Skill recommendations must be provided when users need them most, without being obtrusive or distracting.

BY RYEN W. WHITE

Skill Discovery in Virtual Assistants

VIRTUAL ASSISTANTS LIKE Amazon Alexa, Microsoft Cortana, Google Assistant, and Apple Siri employ conversational experiences and language-understanding technologies to help users accomplish a range of tasks, from reminder creation to home automation. Voice is the primary means of engagement, and voice-activated assistants are growing in popularity; estimates as of June 2017 put the number of monthly active users of voice-based assistant devices in the U.S. at 36 million.[a] Many are "headless" devices that lack displays. Smart speakers (such as Amazon Echo and Google Home) are among the most popular devices in this category. Speakers are tethered to one location, but there are other settings where voice-activated assistants can be helpful, including automobiles (such as for suggesting

convenient locations to help with pending tasks[5]) and personal audio (such as for providing private notifications and suggestions[18]).

Virtual assistant capabilities are commonly called "skills." Skill functionality ranges from basic (such as timers, jokes, and reminders) to more advanced (such as music playback, calendar management, and home automation). Assistant skillsets include both first-party skills and third-party skills. First-party skills comprise the aforementioned basic skill functionality found in many assistants, as well as skills that leverage assistant providers' strengths in such areas as electronic commerce (Amazon Alexa), productivity (Microsoft Cortana), and search (Google Assistant). All major assistants also provide development kits that empower third-party developers to create their own skills for inclusion. Skills can be invoked independently, linked together within a single voice command to invoke a preprogrammed routine, or in a sequence of related skills arranged as required for complex task completion. Despite the significant value virtual assistants can offer, discovery of their capabilities remains a challenge.

Skill Discovery

Skill discovery is a challenge for two primary reasons: the "affordances," or capabilities, of virtual assistants are often unclear and the number of

>> key insights

- Virtual assistants in headless devices (such as smart speakers) do not fully convey their rapidly expanding capabilities, sometimes causing users to struggle to understand when and how to best utilize assistant skills.

- This article evaluates methods for recommending relevant virtual-assistant skills using the current task context, identifying complementary contributions from personal and contextual features.

- It also provides design recommendations for how to offer contextual skill recommendations in a timely and unobtrusive manner.

skills available in virtual assistants is increasing rapidly. It is not easy to communicate all that these assistants can do. Users may develop an expectation from prior assistant use or device marketing that assistants will perform certain tasks well. Discovering new skills, especially those that could help with the task at hand, is considerably more difficult.[b] The number and variety of skills available in virtual assistants is also accelerat-

ing rapidly, especially with the advent of third-party skill creation through tools (such as the Alexa Skills Kit[c] and the Cortana Skills Kit[d]). Amazon Alexa, the most established skills platform, had more than 26,000 skills available as of December 2017. Figure 1 reports the dramatic increase in the Alexa skillset over time. The pace of growth is such that users often struggle to keep track as new skills are released.

Despite the increase in the number of Alexa skills, it is not clear that the ones being added are actively being utilized. To help determine if they are, I ran an offline experiment. Although usage logs from Alexa were unavailable for this study, it was possible to

examine Alexa skill popularity on the http://www.bing.com/ Web search engine. Bing search logs show that over the 18 months between July 2016 and December 2017 inclusive (the maximum time horizon of the logs), the 100 most popular Alexa skills (0.4%) comprised two-thirds of the skill-related search clicks. It is unlikely the 99.6% of skills clicked one-third of the time have little or no utility, addressing only highly specific needs; moreover, there was no correlation between the explicit skill rating on the Alexa skill store and skill use (Pearson $r = -0.05$). A more likely explanation is that users need help to fully understand the capabilities of their virtual assistants and technology to support salient skill discovery is necessary to help them make the best use of these assistants.

