Supporting User Interaction in Flavor Sampling Trials

Edwin Costello, Lorraine McGinty
UCD School of Computer Science and Informatics,
University College Dublin, Belfield, Dublin 4, Ireland
edwin.costello@ucd.ie, lorraine.mcginty@ucd.ie

Abstract

Sensory innovation in the flavor industry is a process that relies heavily on user involvement to carry out critical sampling trials at a high cost. We are interested in how recommender systems, in particular, have an important role to play in view of addressing some of their key challenges and constraints. We describe a purpose-built sensory recommendation prototype with the potential to: (1) help users navigate a vast and complex solution space using very limited user preference information, (2) reduce the time it takes to gather critical information about individual user flavor preferences, and at the same time (3) allow for a broader coverage of the flavor space to be evaluated during inhouse flavor development trials, thus potentially reducing related overheads (e.g., time and money). In addition, we summarize some of the findings following a real-user evaluation of the prototype recently carried out by a leading Flavor industrial partner.

1 Introduction

The prevalent user endorsement of online search and discovery, coupled with the ever-growing expansion of web-based information/product catalogs, has provided a valuable proving ground for recommendation technologies. Noted as a particularly effective solution to the *information overload* problem, recommender technology has been adopted by online business providers keen to help customers find suitable gifts or movies (e.g., Amazon.com, Netflix .com), or keep track of news relevant to their interests (e.g., DailyMe.com). In this paper, we focus on the unusual domain of flavor science and sensory innovation. We discuss how user interaction and feedback is critical to the flavor development process and show how techniques and ideas from recommender technology can be used to enhance the current best practice in the area.

The flavor development and innovation task is subject to a number of challenges. Our work seeks to introduce a layer of intelligence, in the form of a recommender system, which will support the process of gathering user preferences and provide avenues for further enhancement. In particular we aim to improve the form of interaction that users take part in when providing this important feedback.

Our approach is discussed in Section 2 following a summary of the relevant methodology and specialist practice in this area. In Section 3 we describe the operation of our sensory recommender prototype, before summarizing in Section 4 some of our findings following a recent real-user evaluation of the prototype conducted onsite by our specialist industrial partner in the USA. By way of conclusion, in Section 5, we discuss next steps and position our research with reference to related work in the general areas of recommender systems and flavor science.

2 Flavor Science & Innovation

The flavor industry is a complex, far-reaching, highly specialist, and profitable business[1]. While we may not be consciously aware, it is a business that services us all as we conduct our day-to-day lives. Much of the food and drink we consume is likely to contain some flavor or fragrance creation (from soft drinks to salad dressing, ice-cream, beer, etc.). All of these pass through many stages of rigorous testing and validation before they take up residence on the supermarket shelf. We conduct our daily lives largely unaware of the amount of specialist consultation, time, energy, and money that preceded the production of many of the products we buy and certainly the food we eat. The remainder of this section introduces our industrial partner and provides a general overview of their flavor innovation process and the key challenges that are the focus of this work.

2.1 Introducing Our Flavor Innovation Partner

Our partner, Givaudan SA, is the world leader in sensory innovation within the flavors industry[2]. Sensory innovation involving flavors is a very knowledge-intensive business requiring experienced *flavorists*[3], information on customer requirements, user preferences, flavor compounds and ingredients, all of which must be combined effectively to produce a desirable flavor . To this end our industry collaborator, Givaudan SA, has led the way in recognizing, developing and applying

[1] Givaudan SA., Annual report 2007

[2] Givaudan SA was the first company to establish itself as a creator of tastes

[3] A team of highly experienced scientific experts in the discipline.

a range of highly innovative technology solutions to the flavor development process. The basic cyclic process consists 3 key stages:

1. *Preparation*: Experienced flavorists prepare a variety of samples and seek feedback.
2. *Sampling*: The target market (i.e., a group of people randomly selected from the public) evaluate the samples.
3. *Revision*: Experienced flavorists in light of feedback received revise the set of samples and seek feedback.

