Customizing Question Selection in Conversational Case-Based Reasoning

Vahid Jalali and David Leake
School of Informatics and Computing, Indiana University
Informatics West, 901 E. 10th Street
Bloomington, Indiana, 47408

Abstract

Conversational case-based reasoning systems use an interactive dialog to retrieve stored cases. Normally the ordering of questions in this dialog is chosen based only on their discriminativeness. However, because the user may not be able to answer all questions, even highly discriminative questions are not guaranteed to provide information. This paper presents a customization method CCBR systems can apply to adjust entropy-based discriminativeness considerations by predictions of user ability to answer questions. The method uses a naive Bayesian classifier to classify users into user groups based on the questions they answer, applies information from group profiles to predict which future questions they are likely to be able to answer, and selects the next questions to ask based on a combination of information gain and response likelihood. The method was evaluated for a mix of simulated user groups, each associated with particular probabilities for answering questions about each case indexing feature, in four sample domains. For simulated users with varying abilities to answer particular questions, results showed improvement in dialog length over a non-customized entropy-based approach in all test domains.

Introduction

Conversational case based reasoning (CCBR) (Aha and Breslow 1997) is an interactive form of case based reasoning (CBR) (Leake 1996; Mantaras et al. 2005). CCBR retrieves cases by asking users a series of questions about features of a target and choosing questions aimed at discriminating between alternative cases in the case base. Incremental responses narrow the set of candidate cases until a single matching case is retrieved. CCBR has been widely applied to tasks such as help desks (Watson 1997) and E-commerce product recommendation. The standard CCBR approaches to question selection focus on selecting discriminative questions, whose answers will lead to rapid selection of a target case. However, the discriminativeness of a question is only part of whether the question is a good choice to ask: It is not guaranteed that the user will be able to answer all system questions, and asking unanswerable questions needlessly increases dialog length.

This paper proposes refining the standard CCBR process by making question selection sensitive not only to the discriminativeness of the questions, but to the likelihood of the user being able to answer the question, based on classifying the user. This refinement is aimed at improving the efficiency of the CCBR dialog, as measured by the average dialog length required to retrieve the target case. For example, if CCBR is applied to cell phone recommendation, a nonexpert user may be unfamiliar with certain technical characteristics of the phones, making highly technical questions unlikely to yield information, even if they might rapidly discriminate between different types of phones. Normally, CCBR systems provide a small sampling of alternative questions from which the user may choose, providing some robustness to user inability to answer particular questions. Nevertheless, including unanswerable questions can displace answerable questions, and can decrease interaction quality by requiring users to read through unhelpful questions in in each iteration. Thus in question selection it is important to balance the discriminativeness of features against the likelihood of the user being able to supply their values.

This paper proposes a domain-independent question selection method based on modifying standard entropy calculations for CCBR to take into account the user likelihood of answering a question. The approach assumes that the system has been provided with profiles of different user groups which provide the probabilities of members of each group being able to answer questions about each possible case feature. During a question dialog, the approach applies a Naive Bayesian classifier to incremental information about users' past answers to predict group membership, which is then used to predict the user's ability to answer particular future questions. That information is in turn used to guide question selection, based on a method for balancing question discriminativeness against accessibility. An evaluation of the method in four different domains shows improvement over a baseline uncustomized approach, when users do not have perfect ability to answer system questions.

Related Work

Dialog length in CCBR systems depends both on case representation and question selection. For example, Gupta's (2001) taxonomic CCBR focuses on case representation, proposing improving performance by using taxonomies to handle abstraction in domain features. For question selection, many approaches are based on refinements of the general entropy-based approach found in decision tree learning (Quinlan 1986), selecting the most discriminative feature at each step based on the information gain of that feature. For example, Doyle and Cunningham (2000) apply interactive CBR to the field of on-line decision guides, comparing a number of pre-clustering algorithms for calculating information gain of different features in terms of the average dialog length. In the context of fault diagnosis and E-Commerce product selection domains, McSherry (2001) proposes minimizing the dialog length by information splitting criteria based on information gain, for identification trees.

In some domains, there is a tradeoff between the distinctiveness of features selected by the system and other factors that should be considered. For example, in medical diagnosis, the medical tests required for determining classifications may have differing costs. Carrick et al. (1999) propose a CCBR process in which question selection is based on a combination of measures for information quality and the estimated cost of answering questions, where cost is estimated by formulating a hierarchical plan for the process for gathering the needed information and evaluating its cost.