---

b Although the focus is on skill discovery, it is also worth acknowledging there are other factors that can affect the use of virtual assistants, including reliance on far-field speech recognition in smart speakers that is typically less accurate than its near-field counterpart,[25] as well as lack of privacy with the broadcast audio in these devices and concern about the social acceptability of using voice-activated assistants in general.[3]

c https://developer.amazon.com/alexa-skills-kit
d https://developer.microsoft.com/en-us/Cortana

Skill Search

While assistants may support skill searching to locate skills of interest (such as through a dedicated skill like "SkillFinder" on Alexa), these searches may be frustrating and fruitless since smart speakers do not present sufficient clues about their capabilities; many people simply do not know what they can even ask. Although a virtual assistant can help order food from a local restaurant or reserve transportation, unless users are aware of such options, they are unlikely to invoke the skills except by accident or through trial and error. Unclear affordances have long been highlighted by the design community as a reason for inaccurate mental models and the sparse or incorrect use of technology.[14] To address this limitation, assistants support voice prompts (such as "things to try" or "skill of the day") or answer questions (such as "What are your new skills?"). Both are inefficient ways to access new capabilities that require user input and present results through an audio list that is difficult for a typical user to peruse. Moreover, skills discovered through such mechanisms are less likely to relate to users' current tasks since the active context is ignored.

When users do search for specific skills, they encounter a different experience from what they may be familiar with through Web search engines; search engines are designed to handle general-purpose queries and provide rich visual feedback (a list of search results) that can help people understand what worked well in their query and help them refine their search as needed. In contrast, virtual assistants have a fixed set of capabilities, and smart speakers provide users limited information about what worked well in their search. Failure messages (such as "Sorry, I cannot do this for you right now" or "I am sorry, I do not understand the question") are common but uninformative. If they cannot handle the request and have a display (if invoked through, say, a mobile device), some assistants may resort to presenting Web search results, creating a disjointed experience. Users may be affected by "functional fixedness,"[2] a cognitive bias whereby expectations about what the assistant can do limits the breadth of users' requests. Also, complex answers or multiple search results are difficult to convey through audio output alone. Virtual assistants may elect to delegate results presentation to a companion device (such as a smartphone) when they cannot present the results via audio, but such devices are not always available.