The traditional approach to flavor sampling trials involves the recruitment of sets of "panelists" to taste each flavor and rate it along a desirability score range. This second stage of the flavor development cycle is by far the most costly in terms of both the amount of user interaction required and the total time taken to gather this critical feedback on user preferences. A major limitation here is the number of samples that can be tasted before being subject to *sensory fatigue*. Evaluating a large set of flavors therefore requires that panelists have to make several return visits which, in turn, extends the lead time to product development.

As a solution here our industry collaborators pioneered a revolutionary approach to enhance these user sensory sampling trials that eliminates the need to taste flavors. Studies that have shown that as much as 80% of flavor perception is contributed by our sense of smell (see Blake [2004]) and, consequently, their panelists *sniff* flavor aromas using a highly specialized hardware device known as the Virtual Aroma Synthesizer (VAS). By this methodology, participants can sample up to a maximum of 70 flavor aromas in a single sitting, representing a significant increase on the conventional approach. The process consists of the same three stages as before except that sensory feedback based on flavor aromas, instead of taste, is collected at stage 2 (see Figure 1).

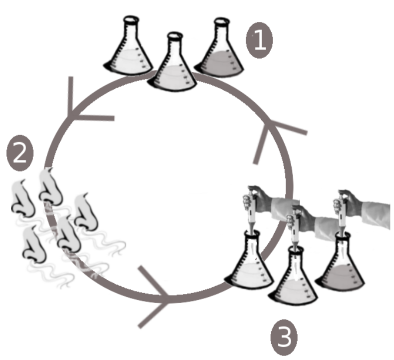

Figure 1: Aroma-based Flavor Development Cycle

2.2 Challenges and Constraints

Maintaining a high rate of innovation is essential to industry leadership and the long-term success of our partner's business. The company's customers, who operate in a very competitive environment, are constantly looking for new and innovative ideas to differentiate their products in the marketplace. It is their ability to develop a new flavor and then quickly transform it into a successful and viable market concept that makes Givaudan SA unique. Thus, it is essential that the processes that lead up to preparing a flavor for consumers is as efficient and effective as possible. The goal we are aiming towards is two-fold, firstly supporting a user interaction process for the flavor trials that enables the collection of preference information over as many alternative flavors as possible within the limited duration of each sensory sampling trial and secondly, to improve upon the quality of this feedback relative to what is currently available.

User Interaction

In the existing process, a key concern relates to the interpretation of ratings-based feedback. That is to say that the numerical ratings (e.g 1-9) gathered during trials, where low ratings imply a dislike and high ratings imply a liking for a flavor, can be subject to noise. Users do not always provide accurate ratings relating to their preference. In fact, a preliminary analysis of ratings collected in this domain (see Costello *et al.* [2007]) shows that users are very inconsistent when providing their subjective preference information. For example, in that particular study only 33.6% of individual panelists consistently gave a specific flavor the same rating).

In short, while our partner was keen to use recommendation to facilitate a customer-driven flavor development process, the only truly reliable feedback that would be available to the recommender would be simple preference information assigned on the bases of comparison (i.e., given a a number of samples one is preferred over the others). Given that user feedback plays such an integral role in flavor development and decision-making, Givaudan SA were interested to see if it were possible to gather richer feedback about individual user preferences and to enable the coverage of even more flavor options. This is the point where the introduction of recommender systems technology can play an integral role in enhancing and guiding a user's interaction with the flavor space.

A significant addition in the flavor trial user interaction process would be the ability to adapt the presentation order of flavor samples in response to each individuals feedback, as opposed to being random (i.e., not all users will be presented with exactly the same samples from the available options as in previous trials). A critical challenge here is how to build a recommender engine capable of fitting into the user trial process while supporting the user by providing adaptive navigation of these complex solution spaces using only minimal preference information.

