

Voice-Rate: A Dialog System for Consumer Ratings

Geoffrey Zweig, Y.C. Ju, Patrick Nguyen, Dong Yu,
Ye-Yi Wang and Alex Acero

Speech Research Group

Microsoft Corp.

Redmond, WA 98052

{gzweig, yuncj, panguyen, dongyu, yeyi-
wang, alexac}@microsoft.com

Abstract

Voice-Rate is an automated dialog system which provides access to over one million ratings of products and businesses. By calling a toll-free number, consumers can access ratings for products, national businesses such as airlines, and local businesses such as restaurants. Voice-Rate also has a facility for recording and analyzing ratings that are given over the phone. The service has been primed with ratings taken from a variety of web sources, and we are augmenting these with user ratings. Voice-Rate can be accessed by dialing 1-877-456-DATA.

1 Overview

Voice-Rate is an automated dialog system designed to help consumers while they are shopping. The target user is a consumer who is considering making an impulse purchase and would like to get more information. He or she can take out a cell-phone, call Voice-Rate, and get rating information to help decide whether to buy the item. Here are three sample scenarios:

- Sally has gone to Home Depot to buy some paint to touch-up scratches on the wall at home. She'll use exactly the same color and brand as when she first painted the wall, so she knows what she wants. While at Home Depot, however, Sally sees some hand-held vacuum cleaners and decides it might be nice to have one. But, she is unsure whether which of the available

models is better: The "Black & Decker CHV1400 Cyclonic DustBuster," the "Shark SV736" or the "Eureka 71A." Sally calls Voice-Rate and gets the ratings and makes an informed purchase.

- John is on vacation with his family in Seattle. After going up in the Space Needle, they walk by "Abbondanza Pizzeria" and are considering lunch there. While it looks good, there are almost no diners inside, and John is suspicious. He calls Voice-Rate and discovers that in fact the restaurant is highly rated, and decides to go there.
- Returning from his vacation, John drops his rental car off at the airport. The rental company incorrectly asserts that he has scratched the car, and causes a big hassle, until they finally realize that they already charged the last customer for the same scratch. Unhappy with the surly service, John calls Voice-Rate and leaves a warning for others.

Currently, Voice-Rate can deliver ratings for over one million products, two hundred thousand restaurants in over sixteen hundred cities; and about three thousand national businesses.

2 Technical Challenges

To make Voice-Rate operational, it was necessary to solve the key challenges of name resolution and disambiguation. Users rarely make an exactly correct specification of a product or business, and it is necessary both to utilize a "fuzzy-match" for name lookup, and to deploy a carefully designed disambiguation strategy.

Voice-Rate solves the fuzzy-matching process by treating spoken queries as well as business and product names as documents, and then performing TF-IDF based lookup. For a review of name matching methods, see e.g. Cohen et al., 2003. In the ideal case, after a user asks for a particular product or business, the best-matching item as measured by TF-IDF would be the one intended by the user. In reality, of course, this is often not the case, and further dialog is necessary to determine the user's intent. For concreteness, we will illustrate the disambiguation process in the context of product identification.

When a user calls Voice-Rate and asks for a product review, the system solicits the user for the product name, does TF-IDF lookup, and presents the highest-scoring match for user confirmation. If the user does not accept the retrieved item, Voice-Rate initiates a disambiguation dialog.

Aside from inadequate product coverage, which cannot be fixed at runtime, there are two possible sources for error: automatic speech recognition (ASR) errors, and TF-IDF lookup errors. The disambiguation process begins by eliminating the first. To do this, it asks the user if his or her exact words were the recognized text, and if not to repeat the request. This loop iterates twice, and if the user's exact words still have not been identified, Voice-Rate apologizes and hangs up.

Once the user's exact words have been validated, Voice-Rate gets a positive identification on the product category. From the set of high-scoring TF-IDF items, a list of possible categories is compiled. For example, for "The Lord of the Rings The Two Towers," there are items in Video Games, DVDs, Music, VHS, Software, Books, Websites, and Toys and Games. These categories are read to the user, who is asked to select one. All the close-matching product names in the selected category are then read to the user, until one is selected or the list is exhausted.

3 Related Work

To our knowledge, Voice-Rate is the first large scale ratings dialog system. However, the technology behind it is closely related to previous dialog systems, especially directory assistance or "411"

systems (e.g. Kamm et al., 1994, Natarajan et al., 2002, Levin et al., 2005, Jan et al., 2003). A general discussion of name-matching techniques such as TF-IDF can be found in (Cohen et al., 2003, Bilenko et al., 2003).

The second area of related research has to do with web rating systems. Interesting work on extracting information from such ratings can be found in, e.g. (Linden et al., 2003, Hu et al., 2004, Gammon et al., 2005). Work has also been done using text-based input to determine relevant products (Chai et al., 2002). Our own work differs from this in that it focuses on *spoken* input, and in its *breadth* – covering both products and businesses.

References

- M. Bilenko, R. Mooney, W. W. Cohen, P. Ravikumar and S. Fienberg. 2003. Adaptive Name-Matching in Information Integration. *IEEE Intelligent Systems* 18(5): 16-23 (2003).
- J. Chai, V. Horvath, N. Nicolov, M. Stys, N. Kambhatla, W. Zadrozny and P. Melville. 2002. Natural Language Assistant- A Dialog System for Online Product Recommendation. *AI Magazine* (23), 2002
- W. W. Cohen, P. Ravikumar and S. E. Fienberg . 2003. A comparison of string distance metrics for name-matching tasks. *Proceedings of the IJCAI-2003 Workshop on Information*, 2003
- M. Gamon, A. Aue, S. Corston-Oliver and E. Ringger. 2005. Pulse: Mining Customer Opinions from Free Text. In *Lecture Notes in Computer Science. Vol. 3646. Springer Verlag. (IDA 2005)*., pages 121-132.
- M. Hu and B. Liu. 2004. Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD international conference*.
- E. E. Jan, B. Maison, L. Mangu and G. Zweig. 2003. Automatic construction of Unique Signatures and Confusable sets for Natural Language Directory Assistance Application. *Eurospeech 2003*
- C. A. Kamm, K. M. Yang, C. R. Shamieh and S. Singhal. 1994. Speech recognition issues for directory assistance applications. *Second IEEE Workshop on Interactive Voice Technology for Telecommunications Applications*.
- E. Levin and A. M. Manš. 2005. Voice User Interface Design for Automated Directory Assistance *Eurospeech 2005*.
- G. Linden, B. Smith and J. York. Amazon.com recommendations: item-to-item collaborative filtering. 2003. *Internet Computing, IEEE* , vol.7, no.1pp. 76- 80.
- P. Natarajan, R. Prasad, R. Schwartz and J. Makhoul. 2002. A Scalable Architecture for Directory Assistance Automation, *ICASSP 2002*, Orlando, Florida.