Kohlmaier et al. (2001) distinguish four different scenarios for user answers to product-selection questions in an E-commerce domain: answer without help, answer with help, have no preference, or fail to understand. Their method then updates the information gain of features by applying a penalty factor to cases where the user is not able to answer a question. This penalty factor represents the information lost.

In their model, user behavior is predicted by using a Bayesian Network and built-in priori distribution of customers' domain knowledge supplied by the shop owner. Factors other than feature selection cost could be balanced against informativeness as well. For example, McSherry (2011) considers the tradeoff between accuracy and efficiency in CCBR dialogs.

The role of customization has been studied by Thompson et al. (2004), who propose a personalized conversational recommendation system to help users select a restaurant (or a destination in general) that meets their preferences. They capture an individual's choices interacting with their system over time to generate a user model, which is then used to guide future question selection. Our approach differs in addressing situations for which the user is unknown to the system, so must be classified during (and based on) the ongoing dialog, and in the new entropy-based approach developed here for question selection.

Ricci et al. (2003) distinguish between content and collaborative features in CCBR. As an example, for hotel recommendation, they consider hotel rating and parking availability as content features and the user nationality as a collaborative feature. They use a hybrid approach for product recommendation which uses content and collaborative features for ranking the cases based on both previous similar sessions and similar cases in case base. However, sometimes the collaborative features are not explicitly captured, or candidate collaborative features may be unknown. For example, in cell phone recommendation, a specific characteristic of the user (e.g. being familiar with technical features) could be potentially more beneficial in conducting the dialog compared to another collaborative feature captured explicitly from the user profile (e.g. nationality of the user). We hypothesize that it is possible to infer relevant implicit collaborative features, based on the answered/skipped questions by the user in a dialog and other previously stored sessions. In this regard, our method is more of a Data-Driven approach rather than a Model-Driven one like Ricci's.

Classic Information Gain

For a set S, divided into c categories, where p_i is the proportion of S belonging to class i, the entropy of the set is defined as (Shannon, Petigara, and Seshasai 1948):

$$Entropy(S) \equiv \sum_{i=1}^{c} -p_i \log_2 p_i \tag{1}$$

If f is a feature of the elements of S, the ability of knowledge of that feature value to predict category membership—the information gain provided by knowledge of the value of feature f—is defined by:

$$Gain(S, f) \equiv Entropy(S) - \sum_{v \in Values(f)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

where Values(f) is the set of all possible values for feature f, and S_v is the subset of S for which feature f has value v. (i.e., $S_v = s \in S | f(s) = v$).

Customizing Question Selection by User Group

Different users may be knowledgeable about different feature types. If the system requests information outside the user's knowledge, the user be unable to furnish information about those features. For example, novice computer buyers may be able to describe preferences for price or types of use for their purchases (e.g., buying primarily for email vs. for gaming), but might be unlikely to know technical differences (e.g., DDR vs. DDR2 vs. DDR3 memory).

For systems used routinely by a fixed set of users who consistently make particular types of purchases (e.g. a focused E-Commerce website with repeat customers), the user ability to answer particular query types may already be known. Exploiting this information depends on the ability to balance information gain and accessibility information to influence question selection.

For systems which may be used by unknown users (e.g., a sight-seeing recommendation system in the lobby of a hotel), or a recommendation system for infrequent purchases or purchases for which the relevant feature set may change over time, with new features added (e.g., a recommendation system for cell phones), it may be necessary to infer likely user characteristics from experience with previous users. Our approach to inferring user characteristics assumes the existence of group profiles, and predicts feature accessibility based on assigning users to particular user groups. Such group-level customization requires addressing two questions: how to classify a user based on incremental information gathered during a dialogs (in order to access group-based accessibility information), and how to balance information gain and accessibility information to influence question selection.

Classifying Users During CCBR Dialogs

In our method, user groups collect users with similar ability/interest in providing information about particular features in response to questions in the CCBR dialog. Given a user whose group is unknown, the method attempts to infer the most appropriate user group, based on which questions were answered or left unanswered in previous steps of the current dialog.