Figure 1. Growth in the number of available skills for Amazon Alexa from November 2015 to December 2017 inclusive; https://www.alexaskillstore.com/. Alexa is a good subject for this analysis given the maturity of Amazon's third-party skills platform.
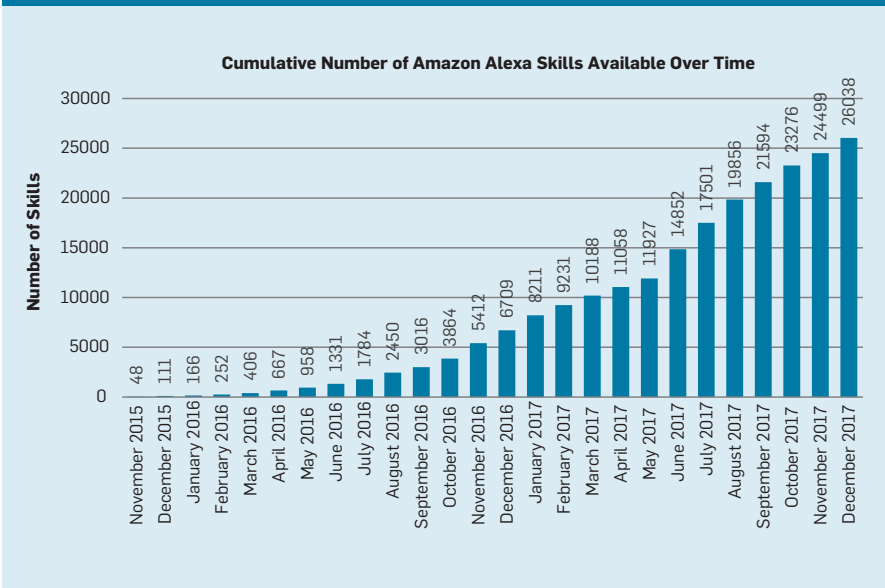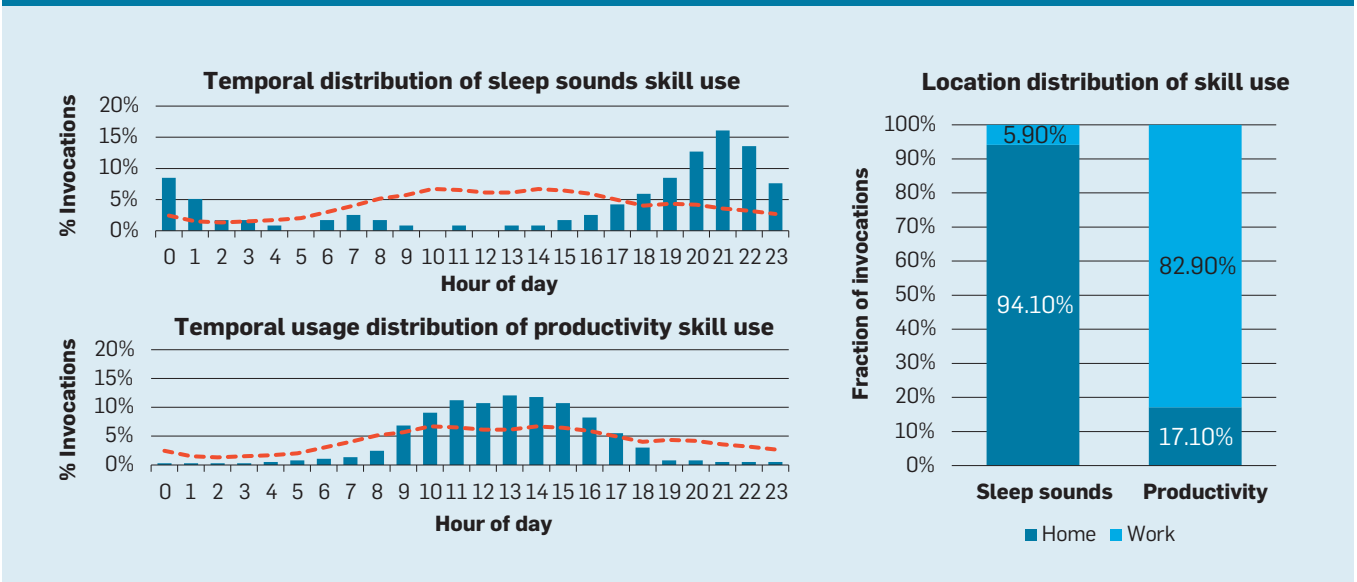

Figure 2. Plots of the time of day at which sleep sounds and productivity skills are invoked. The red dashed line denotes temporal distribution across all skills. Also shown are percentages of all invocations for each skill at two user-defined locations: home and work.

Design Challenges

Developers of virtual assistants face a two-fold challenge: how to set user expectations and educate users about what their assistants can do; and how to help these users while they are learning or even afterward as the assistant adds new skills. Existing onboarding methods provide written instructions in the retail packaging with examples of the types of requests assistants can support, along with periodic email messages that highlight new capabilities over time. These methods mostly promote only first-party skills, yet the power of virtual assistants (and much of the challenge with skill discovery) resides in the silent emergence of tens of thousands (and ultimately many more) third-party skills. Although users can develop effective mental models of products through prolonged use,[20] ever-expanding assistant capabilities make it difficult.

From a consumer-learning perspective, knowledge of virtual-assistant capabilities falls into the domain of "declarative knowledge."[22] Instruction manuals packaged with these products may outline sample functionality, but following the instructions too closely can hinder exploration of product capabilities.[8] Periodic (weekly) email messages may help reveal new skills and reinforce existing skills, but they are shown on a different device from the one(s) used for skill invocation and a different time from when they are needed. The in-situ recommendation of skills (based on the current task) could be an effective way to help ensure users of virtual assistants discover available skills and receive help when it is most needed and welcome.

