A Bayesian Approach to Conversational Recommendation Systems

F. Mangili[*], **D. Broggini**[*], **A. Antonucci**[*], **M. Alberti**[†] and **L. Cimasoni**[†]

[*]IDSIA (Switzerland) - `idsia.ch`
[†]Stagend (Switzerland) - `stagend.com`

Abstract

We present a conversational recommendation system based on a Bayesian approach. A probability mass function over the items is updated after any interaction with the user, with information-theoretic criteria optimally shaping the interaction and deciding when the conversation should be terminated and the most probable item consequently recommended. Dedicated elicitation techniques for the prior probabilities of the parameters modeling the interactions are derived from basic structural judgements. Such prior information can be combined with historical data to discriminate items with different recommendation histories. A case study based on the application of this approach to *stagend.com*, an online platform for booking entertainers, is finally discussed together with an empirical analysis showing the advantages in terms of recommendation quality and efficiency.

Introduction

The task of selecting from a collection of items one which is in some sense optimal for a specific user is a classical AI task. Several algorithms have been explored to perform such tasks and automated recommendations are today included in most modern e-commerce websites (He, Parra, and Verbert 2016; Sarwar et al. 2000). In standard setups, no interaction with users is considered, and the recommendation system bases its decision on some background information about the user, historical records of her choices and of the choices of other similar users, and, more recently, automatically inferred contextual information (Adomavicius and Tuzhilin 2011; Lu et al. 2015).

Yet, modern technologies such as chat-bots or personal assistants need systems able to support and model dynamic and sequential interactions, this leading to a substantial redesign of the traditional approaches. Here we focus on such a newer class of recommendation systems, called here *conversational*, as we term conversation a sequence of dynamically customized interactions between the user and the system, before the latter return the recommendation. This class of systems exploits the knowledge-based recommendation techniques and is based on a strong interaction with the users. Therefore, this type of recommendation system should be considered when the goal is to support the user in

purchasing a high-involvement product. In such a situation, indeed, the user wants to be involved in the decision process, and thus is not bothered by the need of interacting with the system (Jugovac and Jannach 2017).

To achieve that, we take inspiration from existing approaches in the field of computer testing (Wainer et al. 2000; Butz, Hua, and Maguire 2006; Mangili, Bonesana, and Antonucci 2017), whose goal is to determine the skills of a student on the basis of the answers to the questions of a test. In particular we focus on *adaptive* approaches in which the system selects the next question to ask to the student from a given set of questions on the basis of the previous answers by information-theoretic scores, that are also used to decide when the test should be ended. This can be easily achieved with generative probabilistic models such as, for instance, Bayesian networks (Millán et al. 2000), which are sequentially updated any time the answer to a new question is collected. The adaptive concept is converted here in a conversational approach which will lends itself to the (future) development of a dynamic generation of questions and richer interaction models.

The conversational procedure is applied to the recommendation system of Stagend (stagend.com), an online platform that allows organizers of events (playing here the role of the users) to book artists (playing here the role of the items) for their events. The goal of the recommendation system is therefore to select the most suitable artist for a particular event based on a set of needs elicited by the user. Results of some simulations based on data from Stagend platform will be presented to show the advantages provided by the conversational approach compared to earlier procedure based on a static questionnaire and discuss possible improvements.

Basic Concepts

Consider a recommendation system based on a *catalogue* $\mathcal{I}$ including n items, say $\mathcal{I} := \{i_1, \ldots, i_n\}$. The system is supposed to support the user in selecting a single item from the catalogue on the basis of a conversational process. We assume that, at the end of the conversation, the user always picks one item from $\mathcal{I}$. The uncertain quantity I denotes the element of $\mathcal{I}$ to be picked by the user. We consider a Bayesian setup and model the subjective probabilities of picking the different items before the start of the conversational process as a prior mass function $P(I)$.

We call *questions*, the interactions between the user and the system. A generic question is denoted as an uncertain quantity Q taking its possible values, to be called *answers*, from $\mathcal{Q}$. Here we assume a finite set of possible answers for each question, say $\mathcal{Q} := \{q^i, \ldots, q^r\}$. As a model of the relation between Q and I, we might be able to assess a *conditional probability table* $P(Q|I)$ whose columns are indexed by the answers and whose rows are associated to the items. After an answer $Q = q \in \mathcal{Q}$ is collected, the probability mass function over the items can be updated by Bayes theorem, i.e.,

$$P(i|q) = \frac{P(i) \cdot P(q|i)}{\sum_i P(i) \cdot P(q|i)}. \tag{1}$$

This shows that the impact of an answer $Q = q$ to the choice I is only based on the relative proportions of the values of $P(q|i)$ for each $i \in \mathcal{I}$. In particular, setting an element of the conditional probability table equal to zero implies a logical constraint for which the answer associated with column $Q = q$ makes the choice of the item associated with row $I = i$ impossible.

