cj. The revision procedures belong to a set R. Conjectures are formulated according to one or several theories or hypotheses in $\mathcal{T}$, so that each $c_i \in Cj$ is associated to one of them. Figure 1 depicts the described general setting. Note that both F, U and the depicted user modeling tasks are usually considered to be part of the User Model in terms of the overall architecture of adaptive hypermedia systems [21].

For example, in NewsDude [2], the basic facts are the proportion of a story heard by the user and also explicit ratings. The former ones can be expressed in the form $p\_listened(c, value)$, where $value \in [0, 1]$ and $c \in C$, and the"channel" or topic of each c_i is also known. The most elemental conjectures are *scores* that are deduced through rules such as "if story was rated as interesting, $score = 0.7 + 0.3 \cdot p$ ", so that a basic theory $\mathcal{T}_a$ about the correlation between interest and percentage of listening is assumed. In addition, both a short term and a long term user models are built, the former based on a nearest neighbor algorithm ($\mathcal{T}_b$)

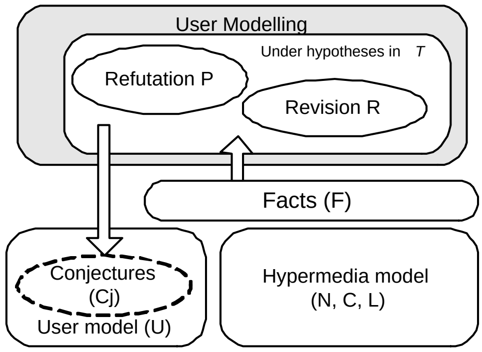

Figure 1. Conjecturing, Refutation and Revision as User Modeling

and the second on probabilities ($\mathcal{T}_c$). Three basic refutation procedures p_i are used, according to three *templates*. All of them entail explaining the user the inferences made and allowing him/her to express with a binary rating if the reasoning matches his preferences. Finally, several revision procedures are implemented, one of them entailing a way of "discarding" stories listened by the user.

2.2. Using the Model for Evaluation

Given the just described framework, each system can be evaluated in several dimensions, including the following:

(i) According to the degree of appropriateness assigned to the theories or hypothesis that guide conjecturing.

(ii) According to the degree of quality assigned to refutation procedures.

(iii) According to the degree of appropriateness assigned to the revision procedures.

Research on these three dimensions of user modeling adequacy constitutes a long–term research programme. Surprisingly, many systems use only "commonsense" approaches for (i), and problems (ii) and (iii) are in many cases overlooked, which — from the philosophical assumption that inquiry is the rational mean for acquiring knowledge —, points out that new approaches for the design of user modeling tasks are required. The just described dimensions could be used as an a *aprioristic* complement to evaluation approaches focused on the final usability of the adaptive system [8].

All the induction procedures found in adaptive hypermedia systems entail some form of uncertainty, which in most cases is handled with the use of probabilities. But when talking about user preferences or attitudes, possibility theory, as proposed by Smithson [20] provides a more realistic framework, since it provides a upper probability bound 'disconnected' with randomness. In general terms, we can state that a set of organized conjectures about a user can be modeled by a possibility measure in the form denoted in (1) and (2).

$$poss : 2^{Cj} \rightarrow [0, 1] \tag{1}$$
$$poss = g_{\mathcal{T}}(F, U) \tag{2}$$

Existing systems like `NewsDude` can be generalized this way. This implies a commitment to empiricism, since the possibility of a given state of conjectures is ultimately determined by the set of observable facts F, although they are "interpreted" in the light of a pre-existent collection $\mathcal{T}$ of hypothesis or theories.

In any case, the forms of uncertainty handled by the hypothesis and materialized in conjectures need to be carefully analyzed. Although such analysis is out of the scope of this paper, the taxonomy proposed by Smets in [19] can be used as a point of departure, taking also into account the relationships between the diverse mathematical frameworks for handling uncertainty as described in [10].

3. Modeling Some Existing Conjecturing Systems

The just described model has been used to analyze half a dozen reported hypermedia systems with regards to its described process of inquiry, using a basic ontology built with `OILEd`[1] (Figure 2 shows an screenshot containing a part of the ontology terms describing the model). The model is able to making explicit the weaknesses of adaptive systems in the long-run whenever a proper mechanism for conjecturing and refutation is not provided. The main benefit of the described model is that it can serve as a reference model for the evaluation of the *epistemological adequacy* — according to the definition in [11] — of user modeling tasks.