Role of Context

Context is an important determinant of skill utility. The capabilities people want to employ differ based on contextual factors (such as time and location). For example, consider two skills, one offering ambient relaxing sounds to help promote sleep and one focusing on work productivity. Analyzing usage logs of these two skills from their internal deployment in Cortana with Microsoft employee volunteers reveals notable differences by time of day and location—home and work—both user-specified.

Although a virtual assistant can help order food from a local restaurant or reserve transportation, unless users are aware of such options, they are unlikely to invoke the skills except by accident or through trial and error.

Figure 2 reports that, as expected, the sleep-sounds skill is more likely to be used in the evening and at night than during the day, and much more likely to be used at home than at work. Use of the productivity skill exhibits a wholly different time-and-location profile. Though just one example, it shows that even for an immobile device like a smart speaker, there are still important contextual factors that should influence skill-recommendation priors. The use of context is even more pertinent in mobile scenarios where the context is more dynamic and user tasks are more context-dependent.

Just-in-time information access has been studied extensively.[16] To recommend the right skills at the "right time," or when they are most useful for the current task, it is understood that developers of skills software need rich models of user context. Fortunately, virtual assistants already employ myriad sensors to collect and model context, including physical location, calendar, interests, preferences, search and browse activity, and application activity, all gathered with explicit user consent.

Skill Recommendation

Models for contextual skill recommendation can leverage a range of signals available to the virtual assistant to recommend relevant skills based on the current context. Recommender systems have been studied extensively,[1] and developers of virtual assistants can draw on lessons from that community to assist in the recommendation of skills to users. Salient skills could be suggested in response to an explicit request for assistance from users (or a "cry for help" in more time-critical scenarios,[12] where the need for assistance is more urgent) or be based on external events. For example, a virtual assistant running on a smart speaker deployed in a meeting room can use the commencement of the meeting as a trigger to suggest ways to help make the meeting more productive (such as by taking notes or identifying action items).

Given a set of rich contextual signals, I have been exploring the use of machine-learned skill-recommendation algorithms (in this case, multiple additive regression trees[4]) to recommend skills that are useful in the current context. The models are trained

using historic skill usage data. This experimental setup resembles click prediction in search and advertising[17] but reflects several differences, including prediction target (skill used vs. search result or advertisement clicked), setting (open-ended assistant engagement vs. search-engine result page examination), and context (richer and more varied contextual signals available for skill usage prediction).

This research uses records with five months of skill invocations from an internal deployment of a smart speaker powered by Cortana with Microsoft employee volunteers (the same dataset as in Figure 2) and data on the context in which those skills were used as collected by Cortana. The data is split temporally and the first 16 weeks are used for training and the last two weeks for testing. Training and test data is stratified by user. The core principal in this specific instantiation of contextual skill recommendation is that if this or a

similar context is observed again, then the skills used previously in that context—by the current user, one or more cohorts, or the population of users—are more likely to be relevant and hence used again. Since the use of historical data puts the focus largely on predicting already-used skills, the study also investigated prediction performance for the subset of test cases where Cortana first observed users trying a skill.

The study further examined the effect of three classes of features used in the learned-skill-recommendation model: popularity, or general popularity of a skill across all users (using only historic usage data from before the skill was invoked); context, or rich contextual signals describing when the skill is used; and personalization, or features corresponding to the user who invoked the skill (such as the popularity of the specific skill for that specific user). These features resemble some that are commonly used in search and

recommendation,[1,12] although there are differences (such as lack of an explicit query and desire to focus on suggestion utility) rather than relevance or "interestingness" as the primary measure of model effectiveness. The table here provides more detail on the feature classes.