On the other side, decisions of the recommendation system are based on the the conditional probability $P(i|q)$ which may be strongly influenced by the prior $P(i)$. Therefore, to better elicit the behaviour of the model, understand its underpinning assumptions and foresee the system behaviors, it will be often more natural to consider the joint probability $P(i, q)$, rather than the conditional one $P(q|i)$. In fact, as the probabilities of the joint events (i, q) completely describe the model, they allow on one hand to derive the probabilities of interest $P(q|i)$ and $P(i)$ and, on the other hand, to reason about the relative weight of the posterior probabilities assigned to different items after conditioning on the answer, as, for strictly positive probabilities, it holds:

$$\frac{P(i', q)}{P(i'', q)} = \frac{P(i'|q)}{P(i''|q)}, \tag{2}$$

for each $i', i'' \in \mathcal{I}$ and $q \in \mathcal{Q}$.

We start by considering a static approach to the elicitation of the user needs, based on a list of m questions $\mathbf{Q} = \{Q_1, \ldots, Q_m\}$, called here a *questionnaire*. Selection of the optimal item to be suggested to a user at the end of the conversation is based on the conditional probability $P(i|\mathbf{q})$ of each item i given the list of answers collected, hereafter described by the vector $\mathbf{q}$. In order to improve the user experience it is desirable to minimize the number of questions necessary to identify the optimal suggestion. To this goal, the list of questions in $\mathbf{Q}$ should be built dynamically as the conversation proceeds. Such a customized list of questions is called hereafter a *conversation*. Before discussing this dynamic approach, we will present the set up that allows the computation of $P(i|\mathbf{q})$, be $\mathbf{q}$ the output of a questionnaire or of a conversation. To model a conversation we initially formulate a naive-like assumption stating the conditional independence of the questions given the item, i.e.,

$$P(q_1, \ldots, q_m|i) = \prod_{j=1}^{m} P(q_j|i). \tag{3}$$

More general models will be discussed later on in this paper. Under this assumption, if the conditional probability tables $P(Q|I)$ associated to each $Q \in \mathbf{Q}$ are available, the probability $P(i|\mathbf{q})$ can be obtained by recursively applying the Bayes theorem to update the probability $P(i|q_1, \ldots, q_k)$ after the first k answers by conditioning also on the next answer q_{k+1}.

To open the discussion, let us present a toy example, inspired by the Stagend case study reported in the final part. The example will be used through the paper to illustrate the main features of the proposed method.

Example 1. *Stagend users are invited to answer a number of questions about the entertainment they are looking for. The two questions Q_1 and Q_2 typically asked to the users are reported in Table 2, while a catalogue of three entertainers-items is in Table 1.*

$\mathcal{I}$	Catalogue
i_1	DJ available for all type of events
i_2	Band available for weddings and corporate events
i_3	Magician available for birthdays and parties for kids

Table 1: A catalogue

Q_1	Which entertainment are you looking for?
q_1^1	DJ
q_1^2	Band
q_1^3	Musician
q_1^4	Entertainer

Q_2	Which event are you organizing?
q_2^1	Wedding
q_2^2	Corporate event
q_2^3	Birthday
q_2^4	Party for kids

Table 2: Two questions of Stagend questionnaire

Elicitation by Logical Compatibility

Consider the background information about the entertainers in Table 1 and the questions (and possible answers) in Table 2. An underlying notion of *logical compatibility* can be therefore formalized as follows. Given an item $i \in \mathcal{I}$ and an answer q to a question $Q \in \mathbf{Q}$, we say that i is compatible with q, and we denote this as $i \models q$, if i satisfy all the logical requirements implied by q. As an example $i_2 \models q_2^2$, i.e., the second antertainer in the catalog is compatible with a corporate event, but $i_3 \not\models q_i^3$, i.e., the third entertainer is not a musician. Given $Q \in \mathbf{Q}$, we can therefore define an indicator function for such a compatibility concept for each pair $q \in \mathcal{Q}$ and $i \in \mathcal{I}$ as follows:

$$\delta(i, q) := \begin{cases} 1 & \text{if } i \models q \\ 0 & \text{otherwise} . \end{cases} \tag{4}$$

Following the above definition of compatibility, each item can be characterized by the set of compatible answers for each of the m possible questions. For instance, artist i_2 in Example 1 is described by set of answers $\{q_1^2\}$ to Q_1 and $\{q_2^1, q_2^2\}$ to Q_2.

In the probabilistic model proposed, the notion of compatibility is translated into the assumption that the joint probability $P(q, i)$ and, following Equation (1), also the conditional probability $P(q|i)$, have the same support of $\delta(i, q)$, i.e., they are zero whenever $i \not\models q$. To elicit $P(q, i)$, and hence $P(q|i)$, when q and i are compatible, we need additional sets of assumption. In the remaining part of this section we discuss strategies to tackle this problem and analyze the underlying assumptions and consequences in terms of posterior inferences.