In this section, we describe some existing systems in terms of the model depicted in Figure 1, in an attempt to provide some evidence for the necessity of making revision and refutation explicit.

3.1. Knowledge Graph–Based Instructional Systems

The `ELM-ART` system described in [22] is an adaptive electronic teaching application for the LISP language. This an other educational systems use a hierarchical model of concepts regarding the domain space of the subject being teach, where concepts and sub-concepts form some sort of

[1]http://oiled.man.ac.uk/

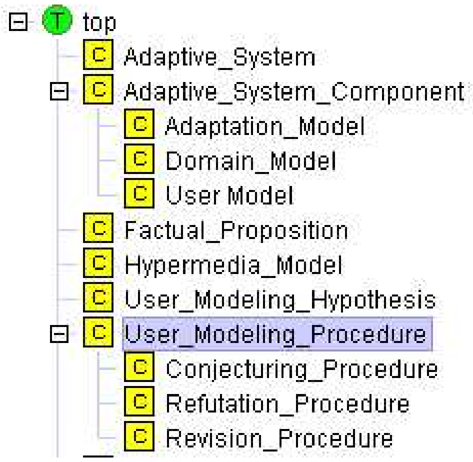

Figure 2. Partial View of the Terminology of the Model

aggregation. Factual propositions about the user include the navigation of the user and the results of exercises and tests that are associated to units.

From elements in F, the following knowledge states for a given unit or concept $c \in C$ and a user $u \in U$ are conjectured:

- $visited(u, c)$. This is really an element belonging to F, since it's a mere recording of the navigation of the user.
- $learned(u, c, confidence)$ represents the degree of confidence the system has about the knowledge of the user regarding the given concept.
- $inferred(u, c)$ represents the fact that the given concept is considered to be known to the user since he/she has worked more advanced units.
- $known(u, c)$ represent the explicit fact that the user has explicitly stated that he/she knows the given concept.

The $learned$ status is obtained through conventional assessing method like tests, so that reliability depends on the quality (i.e. classical validity and reliability) of the measurement instruments, but no difference exists with a non-computer mediated situation. Nonetheless, a more cautious approach may considered that state as provisional, since subsequent activities may reveal that the concept has been forgotten by the student, or the assessment failed (actually, this kind of revision systems are rare in Universities, since knowledge is credited forever once the student has passed formal assessments).

Inferred states are valid to the extent that the relationships between concepts authored (as part of the hypermedia model) by the instructor are. These inferences are drawn from 'inference links' that conform a domain-dependent network of hypothesis about the sequence of acquisition of concepts about the given topic, so that an acyclic graph of hypothesis in the form $\mathcal{T} = \{c_i \stackrel{precedes}{\rightarrow} c_j | c_i, c_j \in C\}$ transitively form a particular instructional theory. This way, an important part of the evaluation of these systems is concerned with the concept structure.

It should also be noted that an essential difference exists between $learned$ and $known$ items, since the latter are weaker due to the subjectiveness of self-assessment. This difference should eventually result in different revision procedures.

3.2. Probabilistic Models

Several probability-based user modeling systems have been proposed (some of them are described in [9]). Here we focus on the model described in [6] for simplicity. In that system, factual propositions about the user include the last visited nodes and the time spent on them, that can be denoted as $last - visited(u, n, t, i)$, where $u \in U$, $n \in N$, i indicates the i–est latest visited noted and t is expressed in seconds. The current node is denoted as $last - visited(u, n, t, r - 1)$. Additionally, the user interact with the system through a given view or stereotype, so that the current stereotype of the user is denoted as $current - stereo(u, k)$ where $k \in \mathcal{P}$. The system makes an 'initial guess' for the first stereotype of the user, that can be denoted by $first - stereo(u, k)$. Then, the system continuously (or in a periodical basis) changes the stereotype to generate the destination of the following link traversal.