3 Our Sensory Recommendation Prototype

In this section we describe the architecture and operation of our sensory recommendation prototype built to integrate with our partners current processes. Their cyclic methodology illustrated by Figure 1 aligns well with the general operation of most *conversational* recommenders (i.e., a 3 stage cyclic process of presentation, feedback elicitation, and revision of user requirements (e.g., see Smyth and McGinty [2003]).

3.1 Architecture Overview

Figure 2 below illustrates the basic underlying architecture which consists of 4 key components:

User Interface: Users interface with the VAS device mentioned in Section 2.1 and provide feedback through a simple client interface.

Session Manager: Responsible for the realtime logging of all system events, user actions and deliberation times.

Data Access: Responsible for reading and representing input (i.e., flavor) data from CSV or XML formatted files.

Recommender Engine: This is the brain-center of our demonstrator. We implement a comparison approach to flavor sampling using preference-based feedback and provides further details in Section 3.2.

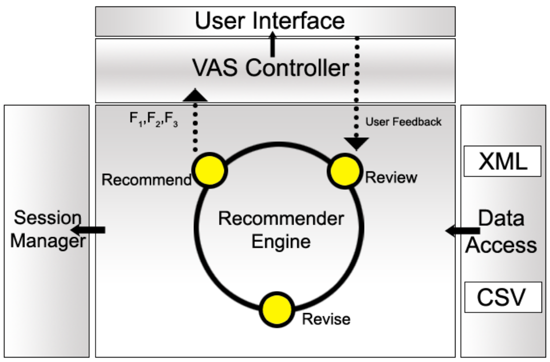

Figure 2: Basic underlying architecture.

3.2 Core Recommendation Algorithm

A pseudocode version of our 3 step recommendation algorithm is given in Figure 3. At the **recommend** stage the system presents a set of 3 flavor aromas to the user. We adapt a version of the *Adaptive Selection* strategy initially proposed by Smyth and McGinty [2003]. Importantly, the retrieval strategy governing the selection of recommendations is capable of altering its retrieval mode (as will be discussed later) on the basis of using only a user's preference-based feedback. At the **review** stage user feedback is gathered for recommendations. Each user is asked to indicate a preference from amongst 3 aroma options. Next, when in the **revise** stage the recommender takes a very simple approach to revising its understanding of where the user wants to go next by updating the query to be that of the most recently selected preference case.

User-Guided Retrieval

A key point that we borrowed from previous implementations of this algorithm is that the most recent preference indicated by the user is carried from each cycle to the next. This facilitates the altering of two retrieval modes: *refine* and *refocus*. If the user does not prefer one of the newly presented aromas in the next cycle (i.e., select a new preference) a more diverse set of recommendations, R, is retrieved in the subsequent cycle. That is to say, samples from very different parts of the flavor

```
define SensoryRecommendation(Q, CB, k)
Do
   R ← Recommend(Q, CB, k)
   c_p ← Review(R, CB)
   Q ← Revise(Q, c_p, R)
until (CB == 0) ∨ (stopCriteria() == true)

define Recommend(Q, CB, k, c_p, c_p-1)
if (c_p != null) && (c_p == c_p-1)
   R ← Refocus(Q, CB, k)
else
   R ← Refine(Q, CB, k)
return R

define Review(R, CB)
  c_p ← user's most preferred case from R
  CB ← CB - R
return c_p

define Revise(Q, c_p, R)
  R' ← R - {c_p}
  Q  ← c_p
return Q

define stopCriteria()
if(session stopping criteria has been met)
  return true
else
  return false

Q    = Query; CB = Casebase\Dataset
k    = Number of recommendations
R    = Recommendations; C_p = Preference case
C_p-1 = Previous Preference R' = Non-Preferred Cases
```

Figure 3: Our sensory recommendation algorithm.

space of alternates are presented (e.g., lime, strawberry, and orange) in order for the user to **refocus** their navigation. The *diversity* of these samples is calculated based on their perceptual similarity to eachother as shown in Equation 1.

$$Diversity(F_i, R) = 1 \; if \; R = \{\}; \qquad (1)$$
$$= \frac{\sum_{j=1}^{m}(1 - Sim(F_i, r_j))}{m} \; otherwise$$

Should a user indicate a new preference in a given cycle, subsequent retrievals are selected from the neighborhood of the most recent preference, and these are likely to be perceptually similar to each other (e.g., similar strawberry aromas). This allows the recommender to **refine** the search by allowing the user to make slight preference revisions based on local comparisons.