Items Compatible with a Single Answer

The above discussed notion of logical compatibility might be sufficient to elicit a conditional probability table $P(Q|I)$ without any additional assumption in the special case in which each item is compatible only with a single answer. Let $q^*(i)$ denote such answer in the case of item i, i.e., $i \models q^*(i)$ and $i \not\models q$ for each $q \neq q^*(i)$. Under these assumptions, the number of non-zero elements of the joint mass function $P(Q, I)$ is equal to the catalogue cardinality $k := |\mathcal{I}|$. We assume all these non-zero probabilities to be equal, this meaning $P(q^*(i), i) = k^{-1}$ for each $i \in \mathcal{I}$, while the corresponding conditional probabilities $P(q^*(i)|i)$ have value one (e.g., see Table 3). After the conditioning on the observed answer q, all the items compatible with q will have the same posterior probability, whereas all non compatible items will receive zero probability mass. Therefore, in this simple setting, the Bayesian approach proposed implement a logical filter, making impossible the items incompatible with the answers. Such a situation occurs in the following example.

Example 2. *In the setup of Example 1, consider the conditional probability table for Q_1. Each entertainer is consistent with only one a single answer to this question. The corresponding quantification of the conditional probability table is depicted in Table 3. If we assume, for instance, that a new user answer 1 to the first question Q_1, we can update the probabilities assigned to each entertainer by Equation (1) and conclude a posterior probability equal to one for artists i_1 and zero for the other two entertainers.*

$P(q_1|i)$	q_1^1	q_1^2	q_1^3	q_1^4
i_1	1	0	0	0
i_2	0	1	0	0
i_3	0	0	0	1

Table 3: Elicitation of $P(Q_1|I)$

Items Compatible with Multiple Answers

Consider a question $Q \in \mathbf{Q}$ such that the assumptions in the previous section are not satisfied and there are items compatible with more than a single answer. In this case, the elic-

itation of the joint mass function $P(Q, I)$ requires further assumptions. Here below we describe two alternative strategies for such elicitation. It is a trivial exercise to note that both strategies lead to the elicitation discussed in the previous case for items compatible with a single answer, that is a special case of the one considered here.

Uniform Joint Strategy (UJS). Let us assume, as in the previous section, that all the compatible item/question combinations, i.e., all the pairs (i, q), with $i \in \mathcal{I}$ and $q \in \mathcal{Q}$, such that $i \models q$ have the same probability. If $N(Q)$ denotes the number of these pairs, this assumption corresponds to assigning $P(i, q) := N(Q)^{-1}$. Let us also define the *versatility* $v_Q(i)$ of an item $i \in \mathcal{I}$ with respect to a question Q as the number of answers of Q that are compatible with the item i. Then, $N(Q) = \sum_{i \in \mathcal{I}} v_Q(i)$. From the joint mass function $P(I, Q)$, we obtain $P(i) = v_Q(i)/N(Q)$ and $P(q|i) = \delta(i, q)/v_Q(i)$ for each $i \in \mathcal{I}$.

It is easy to note that, before the question is asked, the more versatile an item is, the higher is its probability of being selected. However, once the desired characteristic of the item has been elicited by the user, all item compatible with it will be assigned the same posterior probability.

Example 3. *In the setup of Example 1, UJS can be applied to the elicitation of the conditional probability table for question Q_2. This corresponds to the joint probability mass function in Table 4. By summing over the rows we obtain the corresponding marginal probabilities for the entertainers: $P(i_1) = 1/2$, and $P(i_2) = P(i_3) = 1/4$, while the corresponding conditional probability table is the one depicted in Table 5. Note that, before the question, entertainer i_1 has twice the probability of the other to be chosen, as it is the most versatile one. However, once we collect the answer, e.g., $Q_2 = q_2^1$ the two compatible entertainers, i.e, i_1 and i_2, receive the same conditional probability:*

$$P(i_1|q_2^1) \propto P(q_2^1|i_1) \cdot P(i_1) = 1/8, \quad (5)$$
$$P(i_2|q_2^1) \propto P(q_2^1|i_2) \cdot P(i_2) = 1/8. \quad (6)$$

$P(i, q_2)$	q_2^1	q_2^2	q_2^3	q_2^4
i_1	$1/8$	$1/8$	$1/8$	$1/8$
i_2	$1/8$	$1/8$	0	0
i_3	0	0	$1/8$	$1/8$

Table 4: UJS elicitation of $P(Q_2, I)$

$P(q_2|i)$	q_2^1	q_2^2	q_2^3	q_2^4
i_1	$1/4$	$1/4$	$1/4$	$1/4$
i_2	$1/2$	$1/2$	0	0
i_3	0	0	$1/2$	$1/2$

Table 5: UJS elicitation of $P(Q_2|I)$

Uniform Prior Strategy (UPS). As shown by Example 3, UJS might produce non-uniform values for the marginal

probability mass function over the items. Alternatively, we can impose such a "prior" uniformity, i.e., $P(i) = n^{-1}$ for each $i \in \mathcal{I}$. If we also assume that, given the item, the probability mass is uniformly distributed over all the compatible answers, we have $P(i,q) = (n \cdot v_Q(i))^{-1}$ and, for the conditional probability, $P(q|i) = v_Q(i)^{-1}$. A consequence of this model is that compatible items may have different posterior probabilities, with the more versatile ones being less probable than the less versatile ones. Such a model can then be more suitable to situations where specialization of the item can be considered a positive feature.

