The Challenges of Moving from Web to Voice in Product Search

Amir Ingber, Arnon Lazerson, Liane Lewin-Eytan, Alexander Libov and Eliyahu Osherovich
Alexa Shopping, Amazon Research
{ingber, arnonl, lliane, alibov, oeli}@amazon.com

1 INTRODUCTION

In recent years, voice has emerged as a disruptive user interface and is becoming pervasive in our homes, cars, and phones. Tech companies are developing voice assistants and racing to be part of this revolution, with players such as Amazon Alexa, Apple Siri, Google Assistant, and Microsoft Cortana. One of the biggest challenges in this domain is searching and shopping by voice, which we refer to as "Voice Shopping". The voice shopping business is predicted to grow by at least an order of magnitude by 2020, from its $2 billion market today[1].

Intuitively, one would think that it is possible to directly transfer the rich amount of data and signals from Web e-commerce domain to voice shopping. Examples of such could be shopping behavioral signals used to train ranking models for product search. If feasible, this could enable a more efficient process of models development and validation. However, voice shopping, while directly related to Web e-commerce, introduces a new experience: customers buy different things using voice, they describe products differently, and behave differently along their shopping journey. The fact that in voice, both the input and output are spoken, results in customers being exposed to fewer results (typically, 1-2) with much less information. In contrast, a Web shopper is exposed in a single viewport to many offers and limited navigation options for further exploration. In voice, we thus have to optimize for relevance metrics such as success@1-2, which is extremely hard[2].

We discuss here the differences between voice and Web product search, demonstrating that transfer of signals and models from Web to voice cannot be performed directly.

2 THE GAP BETWEEN WEB AND VOICE

We first explore the difference between shopping categories in both domains. Our analysis is based on seven months of product search traffic in Web and voice in a major e-commerce site, considering users who were active in both domains. Figure 1 presents for each category the ratio of its popularity in voice vs. Web, where popularity is measured as the fraction of purchases per category within each domain[3] (the figure depicts a sample of 15 categories). Categories are anonymized, but we note that the categories with the highest voice/web ratio are mostly products that people buy on a regular basis such as supplies, groceries, beauty and health, etc.

Next, we show the different behavioral patterns in the two domains with respect to the numbers of user's actions between a query and a purchase. Voice assistants today allow the user to either perform a purchase action on the offer presented or not, and move to the next offer. In Web shopping, after being presented with a set of offers, the user has the ability to perform further navigation actions such as refinements. Figure 2 depicts the normalized number of actions per category in Web, where categories are ordered from lower to higher with respect to this number. We thus observe that unlike voice, in Web, a user will likely perform several navigation actions before a purchase.

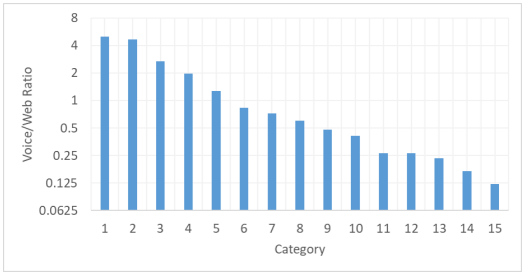

Figure 1: Voice to Web shopping ratio across categories

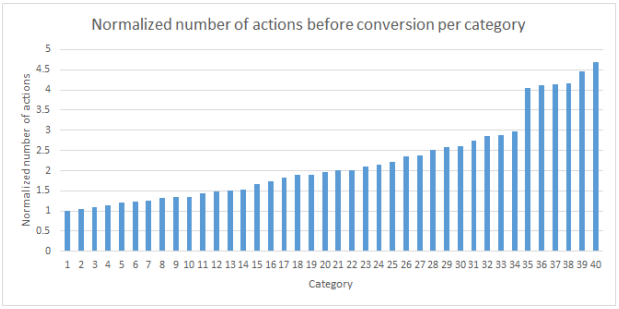

Figure 2: Normalized actions number across categories

Given the query, predicting the purchase category is important for offering the right type of products for the customer. We have observed, that using the same machine learning model, predicting the product category in voice (given the query) is 39% more accurate than in Web (measured in multiclass log-loss). This illustrates that in voice shopping, the user query is closer to the purchase than in Web, where the user can refine the search or browse through related products, until performing the purchase. For example, we may see in the Web experience a user issuing the query *"mickey mouse"* to end up buying a beauty product or going from searching for a lamp to end up buying a mobile app. This is much less common in the voice experience where customers must be more focused.

3 RESEARCH DIRECTIONS

In this workshop, we would like to make the community aware that Web methods cannot be directly applied to the voice domain. We list below some research directions that we would like to discuss with other workshop attendees:

- Characterize and contrast users' behavior in Web vs voice product search
- Revisit user search experience in the voice domain
- Explore transfer learning methods from Web to voice

We hope to attract additional researchers to this emerging field of voice search that is already reaching far beyond early adopters.

[1] PR Newswire, OC&C Strategy Consultants, Feb. 28 2018, https://www.prnewswire.com/news-releases/voice-shopping-set-to-jump-to-40-billion-by-2022-rising-from-2-billion-today-300605596.html

[2] Success@k measures the average number of relevant results among the top k results.

[3] Popularity is taken as a relative number with respect to the whole domain, thus the ratio is calculated between two relative numbers (e.g., if the *groceries* category comprises 10% of the purchases in voice and 5% of the purchases in Web, then the value appearing in the figure would be 2).