Figure 3 reports the receiver operating characteristic curves and precision-recall curves for the skill-usage prediction task across all skill instances in the test data. The feature contribution analysis starts with popularity features (area under the receiver operating characteristic, or ROC, curve or AUROC is 0.651), adds context features (AUROC increases to 0.786), and then adds personal features (AUROC increases further to 0.918), as in Figure 3a. All three models outperform a baseline of always predicting skill utilization, which reflects 19% precision at 100% recall, as traced by the dotted line in Figure 3b. The model that uses only historic skill-usage frequency performs worst. The results also show that algorithm performance improves considerably, given contextual features (yielding gains in precision) and personal features (yielding gains in recall), as in Figure 3b.[e]
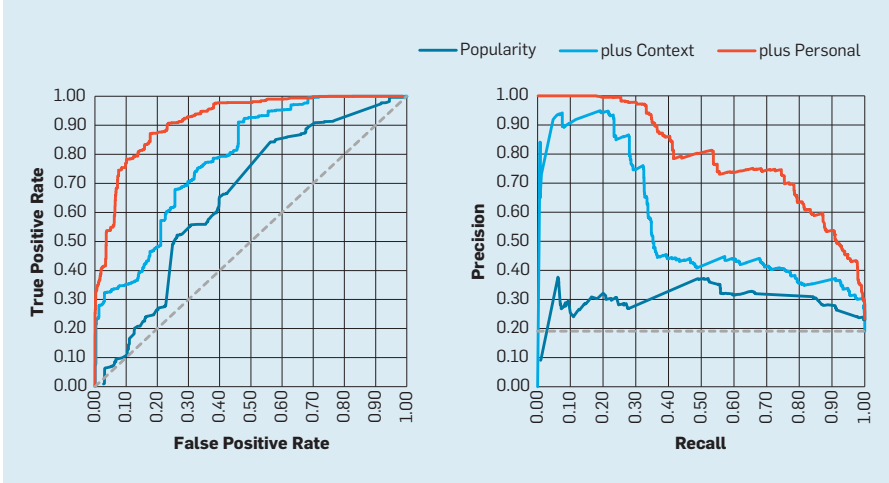
Inspecting the feature weights in the model containing all features (full model) reveals the features with the greatest discriminatory values are those associated with skill popularity (for both the current user and globally), calendar, and short-term interests, in this case, recent Web search queries. While the performance of the recommendation model is promising, reliance on historic data and the importance of such data in the model means there could be limitations on when it can be applied; for example, it could perform worse for new skills for which virtual-assistant developers have little data. To better understand the role of usage data, I reran the experiment, limiting test-data skill invocations to cases where the invoked skill was used by the user for the first time. The results resemble those reported earlier (AUROC = 0.894 for the full model), suggesting the approach may well generalize to unseen

Classes of features used in the skill-recommendation task. Text features with * are first represented in a continuous semantic space (300-dimension concept vector),[6] and the cosine similarities with both skill name and skill description are then computed. Each cosine measure (such as cosine similarity between the vectors for recent queries and for skill name) becomes a feature in the contextual skill-recommendation model.

Feature class	Example features
Popularity	Historic skill usage count, historic skill user count
Context	Calendar (meeting duration, subject*), visited venues (type, duration, name*), local time, to-do-list items*, search queries (in past 30 mins*, in past 30 days*), applications used*
Personal	Historic skill usage by current user within the past 30 mins, within the past 30 days, and all time

Figure 3. Performance curves for contextual skill recommendation. Results are shown for all skill instances in the test data for several feature classes: popularity, popularity plus context, and popularity plus context and personal (full model).

e   Similar trends are were observed when the order was reversed, to first add personal features and then to add contextual features.

skills. Regardless, this usage-based method is meant only as an illustrative example, and many extensions are possible. Complementary methods from recommender-systems research specifically tailored for cold-start scenarios (such as by Schein et al.[19]) may be helpful in tandem with the usage-based approach.