After a question about feature f has been asked by the system, the probability of the user belonging to each user group can be updated based on whether or not the user answered the question. Given a set U of user groups, and for $mem(u)$ denoting the user being a member of group u, and $d(f,v)$ denoting the value provided by the user for f (which may be a feature value or "skip"), the probability of the user belonging to each group can be updated by Bayes' rule as follows:

$$\forall u \in U, P(mem(u)|d(f,v)) = \frac{P(d(f,v)|mem(u)) \times P_{last}(mem(u))}{P(d(f,v))} \quad (3)$$

where $P_{last}(mem(u))$ is the $P(mem(u)|d(f,v))$ calculated from the previous question (for the first question, this uses a prior probability provided to the system with the group profile). $P(d(f,v)|mem(u))$ is provided to the system as part of the group profile for each group. $P(d(f,v))$ can be calculated as follows:

$$P(d(f,v)) = \sum_{u \in U}(P(d(f,v)|mem(u)) \times P_{last}(mem(u))) \quad (4)$$

Using these formulas the system incrementally refines its probabilities of group membership as questions are answered. These probabilities are used to choose which features to ask at each step, as described in the following section.

Adjusting Information Gain to Reflect Accessibility

To reflect the varying likelihood of different users supplying a feature value, we propose the following adjustment of the formula for the information gain of splitting set S by feature f, for user group u (where AGain stands for Accessibility-influenced Gain):

$$AGain(S,f,u) \equiv Gain(S,f) \times \sum_{v \in V} \frac{SelProb(u,v)}{|V|} \quad (5)$$

where SelProb(u,v) is the probability of a user belonging to group u being able to choose value v for feature f, and $|V|$ represents the number possible values for feature f. The sum $\sum_{v \in V} \frac{P(u,v)}{|V|}$ is the average probability that a user who is in group u, provides one or more values of feature f. Therefore, if the average probability of choosing a value for feature f by the user is relatively low, AGain is lower than for an equally discriminating feature which is more likely to be answered by the user.

Accessibility Influenced Attribute Selection (AIAS)

We apply the previous formulas in an approach we call Accessibility Influenced Attribute Selection (AIAS). The goal of AIAS is simultaneously to infer the user group of a user (based on which of the presented questions the user answered) and to apply that information to select future questions which both have discriminating power and are likely for the user to be able to answer.

For a given set of cases S, AIAS ranks features by:

$$FeatureVal(S,f) = \sum_{u \in U} P(mem(u)) \times AGain(S,f,u) \quad (6)$$

This formula is used in the AIAS process for guiding CCBR dialogs, presented in Alg. 1.

Algorithm 1 AIAS Algorithm

Input: Set of cases:S, features:F, User groups: U, feature accessibility information.
Output: Selected case

```
repeat
  for all Feature f ∈ F do
    Calculate featureVal(S,f)
  end for
  F_max ← top m features with highest featureVal
  for all Feature f ∈ F_max do
    if user selects v for f then
      S ← S − {s ∈ S | f(s) ≠ v}
    else if f is skipped by user then
      F ← F − f
    end if
    for all User u ∈ U do
      Update P(u)
    end for
  end for
until either (a) |S| == 1, or (b) ∀ f ∈ F, ∄ s1, s2 ∈ S , f(s1) ≠ f(s2) or (c) user stops dialog
```

AIRS Architecture

We have implemented AIAS in AIRS (Accessibility Influenced Recommender System), a domain-independent conversational CBR system whose architecture is shown in Fig. 1. Based on knowledge of user groups and group-based accessibility, the system presents queries to the user, and responses are input both to (1) a user-based accessibility estimator, and (2), a case retrieval engine. The user-based accessibility detector is a naive Bayesian learner which predicts the group to which the user belongs based on all answered/skipped questions so far. The case retrieval engine retrieves the appropriate cases from the case-base according to the user query. The feature value selector module calculates the ranking of the features based on their information gain and accessibility. Finally, questions about top ranked features and their associated values are presented to the user. The process is repeated until a desired case is found.

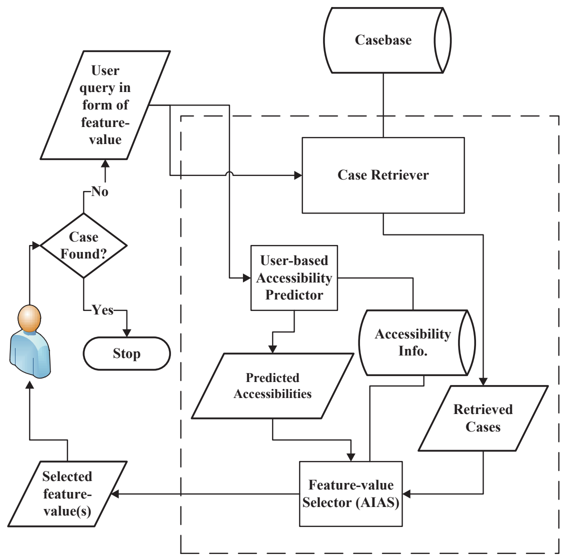

Figure 1: The architecture of AIRS.