Example 4. *Here, still in the setup of Example 1, we apply UPS to the elicitation of the conditional probability table for question Q_2. By definition, the prior is uniform, i.e., $P(i) = {}^1\!/{}_3$ for all entertainers, and the conditional probability table is the same as before. In this case, all items have the same prior probability, whereas, after collecting answer $Q_2 = q_2^1$, being less versatile, user i_2 is assigned a larger probability than i_1, coherently with the fact that the joint distribution assigned a larger probability with the joint event (i_2, q_2^1) than to the event (i_1, q_2^1):*

$$P(i_1|q_2^1) \propto P(q_2^1|i_1) \cdot P(i_1) = \frac{1}{12}, \quad (7)$$

$$P(i_2|q_2^1) \propto P(q_2^1|i_2) \cdot P(i_2) = \frac{1}{6}. \quad (8)$$

$P(i,q)$	q_2^1	q_2^2	q_2^3	q_2^4
i_1	$^1\!/_{12}$	$^1\!/_{12}$	$^1\!/_{12}$	$^1\!/_{12}$
i_2	$^1\!/_6$	$^1\!/_6$	0	0
i_3	0	0	$^1\!/_6$	$^1\!/_6$

Table 6: UPS elicitation of $P(Q_2, I)$

Coping with Multiple Questions

In general, the complete list of m possible questions can be divided into r_J questions $\mathbf{Q}^J$ for which we assume UJS, and a set or r_P questions $\mathbf{Q}^P$ for which we assume UPS (if the single answer compatibility holds for a question, this can be can be arbitrarily included in $\mathbf{Q}^P$ or $\mathbf{Q}^J$). In such general case, because of Equation (3), the joint mass function is given by:

$$P(i, \mathbf{q}) = P(i) P(\mathbf{q}^J|i) P(\mathbf{q}^P|i), \quad (9)$$

where, coherently with the UJS and the UPS strategies described above, we have set

$$P(\mathbf{q}^P|i) = \begin{cases} 1 & \text{if } r_P = 0, \\ \prod_{Q_j \in \mathbf{Q}^P} \frac{\delta(i, q_j)}{v_{Q_j}(i)} & \text{otherwise}, \end{cases} \quad (10)$$

and

$$P(i) P(\mathbf{q}^J|i) = \begin{cases} n^{-1} & \text{if } r_J = 0, \\ \prod_{Q_j \in \mathbf{Q}^J} \frac{\delta(i, q_j)}{N(Q_j)} & \text{otherwise}, \end{cases} \quad (11)$$

from which it follows that $P(q|i) = \delta(i, q_j)[v_{Q_j}(i)]^{-1}$ for all $q \in \mathbf{Q}^J$ and $\mathbf{Q}^P$, and

$$P(i) = \begin{cases} n^{-1} & \text{if } r_J = 0, \\ \prod_{Q_j \in \mathbf{Q}^J} \frac{v_{Q_j}(i)}{N(Q_j)} & \text{otherwise}. \end{cases} \quad (12)$$

Example 5. *In the Stagend example the complete model can be graphically represented by Figure 1, where nodes represent variables, edges describe relations of conditional dependence and the absence of a path connecting two nodes corresponds to conditional independence of the corresponding variables. The conditional probability tables for $P(Q_1|I)$ and $P(Q_2|I)$ are the ones in Tables 3 and 5, respectively, whereas the prior $P(I)$ is given in Table 7 and depends on the strategy adopted for modeling question Q_2.*
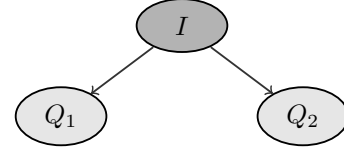

Figure 1: Two questions about the item

Strategy	$P(i_1)$	$P(i_2)$	$P(i_3)$
UJS	$^1\!/_2$	$^1\!/_4$	$^1\!/_4$
UPS	$^1\!/_3$	$^1\!/_3$	$^1\!/_3$