Although the focus in here is on contextual-skill recommendation, the results show that personal features contribute significantly to the quality of the recommendations generated. Personalization differs from contextualization because it is unique to the user, whereas contextualization could apply to all users in the same context, perhaps in the same meeting. More studies are needed on the use of personal signals, as well as how best to apply them to devices (such as smart speakers) that may be used in social settings (such as a meeting room or a family residence) where there could be simultaneous users of the virtual assistant, some known to the assistant and some unknown. More broadly, use of virtual assistants in social situations raises corporate product-development policy questions around whose virtual assistant should be employed at any given time in such multi-user settings. Speaker-identification technology[15] can help distinguish speakers in these settings to help decide what user profile or even what virtual assistant to apply. A centralized group assistant tied to collective activities (such as meetings[21]) can also help serve as a broker to coordinate tasks between individuals and their virtual assistants, even across assistant brands.

Using Recommendations
Despite the plentiful opportunities around developing more accurate contextual skill-recommendation algorithms, generation of skill recommendations is not the only challenge developers of virtual assistants face when working in this area. They need to also consider how to present the recommendations to users at the right time and in a manner that is not too obtrusive or distracting. The detection of trigger events and selection of the appropriate notification strategy are both particularly important.

As noted, the trigger can be user-initiated and intrinsic, as when a user says, "Hey Cortana, help me," or event-driven and extrinsic, as when a weather report warns of an impending severe weather event. Proactive scenarios on headless devices can cause frustration and distraction if a device reaches out with an audio message at an inopportune moment. While such intrusion could annoy users, it also has privacy implications tied to sharing potentially sensitive data with a wide audience, as when, say, accidentally notifying all meeting attendees about an upcoming private appointment. Methods have been proposed to better understand the situation and choose a suitable notification strategy.[7] The need for intelligent notifications is not lost on designers of smart speakers; for example, Google Home and Amazon Echo both support subtler approaches for notifying users (such as illuminating an indicator light on the device as an alert regarding a pending notification). It is only when users notice the notification and engage with the device that the notification is provided. However, this delay reduces notification utility considerably.

Context and User Consent
The performance of the contextual-skill-recommendation algorithms is strongly dependent on the signals that are accessible and the degree of consent users are willing to provide to get access to them. Contextual-skill recommendation focuses on suggesting skills to help people when they are in a context where those skills could be useful. Communicating clearly to users the connection between the provision of consent for data access and the provision of useful recommendations is likely to increase the chances users would be willing to grant data access for skill-recommendation purposes.[10]

As virtual assistants begin to manifest in other applications and devices, the range of contextual signals available to them will expand; for example, Facebook's artificial intelligence assistant, M, indeed chimes in during instant-messaging conversations to suggest relevant content and capabilities.[f] Virtual assistants can leverage signals about the conversation (such as topics discussed and people involved) for contextual modeling. Running skill recommendation within conversations highlights an interesting social dimension to the task of skill recommendations, where the skills suggested could vary based on who is spoken to and that person's relationship with the speaker, in addition to the topic of the conversation. Skills are also not used in isolation; many scenarios involve skills that are interconnected within a task, as in the restaurant-plus-transportation scenario mentioned earlier. Skill-invocation logs can be mined by virtual assistants for evidence of the co-utilization of these skills within a single task (similar to how guided tours and trails can be mined from historical user-activity data[23]) to help generate relevant skill recommendations. Such recommendations can then be presented proactively, immediately following the use of a related skill.

New Horizons
Virtual assistants have traditionally served users independently. In the past few years, assistant providers have started to partner to leverage their complementary strengths (such as the collaboration announced in August 2017 between Amazon and Microsoft on their Alexa and Cortana personal assistants). Although the focus of this article is generally on assistants recommending their own capabilities, opportunities are emerging to recommend skills among multiple assistants in ways where users could have several assistants, each helping them with one or more aspects of their lives. For example, Cortana might aim to capitalize on Microsoft's many productivity assets to excel in the personal-productivity domain. In the partnership between Microsoft and Amazon, Alexa could recommend Cortana for productivity-related tasks, and Cortana could recommend Alexa for more consumer-related scenarios, especially in e-commerce. Such partnerships allow developers of virtual assistants to focus more of their resources on strengthening their differentiating capabilities and less on keeping pace with competitors in other areas. Beyond strategic partnerships between well-known major corporate assistant providers, the virtual-assistant-using public is also