Experimental Design

To evaluate AIAS, we conducted experiments to address two questions:

1. How do varying feature accessibilities affect the dialog length for the different methods?
2. How does the prediction of user group-based feature accessibility affect dialog length?

We applied AIRS to four sample domains with varying numbers of cases and attributes per case:

- Cell phone: The cell phone dataset, distributed by Nokia[1] contains 306 cell-phone models and 436 unique features (this dataset originally contained multi-valued features; we decomposed them into single-valued features and extracted 436 candidate features).
- Restaurant: The restaurant dataset, extracted from chefmoz.org, contains information about 48 restaurants with 52 distinct features. (this dataset originally contained multi-valued features; we decomposed them into single-valued features)
- Automobile: The Automobile dataset (Frank and Asuncion 2010) contains information on 205 automobiles with 26 features.
- University: The University dataset (Frank and Asuncion 2010) contains 285 Universities and 17 features.

[1] $http://www.developer.nokia.com/gen/all\_devices.rss$

For each domain, for each user group in the domain, ten tests were run. In each test, a simulation was run in which a simulated user interacted with the CCBR process to select cases. Tests for a given simulated user were conducted by a "leave one in" method (Aha, Maney, and Breslow 1998). Each case in the case base is used once as a target, but retained in the case base. In all, 24,630 CCBR conversations were simulated.

The simulated user interacts with the system until either the target case is the sole remaining case or all the remaining cases have identical values for their unselected feature-value pairs. The user's ability to answer the question for a given feature was determined randomly, with a probability determined by the feature accessibility information for the user group. The average dialog lengths for all trials are used as the efficiency indicator of each method.

Because CCBR systems often present more than one question to the user, simulated users are "presented" with more than one question at a time. The number of questions presented was chosen based on the differing characteristics of the test domains. For case bases with small numbers of features, to presenting many questions corresponds to placing the feature selection burden on the user, providing a less informative comparision of feature selection strategies. For the cell phone domain, our test domain with the largest number of cases and features, users were presented with the top 5 questions (1% of total features). For the Restaurant, Automobile and University domains, 3, 1, and 1 questions were presented (6%, 4%, and 7% of total features). Whether the simulated user can answer a question is determined randomly, with the probablility associated to its corresponding feature. The simularted user starts at the top of the list and proceeds until answering a question.

User Groups for Test Domains

For each test domain, probabilities were assigned to reflect feature accessibilities for different groups. In the cell phone domain, two groups were defined, novice and expert users. For simulating novice users, features such as "keypad layout" and "video features" were considered accessible. For simulating expert users, technical features such as "CPU clock rate" and "shared memory size" were considered accessible as well.

In the restaurant domain, two groups were defined, users primarily concerned with cost or primarily concerned with the restaurant's food quality and atmosphere. For these users, the failure to respond to a particular feature question reflects not having a preference on that feature. In the automobile domain, groups were defined for users interested in economy cars, users interested in luxury cars, and users interested in family cars. In the university domain, groups were defined for users most concerned about admissions requirements, users most concerned about educational quality, users most concerned about funding and expenses, and users most concerned about acceptance rates.

Conditions Tested We compare results for four selection methods:

- Entropy-based Ideal User: This is a baseline for situations in which feature accessibility is not an issue. The most distinctive questions are asked and the simulated user is able to answer all questions.

- Entropy-based Realistic User: The most distinctive questions are returned to the user and the simulated user's ability to answer questions depends probabilistically on the accessibility information of its group.
- AIAS for Known User Group: Questions are ordered by AIAS (based on their distinctiveness the their likelihood of being answered by users in a given group) and the user group is known to the system (i.e. the system does not need to learn the user group).
- AIAS: Questions are ordered based on their distinctiveness and their likelihood of being answered by the simulated user, based on inference of the user's group.

