Norwegian Artificial Intelligence Symposium, Gjøvik, 22 November 2010

Tourist Without a Cause

Anders Kofod-Petersen[1,2], Helge Langseth[2]
[1]SINTEF ICT
NO-7465 Trondheim, Norway
[2]Department of Computer and Information Science
Norwegian University of Science and Technology
NO-7491 Trondheim, Norway

Abstract—**Location-based recommender systems in the tourist domain are increasingly becoming more and more popular. However, these systems typically suffer from two problems: Not having sufficient information about the user results in the cold start problem; and acquiring suitable user models can be problematic, the knowledge bottleneck problem. The work presented here demonstrates how stereotype modelling can be used as a suitable tool for acquiring knowledge and building user models for a Bayesian network based recommender system.**

1. Introduction

The use of computer supported travelling in the tourist industry has been steadily increasing and has recently attracted considerable interest[1]. Tourism is in many ways the domain most closely connected with personal preferences and by definition connected to (physical) mobility. Hence, not surprisingly personalised location-based information systems are very suitable for this domain. The modern tourists do not only require general guidance and information but also information specifically tailored to their personal preferences. Local guides and guided tours cover many tourists' needs by customising tours. Yet, a location-based personalised recommender systems offers a supplement to the available customised services.

Recommender systems are designed to help users cope with vast amounts of information, and they do so by presenting only a certain subset of items that is believed to be relevant for the user. Traditionally, these systems recommend items like books (amazon.com) or films (movielens.com), but may just as well recommend *points of interest* (POIs) to tourists.

The typical tourist will not linger long in any location. Hence, a location-based information system will not be able to effectively learn the idiosyncrasies of any single tourist. This is a challenge when dealing with recommender systems, as they (most often) rely on a classification of the user and the information it is attempting to recommend. Not having sufficient information is known as the *cold start* problem.

[1]See e.g. Google's City Tours http://citytours.googlelabs.com/

The cold start problem can to some degree be alleviated by employing user models. However, building user models requires (sufficient) knowledge about the specific user. Acquiring this knowledge is subject to the *knowledge bottleneck* problem. That is, it is time consuming (for the user) and not necessarily easily accessible.

The work presented here suggests to combine Bayesian networks [10], [6] with stereotype modelling [12] as a means of mitigating both the cold start and knowledge bottleneck problem.

The rest of the paper is organised as follows: In Section 2 user modelling in the tourist domain by using stereotypes is introduced; Section 3 describes the background on recommender systems and Bayesian networks, and the implemented system; Section 4 describes how an experiment is carried out on an existing mobile learning system. Finally, Section 5 draws up the conclusion and points to future work.

2. User modelling

2.1. Stereotypes

Choosing a specific user model type for adaptive systems requires some considerations as to the nature of the users and system in question. The three most important aspects to consider are: *i*) if users are homogenous or heterogeneous; *ii*) whether users are permanent or not; and *iii*) if their interests are persistent [9].

In domains where the user group is highly homogeneous, canonical user models are likely to be the best option. Whereas in domains where the user group is highly heterogeneous, special user models are likely to