Table 7: UJS and UPS elicitations of $P(I)$

Adaptive Approach

Shaping the Conversation. In classical recommendation systems the assessment of the user preferences with respect to the different items is based on a static block of background information about the user. Such information can be already available in the system or directly obtained from the user after some kind of reduced interaction, e.g., a predefined questionnaire. However, as discussed in the introduction, in modern setups the process of collecting information about the user preferences with respect to the catalogue should be based on a sequence of dynamic interactions. In this view, the questionnaire approach leaves the place to a conversational process taking the form of a personalized sequence of questions dynamically picked from a larger set of questions. The prior probability mass function $P(I)$ is thus sequentially updated each time a new answer is collected, and the updated probability $P(I|\mathbf{Q})$ is used to select the most informative next question. The choice between a possibly huge set of candidate question/interaction can be driven by information-theoretic criteria making any sequence potentially different from the other. In particular, taking inspiration from the literature in the field of adaptive testing, we pick the question that minimizes the conditional entropy (and hence maximizes the expected information gain). This choice is the most natural one as it allows reducing the entropy of the mass function $P(I|\mathbf{Q})$ in order to concentrate most of the probability mass on a limited number of items. More formally, the adaptive conversational process selects the question Q_j^* such that:

$$j^* = \arg \min_{j=1,\ldots,m} H(I|Q_j), \quad (13)$$

where $H(I|Q) := \sum_{q \in \mathcal{Q}} H(I|q)P(q)$ and $H(I|q)$ is the entropy of the posterior mass function over I after the answer $q \in \mathcal{Q}$ to the question Q, and $\{Q_j\}_{j=1}^m$ is the set of questions the system can choose from.

Stopping Rule. This procedure is iterated after any answer and the conversation ends if the posterior entropy $H(I|q)$ decreases under a fixed threshold. As that the entropy of a mass function $P(I)$ is defined $H(I) := -\sum_{i \in \mathcal{I}} P(i) \log_{|\mathcal{I}|} P(i)$, a natural threshold H_s^* is the entropy of a mass function over I which is uniform on s items, while the other ones have zero probability, i.e., $H_s^* := -\log_{|\mathcal{I}|} s^{-1}$. Setting this value in the stopping rule forces the system to halt when most of the posterior probability mass is concentrated on the s most probable items.

Adding Properties

The approach discussed so far allows to properly model the relation between items and questions and dynamically update the probabilities during the conversation. Yet, such an approach is suitable only for small catalogues, as it implicitly requires the assessment of the logical compatibility relations between each item and each questions. As an example, in the case study under consideration, the catalogue includes more than three thousand items (i.e., entertainers), this preventing a straightforward application of the above outline ideas. To bypass such a limitation, we have already introduced a characterization of each item based on the set of compatible answers to each questions. Here, we formalize this characterization and make it independent from the set of questions by introducing the concept of item *properties*. A property C is a random quantity taking its values in a finite set $\mathcal{C} := \{c_1, \ldots, c_r\}$. Let $\mathbf{C} := (C_1, C_2, \ldots, C_p)$ denote the joint set of relevant properties used to characterize an item We assume that this set of properties is a sufficient description of the item I and, consequently, the questions and the item variables are conditionally independent given the properties. Moreover, we initially assume that each question Q refers to a single property C and that questions are conditionally independent given their relative property, as well as properties are conditionally independent given the item I. An example of this augmented setup is reported here below.

Example 6. *In the same setup of Example 1, consider two properties C_1 and C_2 modeling, respectively, the type of enterteiner and the type of event. Question Q_1 refers to property C_1, while Q_2 refers to C_2. Adopting the Markov condition for directed graphs (Koller and Friedman 2009), the above discussed conditional independence assumptions correspond to the graph in Figure 2.*

The simplest way to define properties consists in regarding them as latent (i.e., not directly observable) clones of the questions, which are instead intended as manifest (i.e., observable) variables. In other words, if Q is a question, we define a property C whose possible states $\mathcal{C}$ are in one-to-one correspondence with the answers $\mathcal{Q}$ and we set the logical constraint $C \equiv Q$ meaning that $P(Q = q | C = c) = 1$ if q and c are the elements in the correspondence, and zero
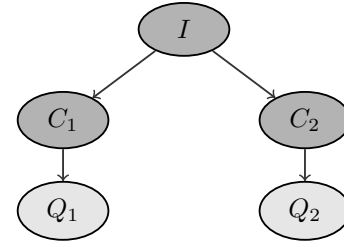

Figure 2: Item, questions and properties relations

otherwise. In terms of compatibility functions, this means $\delta(i, c) = \delta(i, q)$ where q and c are the compatible states of Q and C. Under the above assumptions, the marginalization of the properties simply returns a model equivalent to the one defined in the previous section. More expressive setups can be obtained by specifying properties that are not latent clones of the questions they are associated to. This allows to describe items based on general properties, streamlining the elicitation of compatibility relations between items and questions. For instance, a state c of property C can imply both logical requirements of the answers q' and q''. Such a property would be compatible with both answers, but equivalent to none of them. This situation can be modeled by conditional probabilities $P(q'|c)$, $P(q''|c)$ strictly larger than zero.

Example 7. *In the Stagend platform, it is well known that the organizer of an event may be well satisfied with a band even if she asked only for a musician. This can be modeled by assigning a positive value to the probability $P(q_1^3|c_1^2)$ of asking for a musician when the best matching item i^* is a band (i.e., $C(i^*) = c_1^2$ corresponding to Band).*

Moreover, the possibility of grouping items based on their properties can be exploited to learn model parameters with less fragmentation in the data. E.g., the probability $P(q|c)$ can be estimated from all selected items having property c; instead, to estimate $P(q|i)$ one should consider only the cases where item i has been chosen, which can be very few or even zero for new items.