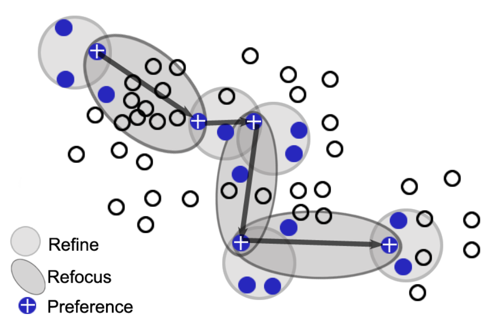

Figure 4: Illustrating the effect of adapting the retrieval mode using only preference-based feedback.

Adapting the retrieval strategy based on user preference-based feedback allows the user to sample a new part of the flavor space very quickly. Figure 4 shows a partial recommendation session where a user reaches their final flavor in only 8 recommendation cycles (using both refine and refocus modes). By this approach they do not need to sample all intermediate flavors between their starting point and the final chosen flavor.

Figure 5 shows an actual similarity profile for a user session from a recent trial discussed in Section 4. It shows a plot of the similarity of the preferred recommendation in each cycle to the ultimate preference. Note from cycle 11 through to 15 the user continues to prefer the same flavor before refocusing on a more desirable region of flavors.

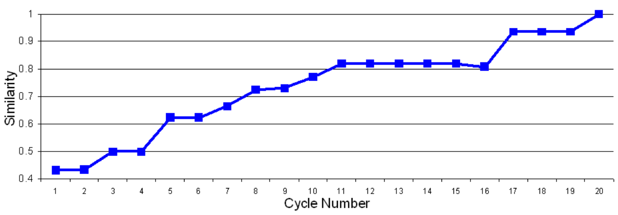

Figure 5: Sample similarity profile for an actual user session.

4 Real-User Evaluation Trial

At this point it makes sense to summarize the challenges our industry partners face in their daily business. First and foremost is the hugely complex and potentially infinite flavor space that they are seeking to explore. Couple this with the problem of trying to gather accurate and informative preference information from groups of panelists and it is clear that the process is difficult. Givaudan SA have already proved highly successful in surmounting these obstacles but they are still keen to innovate and evaluate new opportunities.

In previous work (Costello *et al.* [2007]) we demonstrated the efficiency benefits of taking a conversational recommendation approach in the presence of limited preference-feedback when conducting crucial sensory trials. Our partners were pleased to find that samplers could navigate through the flavor space towards specified target flavors up to 40% faster than the expected norm. While it is important for Givaudan SA to be able to identify convergence points quickly, in this evaluation they were more interested in evaluating the coverage characteristics of the sensory recommendation approach.

Conducting sensory trials in their business takes anything from hours to weeks. While they would like to reduce this time their primary goal is to facilitate a customer-driven feedback process whereby decisions are made on the basis of rich preference feedback that captures the *likes* and *dislikes* of target consumers. Thus, in this trial, they conducted a *blind* evaluation trial whereby we were not told in advance what *stopping criterion* they would impose. Further details of the methodology and observations that followed are discussed further by this section.

4.1 Dataset and Trial Methodology

The dataset used for this trial contains 98 fruit flavors, $F_1....F_{98}$, such that each flavor F_n is composed of at most 5 ingredients $\{I_i, ..., I_5\}$ from a possible 7. Each user was asked to participate in 3 sessions. The first cycle of each session contained 3 preselected flavors with each of the 3 sessions having a different 3 flavors presented. Importantly, these 3 starting sessions were common across all users, in other words all users started with the same starting recommendations.