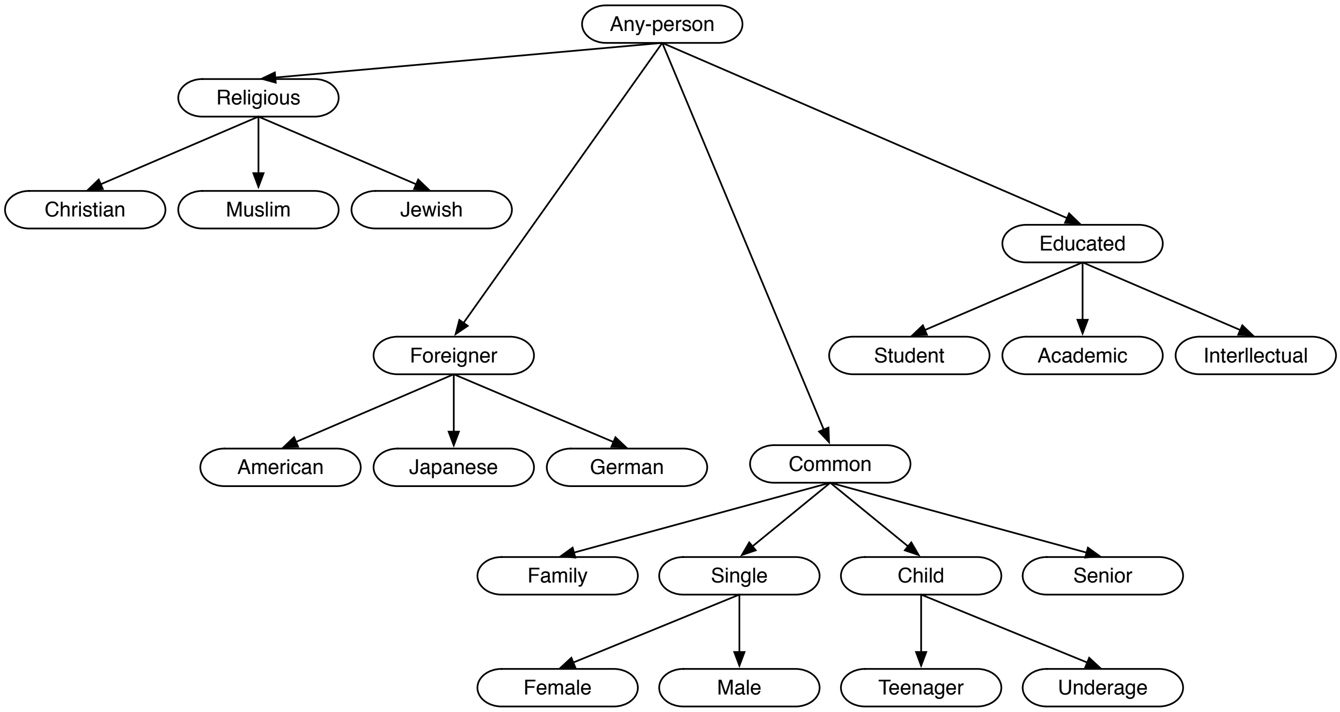

Fig. 1. Example of a stereotype hierarchy.

be the preferred option. In systems where the user is permanent, that is uses the system for a long time, the system has the option to start from scratch and learn a specific user model over time. However, for systems where the user only spends a limited time the user model is likely to be canonical.

Typically, a system with homogeneous and non-permanent users will employ very limited and standardised user models, whereas systems with heterogeneous and permanent users will employ very personalised user models. Recommender systems in location-aware systems for the tourist domain will typically contain a non-permanent and heterogeneous user group. However, we can assume that the interests of the user are persistent throughout the time that the user deals with the system. To conclude, our recommender system will attempt to approximate the user's preferences by employing a suitable user model that will alleviate the cold start problem, but at the same time the user modelling must be of such a quality that it does not cause a knowledge bottleneck problem.

Using stereotypes is a quick and efficient way of building user models [12]. A stereotype contains characteristics ascribed to the user that fits this particular stereotype. The main advantages of using stereotypes are the fact that they are easy to build and quick to use (mitigating the bottleneck problem). Stereotypes contain information on the stereotypes and their characteristics, which are know as *facets*. Traditionally, a facet represents a certain quality, such as interest in art, and is coded with a value ranging from -5 to 5. Each facet has a certainty assign, ranging from 0 to 1000. This rating tells us how certain the system is about the value assigned. Thus a high facet value tells us that the users is very interested in the particular facets, while a high certainty rating tells us that we are very certain of the rating.