Finally, more complex relations between questions and properties or property values can be easily modeled. Some examples are: (i) multiple questions referring to the same property, modeled by assuming conditional independence between the questions, this simply requires to independently elicit the probabilities $P(Q_1|C)$ and $P(Q_2|C)$ for all possible states of Q_1 and Q_2 and C; (ii) multiple properties for the same question, if $\mathbf{C}_Q$ is the set of properties associated with Q, the elicitation of the probabilities $P(q|\mathbf{C}_Q)$ for all values of q and all joint combinations of states for the properties in $\mathbf{C}_Q$ is required; (iii) questions relevant only for some items, e.g., Stagend question *Do you want the musician to play any particular instrument?* should define a preference among musicians, without changing the probability of the property *musician* with respect to *band*, *DJ* and *entertainer* and can be modeled by assuming $P(q|i) = P(q)$ for all items i for which the question is not relevant.

Example 8. *In the Stagend example, the most requested en-*

tertainers for weddings are musicians and DJs. To model this we remove the assumption of marginal independence between the properties Type of even *and* Type of entertainer, *i.e.,* C_1 *and* C_2. *In the graphical language of Figure 2, this corresponds to adding an edge connecting the two properties (Figure 8). To define the corresponding model, the elicitation of the conditional probabilities* $P(c_1|c_2, i)$ *for all possible values of* C_1 *conditioned over all possible combinations of values* (c_2, i) *is required. Then, in this toy example, where only three entertainers have been considered, adding the above dependency between properties requires elicitation of 36 parameters values which can become many more in real applications. It would be much easier to reason about the marginal probability* $P(C_1|C_2)$ *of the type of entertainer given the type of event.*
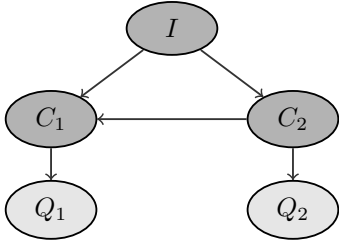
Figure 3: Graph representing the model in Example 8

Elicitating Dependencies between Properties

In the previous section we noticed that adding dependencies between properties increases the number of model parameters (e.g., conditional probabilities), and this prevents a fast elicitation for huge catalogues.

As we did for the conditional probability tables of the questions (i.e., $P(Q|C)$), we could better perform a single elicitation for all the items with the same value for the properties of interest, that is, focusing only on properties (and not on items) in the knowledge-based elicitation of probabilities, and grouping all items with the same property values when learning from data.

This can be achieved as follows. Assume we wish to model the conditional probability table for property C given a set of parent properties $\mathbf{C}_p$, while the remaining properties are assumed independent from C and denoted by $\mathbf{C}_{np}$. The procedure for eliciting the (possibly huge) number of conditional probabilities $P(c|\mathbf{c}_p, i)$ for all values c of C and all joint combinations $(\mathbf{c}_p, i)$ consists of two steps. First, the conditional probabilities $P(c|\mathbf{c}_p)$ marginalized over i are elicited based on prior knowledge, data or both, while the conditional probabilities $P(i|\mathbf{c})$, where $\mathbf{c} = [c, \mathbf{c}_p, \mathbf{c}_{np}]$ is the vector of all property values, is elicited based on logic constraints. Afterwards, probabilities $P(c|\mathbf{c}_p, i)$ are then be derived from the relation:

$$P(c|\mathbf{c}_p, i) = \frac{P(i|\mathbf{c})P(c|\mathbf{c}_p)P(\mathbf{c}_p)}{\sum_c P(i|\mathbf{c})P(c|\mathbf{c}_p)P(\mathbf{c}_p)} , \qquad (14)$$

while the prior is derived from

$$P(i) = \sum_{\mathbf{c}} P(i|\mathbf{c})P(c|\mathbf{c}_p)P(\mathbf{c}_{np}) . \qquad (15)$$

Concerning the conditional $P(i|\mathbf{c})$, it is derived from the joint $P(i, \mathbf{c})$ which has the same form as $P(i, \mathbf{q})$ in Equation (9) with $\mathbf{q}^J$ and $\mathbf{q}^P$ replaced by the vectors of properties $\mathbf{c}^J$ and $\mathbf{c}^P$ for which we assume, respectively, UJS and UPS. $P(i|\mathbf{c})$ is then given by:

$$P(i|\mathbf{c}) = \frac{\prod_{c_j \in \mathbf{c}} \delta(i, c_j)}{\sum_i \prod_{c_j \in \mathbf{c}} \delta(i, c_j)} = \frac{\delta(i, \mathbf{c})}{N(\mathbf{c})} , \qquad (16)$$