The key conjecture elaborated in the traversal algorithm is the user's discrete probability density function (PDF) $A(k)$ that models the 'belonging probability' to each profile. This elaboration is a weighted medium of four elements:

- Initial user choices, synthesized in $first - stereo$.
- The story of interaction, represented by $current - stereo$.
- The (dynamic)relevance of the stereotypes to the recent *clickstream*, i.e., to $last - visited$.
- Structural properties of the hypermedia, independent of user navigation.

The new profile is selected randomly or referring the highest $A'(k)$ value. Obviously, this selection process requires some form of hypothesis that is not described in [6] to be reasonable. Without such rationale, the heuristic is selected in a blind way, and only exhaustive experimentation can drive the final selection of the best approach. In this case, revision takes place at each navigation event, but refutation procedures are not devised. It should be noted also that probability is here expressed in terms of *a priori* judgments about the 'probability that a user belonging to profile *k* follows each link', and updating the probability distribution proceeds by an heuristic adjustment not necessarily connected with randomness of events.

3.3. Collaborative Filtering Systems

Collaborative filtering algorithms are in essence predictive models of user preferences that are built from known preferences of other users. We'll discuss here the original `GroupLens` model [15] again for the sake of simplicity. Factual propositions for that model are simple assessments of messages in a [1,5] integer scale, that can be denoted as $rate(u, m, v), u \in U, m \in C, v \in [1..5]$. From that ratings, the system infers two kinds of information:

- The first hypothesis ($\mathcal{T}_r$) is that user preferences are globally correlated according to their ratings about messages. A correlation coefficient $r_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ is computed for each pair of users X and Y (in more recent models, similarity relations are used instead).
- The second hypothesis ($\mathcal{T}_p$), based on $\mathcal{T}_r$ is that the preference of a user can be predicted from his/her correlation with the ratings of other users. An inferred rating $p_{(X,M)}$ can be computed for each user X and message M.

The first hypothesis assumes that users that agree in their rating on a subset of items would possibly agree about other items. This heuristic is by no means more than a belief that is governed by subjectivity and that is also dependent on the casual combination of ratings available at a given time. In consequence, predictions can be interpreted in terms of epistemic possibilities based on commonsense, so that the possibility for user X to rate message M with rating $p_{X,M}$ is bounded by a distribution on the domain of ratings so that $poss_{(X,M)} = 1$ for $x = p_{(X,M)}$, and $poss_{(X,M)}$ decreases for greater and lower values, with a shape determined by the amount of evidence used to compute the prediction.

Knowledge revision is inherent to collaborative filtering system, since the computation of correlation measures is (ideally) updated every time a new $rate$ fact is asserted. In practical settings, this update is actually carried out by more efficient techniques, including limiting correlation to a local neighborhood of users [16], but the user model is capable of modeling preference changes by its sole way of functioning.

Refutation procedures are not described in the original `GroupLens` system, but were added in subsequent research. The simplest form of refutation for this kind of systems is currently used by the on-line bookstore *Amazon*[2],

[2]http://www.amazon.com

which allows the user to explicitly discard an item.

Collaborative filtering systems provide a good example of a realistic user modeling system, since they focus on a concrete task, and provide a 'best effort' approach for a complex problem regarding changing preferences. These systems seamlessly accommodate user-initiated refutations, and richer hypothesis (like considering item characteristics [17]) have been added to incrementally extend the scope of F.

As shown in the just described examples, the hypothesis of user behavior that are needed to justify conjectures are not considered as first-class citizens in the models of many adaptive hypermedia systems. In addition, revision and refutation procedures are not engineered with regard to a theoretical background, but proceeds heuristically. The model described in the previous section can be used to direct adaptive system design activities to devising sound conjecturing and revision procedures, which in turn would eventually result in more informed and realistic behaviors.

4. Conclusions and Future Work

User modeling techniques based on implicit or basic evidence require sound revision and refutation procedures to be minimally credible in terms of rationality of inferences. Surprisingly, many existing personalized hypermedia or Web systems lack an explicit model for such kind of revision (and eventually, refutation) of beliefs about their users, although some of them provide mechanisms to do so.