In all, 26 users performed the evaluation but instead of having specific *target* navigation points, on this occasion, a severe stopping criterion was imposed by Givaudan SA. That is, users were asked to continue using our sensory recommender, indicating their individual path of preferred preferences through the flavor sample space, for no more than 15 minutes or until they exhausted all of the flavor space (note, no user exhausted the flavors). Importantly, for the traditional approach users can take up to 45 minutes to sample only 70 flavors, so 15 minutes is very restrictive for a dataset of 98. They were interested to see how users would navigate and how much of the dataset could be explored. A total of 78 sessions were recorded, equivalent to over 19 hours of actual user contact time with 1848 individual recommendation cycles recorded.

4.2 Trial Results

Table 1 below provides a summary of the important results gathered. As one might expect, users behave very differently when interacting with the system, some can complete many cycles and spend little time deliberating while others can take considerably longer. Importantly it is the collective feedback that is most important, and in particular, the goal of covering and gathering this feedback over a far greater piece of the flavor space than is currently possible within the restrictive timeframe of 15 minutes.

Table 1: Average, max and min statistics.

Description	Average	Max	Min
No. of Cycles / Session	24.18	46	7
Unique Flavors / Session	47.57	91	13
Cycle Duration	42s	307.36s	14.66s

Flavor Space Coverage

Here we discuss particulars relating to the evaluation of the coverage characteristics of our sensory recommendation prototype. A key point to highlight here is that the entire 98 samples were presented and sampled by the users, the minimum being seen 26 times and the most sampled flavor occurring 148 times. There was a concern that some flavors would not be sampled at all by any of the users, but this was not the case. The traditional aroma-based approach is currently limited to approximately 70 flavors but based on these results, using our recommendation approach, the possibility to expand the set of potential flavors exists allowing for more cover of the flavor space during a trial.

Does this dilute the quality of informative feedback gathered about what flavors samplers are drawn to? The answer here is no. It can be seen that, despite the fact that users

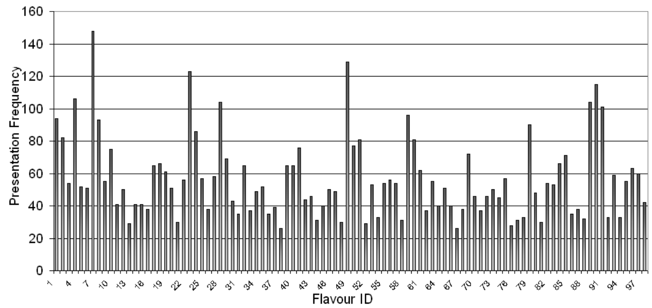

Figure 6: Number of times each flavor was presented.

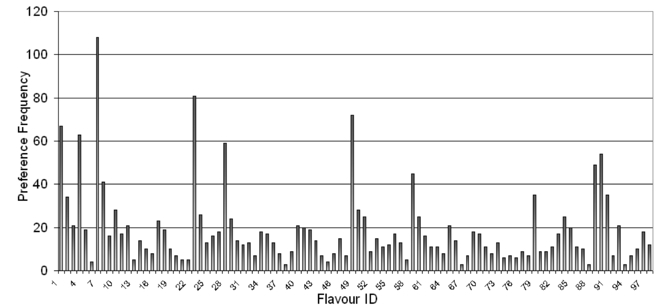

Figure 7: Number of times each flavor was preferred.

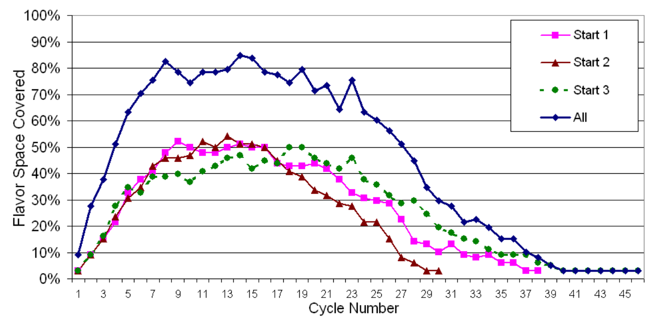

Figure 8: Percentage of the Flavor Space being covered during any given Cycle, across all users for each of the 3 session starting points and the combined coverage at each cycle.