Experimental Results

AIAS vs. Entropy-based Feature Selection

Table 1 summarizes the average dialog length of four sample feature selection methods. As expected, the best average dialog length belongs to the Entropy-based Ideal User case in which the simulated user is able to answer all questions. However, when the user cannot answer all questions, performance of the entropy-based approach is significantly degraded in all domains. When the user group is known, the AIAS strategy provides substantial improvement. Even when the user group is initially unknown, the AIAS method for inferring user group provides almost identical performance. This supports the ability of the approach to classify users sufficiently accurately and rapidly to benefit the dialog process when features have varying accessibility.

	Domain			
Approach	Cell-phone	Restaurant	Automobile	University
Entropy-based Ideal User	1.41	2.71	3.0	2.08
Entropy-based	3.84	6.96	6.59	4.19
AIAS Known User Group	2.69	6.11	5.55	2.38
AIAS	2.74	6.14	5.59	2.67

Table 1: Average dialog length of studied feature selection methods in four sample domains

Fig. 2 compares the efficiency of AIAS compared to the Entropy-based approach for the realistic (non-ideal) user. AIAS decreases the average dialog length compared to Entropy-based approach in all sample domains.

When a CCBR system presents multiple questions simultaneously, the dialog length is not the only factor of interest: The number of questions the user must consider before being able to provide an answer is important as well. Table 2 shows the average number of questions skipped by simulated users in the sample domains. The average number of skipped questions for the entropy-based approach is always greater than for AIAS, due to AIAS avoiding asking questions expected to be difficult for the user to answer.

Fig. 3 illustrates the percentage of increase in the average number of skipped questions of the Entropy-based approach over AIAS.

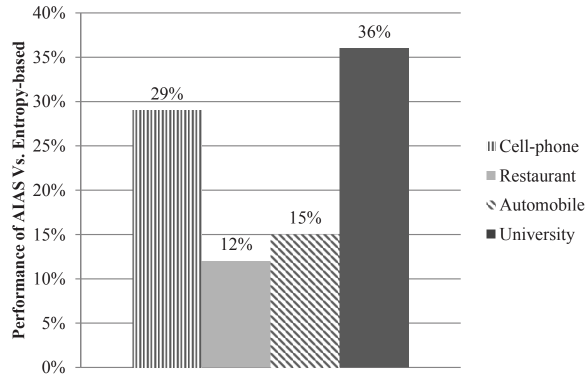

Figure 2: Percent improvement of dialog length of AIAS compared to the entropy-based approach when feature accessibility varies.

	Domain			
Approach	Cell-phone	Restaurant	Automobile	University
Entropy-based	4.21	2.39	5.8	1.99
AIAS	1.96	2.42	2.06	0.02

Table 2: Average number of skipped questions by AIAS and entropy-based methods when feature accessibility varies.

Discussion and Future Work

The previous section shows that using accessibility information in choosing the questions to be presented to a user can improve CCBR dialog efficiency when features have different accessibilities. In experiments in four sample domains, the average dialog length of AIAS showed between 12% to 36% improvement over the entropy-based feature selection approach for this situation.

In future work, we intend to explore learning user groups and assessing answer probabilities for group members on the fly, rather than requiring that these are provided to the system. Even when groups are given, it would be desir-

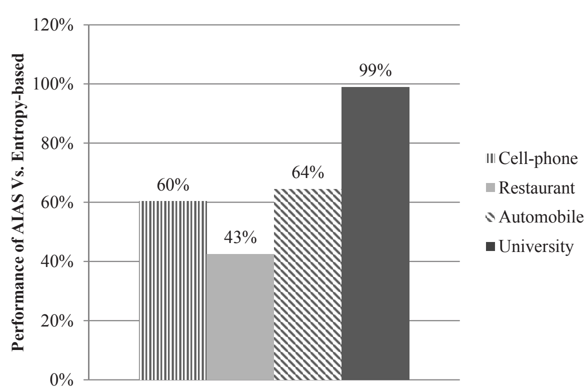

Figure 3: Percent of AIAS improvement over Entropy-based approach in terms of the average number of skipped questions in four sample domains.

able for the system to dynamically update those probabilities based on the user interactions with the system. However, this would require additional considerations such as the system's confidence in the classification of a user, as the system should only update the probabilities for a group based on a user's performance if it is confident about the user's group.

Another issue is how to balance the value of asking questions aimed at identifying user groups, versus aimed at identifying the target item. In the simplest scenario, if two questions are equally discriminating and have equal accessibility, the system should ask the question expected to help identify the user's group faster, in order to improve focusing of future questions.