where $N(\mathbf{c}) = \sum_i \delta(i, \mathbf{c})$ is the number of items compatible with the combination of property values in $\mathbf{c}$. Notice that in case no item is compatible with $\mathbf{c}$ the probability $P(i|\mathbf{c})$ cannot be derived from Equation (16). This inconsistency arises from the fact that only the vectors of property values $\mathbf{c}$ that are compatible with at least one of the item in the catalogue, are, indeed, possible. Let $\mathcal{J}_{\mathbf{c}}^*$ be the set of all possible $\mathbf{c}$ for which $\exists i : i \models \mathbf{c}$. To solve the above inconsistency and account for the logical impossibility of all $\mathbf{c} \notin \mathcal{J}_{\mathbf{c}}^*$, an initial, eventually inconsistent, elicitation of empirical estimate $P'(\mathbf{c})$ of the probability of the joint properties needs to be revised as follows:

$$P(\mathbf{c}) = \begin{cases} \frac{P'(\mathbf{c})}{\sum_{\mathbf{c} \in \mathcal{J}_{\mathbf{c}}^*} P'(\mathbf{c})} & \text{if } \mathbf{c} \in \mathcal{J}_{\mathbf{c}}^* \\ 0 & \text{otherwise} . \end{cases} \qquad (17)$$

Finally, by replacing Equation (16) into Equation (14) we obtain the conditional probabilities for the dependent property C:

$$P(c|\mathbf{c}_p, i) = \frac{\delta(i, c)\delta(i, \mathbf{c}_p)P(c|\mathbf{c}_p)}{\sum_{c^j : i \models c^j} P(c|\mathbf{c}_p)} , \qquad (18)$$

which corresponds to the conditional $P(c|\mathbf{c}_p)$ re-normalized considering only the values of c compatible with i. By replacing 16 into 15 we obtain the marginal distribution of i:

$$P(i) = \sum_{\mathbf{c}:i \models \mathbf{c}} \frac{P(\mathbf{c})}{N(\mathbf{c})} . \qquad (19)$$

Notice that the probability of each combination $\mathbf{c}$ of properties values is uniformly distributed over all items that are compatible with that combination.

Example 9. *In the Stagend example the two steps above correspond to the elicitation of the conditional probabilities of the model in figure 9. The only difference compared to the desired model (figure 8) is the direction of the edges connecting* I *to* C_1 *and* C_2. *Therefore, the two models define the same set of independence assumptions. Assume that the relation between properties can be modeled by the conditional probabilities in the square brackets of Table 9, which describe a situation where the preferred types of entertainers are Djs and musicians for weddings, bands and entertainers for corporate events, Djs and entertainers for birthdays, Djs and bands for parties. Moreover, we assume for* C_2 *the prior probabilities* $P(c_2^1) = 2/3$, $P(c_2^2) = P(c_2^3) = P(c_2^4) = 1/9$, *according to which weddings are twice more popular than any other type of event. The corresponding joint model* $P(C_1, C_2) = P(C_1|C_2)P(C_2)$ *however, does not comply with the logical impossibility of all joint states* (c_1, c_2) *that are not compatible with any item, requiring such joint states*

to have zero probability. As all values of C_2 are compatible with at least one item, we focus on $P(C_1|C_2)$. Cells corresponding to impossible joint states are highlighted in grey in Table 9. Their values are set to zero and the re-normalized probabilities are shown in the table next to the initial ones.

$P(C_1|C_2)$	c_1^1 =DJ	c_1^2 =Band	c_1^3 =Musician	c_1^4 =Entert.
c_2^1 =Wedding	$2/3$ $(1/3)$	$1/3$ $(1/6)$	0 $(1/3)$	0 $(1/6)$
c_2^2 =Corp.Ev.	$1/3$ $(1/6)$	$2/3$ $(1/3)$	0 $(1/6)$	0 $(1/3)$
c_2^3 =Birthday	$1/2$ $(1/3)$	0 $(1/6)$	0 $(1/6)$	$1/2$ $(1/3)$
c_2^4 =Party	$2/3$ $(1/3)$	0 $(1/3)$	0 $(1/6)$	$1/3$ $(1/6)$

Table 8: Elicitation of $P(C_1|C_2)$

From this elicitation the prior probabilities $P(i_1) = 11/18$, $P(i_2) = 8/27$, $P(i_3) = 5/54$ and the conditional $P(C_1|C_2, i) = \delta(i, c_1)$ follow. Therefore, in this simple case where each artist is compatible with one single value of property C_1, we have that $P(C_1|C_2, i)$ reduces to the zero-entropy conditional $P(C_1|i)$ based only on logical constraints, whereas the defined dependence between C_1 and C_2 only affects the prior distribution over items. If a further entertainer i_4 performing at weddings both as DJs and as a band were included in the catalog, it would result in strictly positive probabilities $P(c_1|c_2^1, i_4) = P(c_1|c_2^1)$ both for c_1 equal to band and Dj.
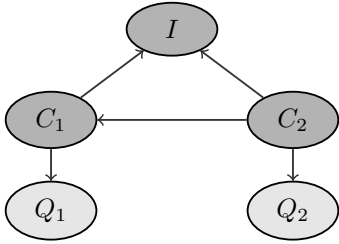
Figure 4: Graph representing the model in Example 9

Learning from Data

The above discussed procedure for the elicitation of the model parameters in the recommendation system is based on structural judgement about the logical compatibility between items, properties and the answers to be possibly integrated by judgements of a domain expert about the relations between the properties (irrespectively of the items). Yet, historical data involving observations of the parameters to be quantified might be available. Following a Bayesian paradigm, we can naturally merge these two sources of information by using the outputs of the elicitation process as the parameter of a multinomial Dirichlet prior to be combined with the likelihood of the observed data. This has the potential of further increasing the discriminative power of the system and the quality of the recommendations.