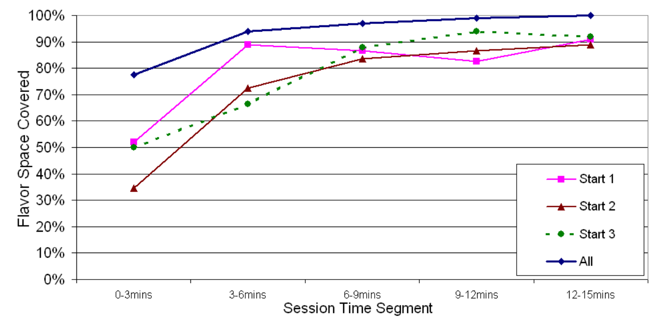

Figure 9: Percentage of the Flavor Space being covered during 3 minute segments,across all users for each of the 3 session starting points and the combined coverage for each segment.

were each taking very different *navigation paths* through the space of alternatives there were obvious flavors that stood out from amongst the others in terms of their *collective preference*. Figure 6 illustrates the number of times individual flavors were presented as a recommendation across all user sessions. Clearly all the flavors are being sampled multiple times, with the average frequency being about 57. Figure 7 illustrates the number of many times each flavor was chosen as a collective preferred case.

Evaluating coverage and having a more rigorous evaluation than simply looking at frequency is desirable therefore, data relating to how many unique flavors were present during a particular cycle number was gathered.

It is interesting to look at how the rate of coverage progresses in line with the (limited) feedback collected over the course of the recommendation sessions. Figure 8 above shows that in cycle 8 *alone*, over 80% of the flavors in the dataset were being sampled across all sessions. From this point on the rate of coverage increase is relatively steady and as Figure 6 shows, all the flavors are sampled *at least* 26 times. The other 3 data series, *Start 1, Start 2, Start 3* in Figure 8 show how the total coverage per cycle for each of the 3 starting points used approaches 50% by cycle 11. The drop in the coverage as cycles progress is due to the number of users that are still active at these cycle stages.

Figure 9 shows a time series analysis of our findings. This shows that the coverage is more regular and displays an upward trend resulting in overall coverage of 94% being achieved by the end of the sixth minute of the trials. This result is important as it provides a starting point for defining a more informed stopping criteria for user sessions. In fact, in terms of covering this particular flavor space, it suggests that 15 minutes may not be needed for this size of dataset. This could provide further cost reduction opportunities for these essential user-trials while allowing our partners to significantly expand the candidate set of flavors they can test beyond what is currently possible.

5 Related Work

While recommender systems have been adopted as an effective solution to the *information overload* problem, to the best of our knowledge their use in the flavor industry is nonexistent. Online business providers have been keen to improve the online experience, be it helping customers find suitable gifts or movies (e.g., Amazon.com, Netflix .com), or keep track of news relevant to their interests (e.g., DailyMe.com), for instance (see also, [Sarwar *et al.* [2000], Schafer *et al.* [2001]]). In this work, while we are still dealing with the similar notion of navigating through a potentially infinite options space we are dealing with a creative non-web based application of conversational recommender systems (see for other examples Shimazu [2001], Burke *et al.* [1997], Ricci [2002]).

While the flavor domain itself is extremely novel, it shares a number of characteristics and challenges that are very familiar to the recommender community. Flavors are described by a set of feature value pairs, i.e., their chemical ingredients and the quantities in which they are present. For most content-based recommendation applications, users can readily map their preferences onto the features and their values (e.g. *"I like this apartment because it has bedrooms"*) and this is readily used by many systems (see Shearin and Lieberman [2001]). When discussing flavors and their ingredients however there exists a vast *'vocabulary-gap'* (see McGinty and Smyth [2002]). For example, users are not going to be able to say *"I like this flavor because it has 2-Ethyl-5-Methylpyrazine"*.