The architecture of existing adaptive hypermedia systems must be further elaborated to come up with a realistic setting for user models, since evaluation criteria are often expressed in terms of that architecture as described in [3]. Such an elaboration must place emphasis in the temporary nature of many of the inferred user characteristics, and consequently, on revision or refutation procedures fitted to their degree of reliability or volatility. In addition, the form and interpretation of uncertainty or imprecision adopted should be justified in connection with the underlying theories used for inferencing.

Future work in the evaluation of personalized systems should stress the importance of justifying or explaining inferences and classifications in terms of some kind of explicit hypotheses of theories of the interaction of the users with the system.

References

[1] L. Ardissono and A. Goy. Tailoring the interaction with users in web stores. *User Modeling and User-Adapted Interaction*, 10(4):251–303, 2000.

[2] D. Billsus and M. Pazzani. A personal news agent that talks, learns and explains. *Proceedings of the Third Annual Conference on Autonomous Agents*, pages 268–275, 1999.

[3] P. Brusilovsky and C. Karagiannidis and D. Sampson. Adaptive User Interfaces Models and Evaluation. *1st Pan-Hellenic Conference on Human-Computer Interaction (PC-HCI 2001)*, 2001.

[4] P. Brusilovsky and M. Maybury. From adaptive hypermedia to the adaptive web. *Communications of the ACM*, 45(5):30–33, 2002.

[5] Brusilovsky, P. Adaptive hypermedia. *User Modeling and User Adapted Interaction*, 11 (1/2), 87–110, 2001

[6] Cannataro M. and A. Cuzzocrea and A. Pugliese. A Probabilistic Approach to Model Adaptive Hypermedia Systems. *Proceedings of the 1st International Workshop on Web Dynamics (WebDyn), in conjunction with the 8th International Conference on Database Theory*, 2001.

[7] J. Fink, A. Kobsa, and A. Nill. Adaptable and adaptive information provision for all users, including disabled and elderly people. *New Review of Hypermedia and Multimedia*, (4):163–188, 1998.

[8] Höök, K. Evaluating the Utility and Usability of an Adaptive Hypermedia System. *Journal of Knowledge Based Systems* 10(5), 1998.

[9] Jameson, A. Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues *User Modeling and User-Adapted Interaction*, 5, pages 193–251, 1996.

[10] Klir, G., Wierman, M.: Uncertainty-Based Information: Elements of Generalized Information Theory. Springer-Verlag (1998).

[11] J. L. McCarthy. Epistemological problems of artificial intelligence. pages 1038–1044, 1977.

[12] A. Moreno. Inquirers: a general model of non-ideal agents. *International Journal of Intelligent Systems*, 15(3):197–215, 2000.

[13] J. E. Pitkow. Summary of WWW characterizations. *Web Journal*, 2(1–2):3–13, 1998.

[14] K. Popper. *The Logic of Scientific Discovery*. Routledge, 1977.

[15] Resnick, P. and Iacovou, N. and Sushak and M. and Bergstrom, P. and Riedl, J. GroupLens: An open architecture for collaborative filtering of netnews. Proceedings of the 1994 Computer Supported Collaborative Work Conference, pages 175–186, 1994.

[16] Sarwar, B. M. and Karypis, G. and Konstan, J. A. and Riedl, J. Analysis of Recommender Algorithms for E-Commerce. *Proceedings of the ACM E-Commerce Conference*, pages 158–167, 2000.

[17] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International World Wide Web Conference*, 2001.

[18] M. Sicilia, E. García, P. Díaz, and I. Aedo. Fuzziness in adaptive hypermedia models. *Proceedings of the North American Fuzzy Information Processing Society Conference*, pages 268–273, 2002.

[19] Smets, P.: Imperfect information: Imprecision-Uncertainty. Uncertainty Management in Information Systems: From Needs to Solutions. Kluwer Academic Publishers (1997), 225-254.

[20] M. Smithson. *Possibility theory, fuzzy logic, and psychological explanation*, volume 12, pages 1–50. North Holland, 1988.

[21] Wu, H., De Kort, E., De Bra, P. Design Issues for General-Purpose Adaptive Hypermedia Systems. Proceedings of the ACM Conference on Hypertext and Hypermedia, pages 141–150, 2001.

[22] G. Weber and P. Brusilovsky. ELM-ART: An adaptive versatile system for web-based instruction. *International Journal of Artificial Intelligence in Education*, 12(4):351–384, 2001.