Stereotypes are traditionally organised as a directed acyclic graph (DAG), where the root node, *any-person*, contains all the facets with average values, and more descriptive values for the facets are given for the more specialist stereotypes. As an example, consider Figure 1 and assume the facet we are interested in is "`Fondness of wurst`". For the *any-person* we would give a fairly non-descriptive value (say zero), whereas the facet would be given a rather large positive value for a *german*, and a vaguely negative value for a *japanese*. Knowing a person's nationality therefore helps indicating this person's "`Fondness of wurst`".[2] Note that a given person can belong to several stereotypes, e.g., a *german male*, who is also a *muslim*.

Stereotypes typically have their strength as a model representation tool; they are quite easy to cope with for those who are to provide the information in a system. Often stereotypes are defined by simplifying subjective perceptions on what aspects that describe certain groups of people. Employing stereotype modelling allows for

[2]One may argue that it is just as easy to query the user about his facet values as it is to find his stereotypes, but as we are interested in many facets (several hundred facets would be required for the tourist domain), this is not a reasonable path to follow.

more personalised user models also when dealing with homogeneous and non-permanent users.

The main challenge for stereotypes is that they make rather rough statements, such as: "you are a typical *German*". However, as also noted by Rich: "*...[An intelligent system] must not regard the user model as fixed, but rather as something upon which it can continuously improve by collecting feedback from the user on each interaction.*" [12]. Thus a *sound* knowledge maintenance technique is required. The approach described here will initially use the stereotypes to make an initial user classification, that must later be refined as the system learns more about the user.

Using a sound knowledge maintenance technique, allows us to update our belief in facets being more-or-less present (value of the facets changing) as new information is collected. Rich's description of facets use `<value, rating>`-pairs [12], and to the best of our knowledge, a sound knowledge maintenance technique of this representation has not yet been found. We therefore propose to rather use a random variable to represent the value of a facet, and believe that this variable's distribution (e.g., represented as `<expectation, variance>` when possible) can be used to cover the same information. Furthermore, this formulation with random variables enables the use of a sound inference scheme (standard manipulation of multivariate statistical distributions).

To summarise our description of user models so far: our approach will use the stereotypes to make an initial user "classification". This classification will amount to a set of stereotypes the user is ascribed to. Next, the stereotypes give values to all user *facets*. The facets are represented as random variables, hence the description of the relevant stereotypes will then be translated into a simple Bayesian network, as described next.

2.2. *The facet model*

From the description of the stereotype hierarchy it should be clear that all users will always be at least a member of the *any-person* stereotype, and all the facets will therefore be defined for that user, as they are all defined using "averaged" values for that stereotype. These gross average descriptions are typically modelled by letting the variances of the corresponding distributions be large. A user, who is also a member of a more specific stereotype, like *german*, will get the distribution of some facets (like "`Fondness of wurst`") changed to a larger value and with a smaller variance. Defining the facet distributions for each stereotype in the hierarchy is part of the modelling that can be performed off-line, and therefore not something the user will be exposed to.

Following this approach, the distribution of each facet is well-defined *as long as* the user is only member of stereotypes in one branch of the stereotype hierarchy

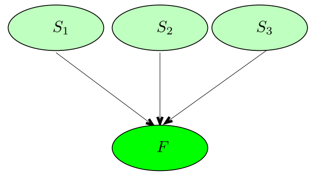

Fig. 2. A Bayesian network fragment to infer a user's value of facet F from the stereotypes S_j, $j = 1, 2, 3$.