This vocabulary gap is an important limiting factor that needs to be considered. Applications exist that attempt to use humans to aid and direct optimization systems. For example the job scheduling application (Lesh *et al.* [2003]) based on the Human-Guided Search (HuGS) framework (Anderson *et al.* [2000]). These applications do require domain knowledge on the part of the user and in the flavor domain such knowledge is not available and cannot be expected on behalf of the user. Therefore any feedback must take this into account, which is a key motivation for using preference-based feedback over sensory-based examples.

Mindful that the gathering of crucial sensory preference information must be performed such that it can bridge this gap without overwhelming the user, we have proposed using a novel content-based recommendation solution. Importantly, our solution uses minimal preference information and implements an interesting approach to similarity and retrieval that is highly sensitive to subjective preference. In addition it allows for the serendipitous discovery of desirable recommendations.

6 Conclusion

User interaction and preference feedback is central to the flavor development process. We have presented a sensory recommendation prototype capable of enhancing the currently used interaction process required to evaluate a target flavor space while at the same time, gathering high quality user preference information. With the introduction of this recommendation approach to the user trials, the amount of the flavor space that can be evaluated during sensory trials, which are routinely carried out by our industrial partner, has also increased. This is achieved by allowing users to flexibly navigate the flavor space and follow a path that allows them to make serendipitous flavor discoveries while providing limited preference-based feedback. The benefit to our industry partner is that they can now look towards conducting further consumer-driven sensory trials, on limited feedback over more extensive datasets, than is standard at the moment.

7 Acknowledgements

We would like to gratefully acknowledge the contribution of our collaboration partner Givaudan SA. Special thanks to Matt Burland, Christopher Simons, Rachel Mansencal and Jason Parcon for their insightful comments, advice, assistance and their role in facilitating the extensive user trials.

References

David Anderson, Emily Anderson, Neal Lesh, Joe Marks, Brian Mirtich, David Ratajczak, and Kathy Ryall. Human-guided simple search. pages 209–216, 2000.

Anthony A. Blake. Flavor perception and learning of food preferences. *Flavor Perception*, pages 172–202, 2004.

Robin D. Burke, Kristian J. Hammond, and Benjamin C. Young. The FindMe Approach to Assisted Browsing. *IEEE Expert: Intelligent Systems and Their Applications*, 12(4):32–40, 1997.

Edwin Costello, Lorraine McGinty, Matt Burland, and Barry Smyth. The role of recommendation for flavor innovation and discovery. In *Proceedings of the 2007 International Conference on Artificial Intelligence (ICAI-07)*, June 2007.

Neal Lesh, Leonardo B. Lopes, Joe Marks, Michael Mitzenmacher, and Guy T. Schafer. Human-guided search for job-shop scheduling. 2003.

Lorraine McGinty and Barry Smyth. Comparison-based recommendation. In S. Craw, editor, *Proceedings of the Sixth European Conference on Case-Based Reasoning (ECCBR-02)*, Aberdeen, Scotland, 2002. Springer.

Francesco Ricci. Travel recommender systems. *IEEE Intelligent Systems*, 17(6):55–57, 2002.

Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John Reidl. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM conference on Electronic commerce*, pages 158–167. ACM New York, NY, USA, 2000.

Ben Schafer, Joseph A. Konstan, and John Riedl. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5(1-2):115–153, 2001.

Sybil. Shearin and Henry. Lieberman. Intelligent profiling by example. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pages 145–151. ACM New York, NY, USA, 2001.

Hideo Shimazu. Expertclerk: Navigating shoppers buying process with the combination of asking and proposing. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 1443–1450, August 2001.

Barry Smyth and Lorraine McGinty. The power of suggestion. In G. Gottlob and T. Walsh, editors, *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, (IJCAI-03)*, pages 127–132, 2003.