Experiments

For an empirical validation we consider the Stagend recommendation system. Currently, the platform includes $n =$
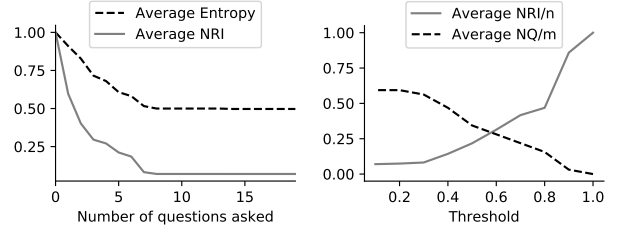
Figure 5: Conversational approach performance

3520 entertainers. In its static version, the system asks to all the interested users a questionnaire including $m = 32$ questions intended to identify the entertainer that matches at best the needs of the users. Stagend advisors select a small subset of artists (in general less than ten) to be presented to each user. The questions identify 13 properties used to characterize the entertainers. We model the relation between questions and items by means of the compatibility-based elicitation and simulate a conversational process using the answers collected in 100 questionnaires filled by actual Stagend users. At the end of each simulated conversation, a set of items is retained by our model. When all questions are asked, in 9 cases over the 100 simulated, the final set is empty. This shows that the notion of logical compatibility can be too strict in some cases. For the remaining questionnaires, we compare the decision of our algorithm to the subset of items selected by Stagend advisors. Let FI denote the fraction of items suggested by the advisors that are actually retained by our model, the average FI over the 91 remaining simulations is 70%. Again, this can be explained by the fact that logical compatibility is not always fully respected by the advisors suggestions, sometimes due to a poor knowledge of the entertainers in the catalogue, sometimes to better diversify the offer. Figure 5 (left) shows how the entropy of $P(I|\mathbf{q})$ as well as the *number of retained entertainers* (NRI) decreases with the number of question asked. However, after a certain number of questions, both the entropy and the NRI stop decreasing. Notice that the order of the question is defined in an adaptive way. In the right-hand side of Figure 5 it is shown how the fraction of retained entertainers (NRI/n) and the fraction of questions asked (NQ/m) varies with respect to the entropy threshold.

Conclusions

A new approach to automatic recommendations which assumes a dynamic interaction between the system and the user to provide customized and self-adaptive recommendations has been developed on the basis of a pure Bayesian approach. The framework introduced in this paper sets the ground to several future developments, among which the dynamic generations of questions in order to improve the conversational nature of the system. This could be based on a natural language generation system interacting with the structured probabilistic description of item properties and elicitation of user needs, (Mostafazadeh et al. 2016).

References

Adomavicius, G., and Tuzhilin, A. 2011. Context-aware recommender systems. In *Recommender systems handbook*. Springer. 217–253.

Butz, C. J.; Hua, S.; and Maguire, R. B. 2006. A web-based Bayesian intelligent tutoring system for computer programming. *Web Intelligence and Agent Systems: An International Journal* 4(1):77–97.

He, C.; Parra, D.; and Verbert, K. 2016. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications* 56:9–27.

Jugovac, M., and Jannach, D. 2017. Interacting with recommenders - overview and research directions. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7(3):10.

Koller, D., and Friedman, N. 2009. *Probabilistic graphical models: principles and techniques*. MIT press.

Lu, J.; Wu, D.; Mao, M.; Wang, W.; and Zhang, G. 2015. Recommender system application developments: a survey. *Decision Support Systems* 74:12–32.

Mangili, F.; Bonesana, C.; and Antonucci, A. 2017. Reliable knowledge-based adaptive tests by credal networks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, 282–291. Springer.

Millán, E.; Trella, M.; Pérez-de-la Cruz, J.-L.; and Conejo, R. 2000. Using Bayesian networks in computerized adaptive tests. In *Computers and Education in the 21st Century*. Springer. 217–228.

Mostafazadeh, N.; Misra, I.; Devlin, J.; Mitchell, M.; He, X.; and Vanderwende, L. 2016. Generating natural questions about an image. *arXiv preprint arXiv:1603.06059*.

Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J.; et al. 2000. Analysis of recommendation algorithms for e-commerce. In *EC*, 158–167.

Wainer, H.; Dorans, N. J.; Flaugher, R.; Green, B. F.; and Mislevy, R. J. 2000. *Computerized adaptive testing: a primer*. Routledge.