(e.g., if the user is only member of *any-person*, *foreigner*, and *german*). Problems arise when, for instance, our *german* friend is also a *muslim*. *Germans* are fond of wurst, whereas muslims do not eat them. How should this be incorporated in our system? We propose to use a small Bayesian network as depicted in Figure 2 to handle the different sources of information about facets. In this model, F represent the facet value, and S_j, $j = 1, 2, 3$, are the stereotypes in question (in our example, F represents "`Fondness of wurst`", S_1 the stereotype *german*, S_2 represents *male*, and S_3 *muslim*. Note that the stereotypes the person is *not* a part of are not included in the model. Also, only the most specific stereotypes are included (meaning that, e.g., *foreigner* is overridden by *german* in this example). During model building, the domain expert will have to parameterise the distributions of the different Bayesian networks of the type shown in Fig. 2. In principle, this amounts to defining: i) the marginal (unconditional) probability of the user being member of any stereotype, and ii) the conditional distribution for the facet F given all configurations of the conditioning variables S_j. This would lead to a daunting knowledge acquisition process, as we may potentially have many stereotype memberships, and the complexity of the conditional distribution of F grows exponentially in the number of stereotypes.

To simplify, we first make the observation that we will include the stereotypes that are relevant for the user, meaning that we know S_j is true if S_j is in the model. We will therefore not have to define the distribution over S_j, and also not have to define the probability distribution for F given S_j when S_j is false. To simplify further, we will first assume that F is continuous, and more specifically, a Gaussian variable. Then, we assume a simple instantiation of the *independence of causal influence*-model [4]. In practice, we assume that each stereotype is annotated with a *mean-shift* and a *variance-scale*, where the former relates to the facet value, and the latter is connected to the confidence rating. The mean of F given a set of parents is then found by using the unconditional mean (i.e., taken from the *any-person* definition) and add the different mean-shifts. The conditional variance is found by taking the unconditional variance, and multiply with all relevant variance scales.

Finally, the facet model is extended to a "full-fletched" recommender system, as described in the next section.

3. Recommender systems

3.1. Background

The main idea of our system is to alleviate the knowledge acquisition bottleneck by also adapting the user models based on implicit measurements of the users' preferences. These measurements are gathered by monitoring whether or not the user follows the recommendations given by the system.

In the current version of the model we focus on the user's *geo-position*, in the future we will also include explicit user feedback in terms of ratings or free-text tagging of a POI. One way to fusion the explicit information from user interviews with implicit behaviour information is to build a *recommender system*, and this will be the topic of the rest of the paper.

Recommender systems are usually grouped into two categories: Content-based systems and collaborative filtering. Content-based systems make recommendations based on a user preference model that combines the user's ratings with, e.g., content information and structured descriptions of POIs. Contrary, collaborative filtering uses the "ratings" of like-minded users to make recommendations for the user in question. Ratings can be explicit (e.g., a number of stars given to a POI), or implicit (e.g., the amount of time spend at the POI that was recommended). In the current set-up we use implicit ratings: users who spend more time at a POI are assumed to rate the POI higher than a user who spends less time at that particular location.

Over the last decade recommender systems based on collaborative filtering have enjoyed a great deal of interest. Collaborative filtering systems are often characterised as either being model-based or memory-based [2], although hybrid systems have also been developed [11]. Roughly speaking, memory-based algorithms use the whole database of user ratings and rely on a distance function to measure user similarity. On the other hand, model-based algorithms learn a model for user preferences, which is subsequently used to predict a user's rating for a particular item that he or she has not seen before.

3.2. The full model

We propose to use a model-based collaborative filtering system in the present case. The system described here employs a Bayesian networks based collaborative filtering model called "*the user centric model*" in [8], see also [5], [7].

This specific model is specifically geared towards avoiding the cold start problem by taking advantage of implicit user models in terms of the list of facets described in Section 2. In this model, a user is represented by a random variable in q-dimensions, where each dimension corresponds to one of the facets in the user model; let $\boldsymbol{F} \in \mathbb{R}^q$ be the representation of a user. For mathematical convenience, we assume that the facets follow the Gaussian distribution, with expectation and variance according to the stereotype(s) the user was allocated to. Now, it is assumed that R_t, the "rating" the user will give to POI t, is Gaussian distributed according to the relationship

$$R_t|\{\boldsymbol{F}=\boldsymbol{f}\} \sim \mathcal{N}(\mu_t + \boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{f}, \sigma_t^2)$$

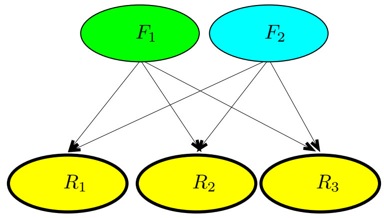

Fig. 3. The CF model. A user giving ratings R_1 and R_2, say, will give an updated posterior distribution over (F_1, F_2), and hence improve predictions (i.e., calculate the posterior over R_3).