Additional factors influencing question selection could be included in the question selection calculations as well, such as the cost for users of responding to the questions they are able to answer (cf. (Carrick et al. 1999)).

Conclusion

This paper proposed AIAS, accessibility-influenced attribute selection for conversational case-based reasoning, a method which considers both entropy-based feature discriminativeness and accessibility when selecting questions. AIAS predicts the user group based on user's interactions with the system, and uses group-based information at each step to select features. The method uses a naive Bayesian classifier for predicting the group to which the user belongs. Predicting the user group enables the approach to use the appropriate accessibility information in calculating the information gain of the features. In order to assess the efficiency of AIAS, experiments with simulated users were conducted in four different recommendation domains. Using the average dialog length as the efficiency indicator of each studied method. Results show 12%, 15%, 29% and 36% improvement for AIAS over the baseline in the studied domains. Possible future directions include strategic question selection for rapid identification of user groups and consideration of additional factors which may influence the desirability of asking particular questions.

Acknowledgments

This work is supported in part by a grant from the Data to Insight Center at Indiana University.

References

Aha, D., and Breslow, L. 1997. Refining conversational case libraries. In *Proceedings of the Second International Conference on Case-Based Reasoning*, 267–278. Berlin: Springer Verlag.

Aha, D.; Maney, T.; and Breslow, L. 1998. Supporting dialogue inferencing in conversational case-based reasoning. In Cunningham, P.; Smyth, B.; and Keane, M., eds., *Proceedings of the Fourth European Workshop on Case-Based Reasoning*, 262–273. Berlin: Springer Verlag.

Carrick, C.; Yang, Q.; Abi-Zeid, I.; and Lamontagne, L. 1999. Activating CBR systems through autonomous information gathering. In *Proceedings of the Third International Conference on Case-Based Reasoning*, 74–88. Berlin: Springer Verlag.

Doyle, M., and Cunningham, P. 2000. A dynamic approach to reducing dialog in on-line decision guides. In *EWCBR*, 49–60.

Frank, A., and Asuncion, A. 2010. UCI machine learning repository.

Gupta, K. 2001. Taxonomic conversational case-based reasoning. In *Proceedings of the Fourth International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development*, ICCBR '01, 219–233. Berlin: Springer-Verlag.

Kohlmaier, A.; Schmitt, S.; and Bergmann, R. 2001. A similarity-based approach to attribute selection in user-adaptive sales dialogues. In *Case-Based Reasoning Research and Development: Proceedings of the Fourth International Conference on Case-Based Reasoning*, 306–320. Berlin: Springer Verlag.

Leake, D., ed. 1996. *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. Menlo Park, CA: AAAI Press/MIT Press.

Mantaras, R.; McSherry, D.; Bridge, D.; Leake, D.; Smyth, B.; Craw, S.; Faltings, B.; Maher, M.; Cox, M.; Forbus, K.; Keane, M.; Aamodt, A.; and Watson, I. 2005. Retrieval, reuse, revision, and retention in CBR. *Knowledge Engineering Review* 20(3).

McSherry, D. 2001. Minimizing dialog length in interactive case-based reasoning. In *Proceedings of the seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, 993–998. San Mateo: Morgan Kaufmann.

McSherry, D. 2011. Conversational case-based reasoning in medical decision making. *Artificial Intelligence in Medicine* 52(2):59–66.

Quinlan, J. R. 1986. Induction of decision trees. *Machine Learning* 1(1):81–106.

Ricci, F.; Venturini, A.; Cavada, D.; Mirzadeh, N.; Blaas, D.; and Nones, M. 2003. Product recommendation with interactive query management and twofold similarity. In *Proceedings of the 5th international conference on Case-based reasoning: Research and Development*, ICCBR'03, 479–493. Berlin, Heidelberg: Springer-Verlag.

Shannon, C.; Petigara, N.; and Seshasai, S. 1948. A mathematical theory of communication. *Communication, Bell System Technical Journal* 27:379–423.

Thompson, C.; Göker, M.; and Langley, P. 2004. A personalized system for conversational recommendations. *Artificial Intelligence Research (JAIR)* 21:393–428.

Watson, I. 1997. *Applying Case-Based Reasoning: Techniques for Enterprise Systems*. San Mateo, CA: Morgan Kaufmann.