The model is also shown in Fig. 3. The model represents POI t using three parameters: μ_t, $\boldsymbol{w}_t$, and σ_t. μ_t is the expected time spent at POI t for a "standard" user, and is hence estimated as the gross average of time all users visiting POI t uses there. The vector $\boldsymbol{w}_t$ describes the POI in terms of the facets in the user model, where a positive number at position i in the parameter vector means that POI t offers an experience that will be well received by users with a high value in facet i. Finally, σ_t is a notion of the unexplained variation in the measurements.

Assume that the user has rated items POI$_1$, POI$_2$, ..., POI$_k$ giving us observations $\mathcal{R} = \{r_1, \ldots, r_k\}$. Since the user model $\boldsymbol{F}$ is Gaussian a priori, the posterior distribution $\boldsymbol{F}|\mathcal{R}$ remains Gaussian with an updated mean value and a reduced variance. Thus, the prediction for a new POI t^* $\mathbb{E}[R_{t^*}|\mathcal{R}]$ is easily calculable, and typically different from the unconditional prediction $\mathbb{E}[R_{t^*}]$.

The last piece of the puzzle is glueing together the two model fragments we have discussed so far to obtain the overall model description that we use. This is simply obtain by letting all *facet* nodes get their relevant parents, as described in Section 2.2. The full model, assuming $q = 2$ facets and $T = 3$ POIs, is depicted in Fig. 4.

Employing this full model gives us the opportunity to: *i*) represent all we know about the user by explicitly code it in the prior distributions for the facets; *ii*) calculate the expected time the user will spend in each POI; *iii*) *rank* the different POIs based on the expected time the present user would spend divided by the gross average expected time of all users; and *iv*) update our belief about the facets (and therefore the predictions – and thereby ranking of POIs) as new information comes in.

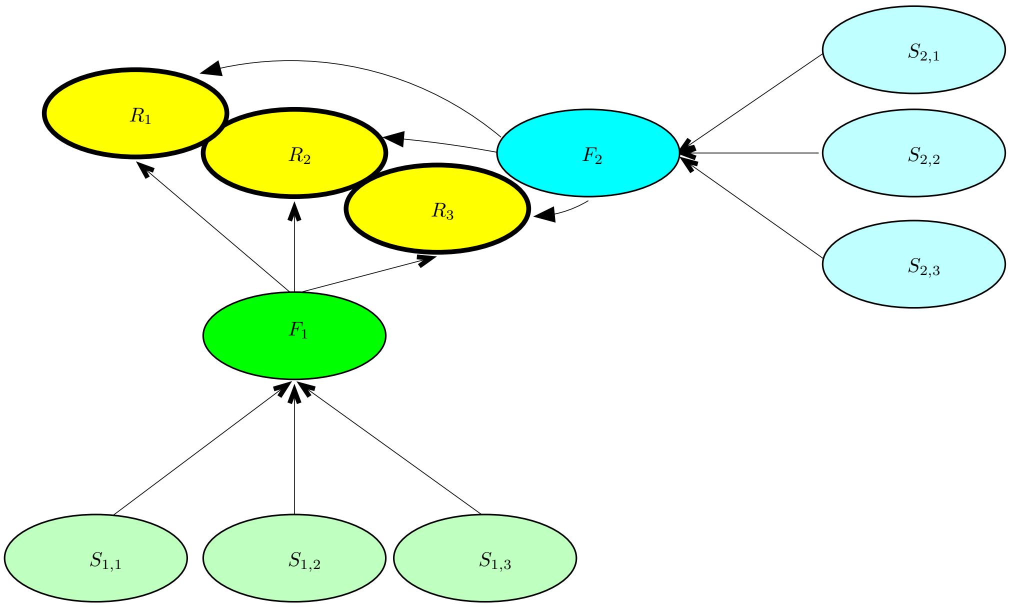

Fig. 4. The full Bayesian network model

3.3. *Defining the model parameters*

As already described, the model contains the following parameters:

- **Facet model:** The *mean-shift* and *variance-scale* for each facet and each stereotype that redefines that facet.
- **Rating model:** μ_t, the *mean* time spent at a POI t; $w_{t,i}$, the strength of influence facet i has on the rating of POI t, and σ_t^2, the conditional variance.

Initially, we define all these parameters off-line (before the system is employed). However, as the system has gathered experience, we can evolve the parameters using standard maximum likelihood learning. As some of the variables are hidden (we will never observe the "true" value of the *facet* variables), we must use the EM algorithm [3], but as the model is fully Gaussian, the EM algorithm can be easily implemented, and is available in the state-of-the-art Bayesian network systems that are available.[3]

4. Experiment

The initial testing of our proposed approach is based on data generated for tourist application in Trondheim, Norway. For the example described we start of from the model of the stereotypes depicted in Figure 1.

A new user starts using our system, and is recognised as being a member of the prototypes *male* (see Table I for a description of the male stereotype using Rich's terminology of values and ratings) and *senior* (Table II).

[3]One example is the excellent Hugin system, http://bayesian.net/.

TABLE I
MALE STEREOTYPE

Facet	**Value**		**Rating**
Interests			
	Sports	4	800
	Nature	2	700
Motivation			
	Excite	4	600
Personality			
	Tolerance-sex	5	1000
	Tolerance-violence	4	700
	Sympathy	-2	500
	Kindness	-2	400

TABLE II
SENIOR STEREOTYPE

Facet	**Value**		**Rating**
Interests			
	Sports	4	700
	Nature	4	400
	History	4	600
Motivation			
	Relaxing	4	700
Personality			
	Tolerance-sex	-4	600
	Tolerance-violence	-3	500

For completeness, the *any-person* is described in Table III. Translating the two stereotypes into random variables, and combining them will give us the following values for the most important facets in the user profile (see Section 2.2):

- `Sports`: 4 (Variance: 1)
- `Nature`: 3 (Variance: 2)
- `History`: 3 (Variance: 2)

Initially, the user in question appears to be into sports.

TABLE III
ANY-PERSON STEREOTYPE

Facet	Value		Rating
Interests			
	Sports	2	300
	Nature	1	200
	History	1	200
	Architecture	0	400
	Clubbing	1	100
	Arts	-1	100
	Food	2	500
Motivation			
	Learning	-2	200
	Relaxing	2	300
	Excite	0	100
	Amusement	1	300
Personality			
	Tolerance-sex	-2	500
	Tolerance-violence	-4	400
	Intelligence	0	300
	Perseverance	0	300
	Sympathy	0	300
	Kindness	0	200x

To deduce if the user is really a sport fanatic the system can ask some relevant questions, which can be used to reduce the variance of the `Sports` facet. Let us assume that the user's answers give further evidence in the direction of him being a sport fanatic. Thus, the sports facet is now:

- Sports: 4.1 (Variance: 0.5)

We can use this to calculate the expected time to us in each location (given user model). The results are that **Lerkendal**, the local football arena, is the highest rated POI, **The horse track** is second, and **Ringve** the music museum is the third.

For the current example, let us assume that we expected the user to spend 43.5 minutes at **Lerkendal**, but observe that only 20 minutes are spent there. This is still more than average (15 minutes), so it is still regarded as a positive experience for the user, but not to the extent we previously anticipated. The result in the model is that we explain this mismatch by updating the *sports* facet (new value: 2.5). Furthermore, since **Lerkendal** is related to the `history` facet, the model also explain the observation by reducing the belief in the user being interested in historic sites. Also, Lerkendal is negatively correlated with `intelligence`, hence the model estimate of intelligence (previously 0 – inherited from *any-person*, see Table III), and now increased to 3.55. Note that the big shift in this value, which is because we had no information or belief about intelligence up to now, and therefore would have a rather large a priori variance connected to this facet. `Nature` has no relation to **Lerkendal**, hence our belief about that facet remains the same.

As the `intelligence` facet takes centre stage, the next recommendation is based mainly on that particular facet. The updated recommendation list has **Industrimuseet** (museum of industry) as the top POI, with **Ringve Museum** second, and the **Horse tracks** only third. This shows how the recommendation system is able to take new information into account and adapt to it.

TABLE IV
EXAMPLE RUN

		Before 1st POI		After 1st POI	
Loc.	Avg. Time	User Time	Score	User Time	Score
1	15	41	2.73	20.00	1.33
2	30	22	0.73	39.05	1.30
3	15	13	0.87	12.11	0.81
4	45	57	1.27	52.56	1.17
5	30	33	1.10	37.44	1.25

Table IV describes how the model evolves. The locations are: 1) **Lerkendal football arena**, 2) **The Industry museum**, 3) **Nidaros dome**, 4) **Horse tracks**, and 5) **Ringve musical museum**. The gross average time spent at each location is given in the second column (so, people spend on average 15 minutes when visiting **Lerkendal**); the third column gives the expected time calculated from the user model (that is, before the first POI is visited), and thereafter each POIs score is given in the next column. Finally, the two last columns give Expected time for the user at each POI and the related score, calculated *after* the visit to **Lerkendal**.

5. Conclusion and future work

The work presented here has argued that stereotype modelling is very useful as a knowledge acquisition tool. It allows the information supplier to think in terms of typical traits of users and points of interests, and not worry about the underlying Bayesian representation. We have further described a Bayesian net representation that offers sound inference over the facets in the stereotypes. This approach move the stereotypes from a somewhat static user model into a model that is continuously improved. Finally, we have discussed how this Bayesian network model can be extended to capture implicit information. All of this has been demonstrated with a simple example.

Future work revolves around the integration of this model into an existing infrastructure in Wireless Trondheim [1]. The stereotypes are to be grounded in the knowledge contained at the tourist office in Trondheim. We expect to have a working prototype ready during the fall of 2010.

References

[1] S. Andresen, J. Krogstie, and T. Jelle. Lab and research activities in wireless trondheim. In *Proceedings of IEEE International Symposium on Wireless Communication Systems*, pages 385–389, Trondheim, Norway, 2007.

[2] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predicitive algorithms for collaborate filtering. Technical Report MSR-TR-98-12, Microsoft Research, 1998.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

[4] D. Heckerman and J. S. Breese. A new look at causal independence. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 286–292, San Francisco, CA., 1994. Morgan Kaufmann Publishers.

[5] T. Hofmann. Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, 22(1):89–115, 2004.

[6] F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer-Verlag, Berlin, Germany, 2007.

[7] M. Kendall. *Multivariate Analysis*. Charles Griffin & Co., London, UK, 2nd edition, 1980.

[8] H. Langseth and T. D. Nielsen. A latent model for collaborative filtering. Technical Report 09-003, Department of Computer Science, Aalborg University, 2009.

[9] H. Lieberman. Letizia: An agent that assists web browsing. In C. S. Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 924–929, Montreal, Quebec, Canada, 1995. Morgan Kaufmann Publishers.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, CA., 1988.

[11] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles. Collaborative filtering by personality diagnosis: A hybrid memory- and model-based approach. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 473–480. Kaufmann, 2000.

[12] E. Rich. Users are individuals: Individualizing user models. *International Journal of Man-Machine Stuides*, 18:199–214, 1983.