---

f   https://www.theverge.com/2017/4/6/ 15200836/facebook-messenger-m-suggestions-ai-assisant

likely to see increased collaborations among multiple skill developers to create compelling new skills and skill combinations. Such partnerships can capitalize on complementary technologies, shared domain knowledge, and other assets (such as data and human capital) that can unlock significant skill differentiation and utility for users. Services that offer skill federation across multiple assistants—much like search engine recommenders (such as Switcheroo,[24] which directs searchers to the optimal search engine for their current query)—will also emerge for virtual assistants, guiding users to the assistant best able to handle the current task or their tasks in general. Interoperability among multiple assistants might also yield considerable user benefit; for example, assistants could share contextual signals to offer skill recommendations and other services of greater utility than any individual virtual assistant alone.

Helping people understand how their assistants can help them is an important step in driving their uptake at scale. This is especially important in smart speakers and similar devices, where capabilities are not immediately obvious given limited display capacity. Looking ahead, I offer the following eight recommendations for virtual assistant developers:

Be proactive. The effectiveness of search (reactive) experiences for the skill-discovery task is influenced by users' expectations regarding affordances in virtual assistants. Proactive skill recommendation methods that understand the current context are a necessary complement to user-initiated skill discovery. Proactive methods may eventually supersede reactive methods as the primary means of engaging with assistant skills, contingent on the emergence of intelligent notification strategies. Virtual assistants could offer proactive support when certain criteria are met, including the availability of rich contextual signals, high confidence scores from recommendation algorithms, and low cost of interruption (such as when the user is assumed by the assistant to not be engaged in another task on the speaker or companion device, as in the recommendation on leveraging companion devices).

Skill-development kits should allow developers to specify for each skill during skill creation the context(s) in which the skill should be recommended.

Timing is everything. Surfacing salient skills means users can more fully leverage the range of support virtual assistants can provide. Presenting users with skill suggestions at the right moment (when they need them) means assistant capabilities are more likely to be remembered in the future.[9] Assistant providers could start by offering support for easily detectable events (such as the start of a scheduled meeting, receipt of a severe-weather alert, or following use of a related skill) and broaden trigger-event coverage thereafter based on task models built from contextual signals, users' contact preferences, and implicit and explicit feedback data.

Use contextual and personal signals. Skills are relevant in one or more contexts. The results of the study showed both contextual and personal signals are important in skill recommendation. A combination of the current context and long-term user activities and interests should be used for this task if that data is available. In addition, skill-development kits should allow developers to specify for each skill during skill creation the context(s) in which the skill should be recommended.

Examine additional signals. There is a range of contextual and personal signals virtual assistants do not have access to today (such as conversations in the room where a smart speaker is located, food being consumed, and television shows being watched) that could correlate with the invocation of skills and enable more targeted recommendations. Virtual-assistant developers should investigate what subset of these contexts is most likely to yield the best improvements in the accuracy of skill recommendations and explore the feasibility of collecting these signals at a large scale. They also need to engage with users to understand what new signals they are comfortable sharing with their assistant.

Consider privacy and utility. User privacy is paramount. If developers and their employers expect users to provide access to the contextual and personal signals required by skill-recommendation algorithms, they must clearly show signal value. Offering the right help at the right moment and attributing it to the permissioned data access via recommendation explanations could serve

to demonstrate the utility that can be derived from data sharing. Virtual assistants could offer explanations for each skill recommendation to help users understand how and why it was generated.

Permit multiple recommendations. The focus in this article is the task of predicting the single skill that users would be most likely to use in a given context. Regardless of the richness of any contextual model, the model is often incomplete and lacking in some information about the current task. Having only limited information could thus affect recommendation quality. When confidence in the recommendation model is below a threshold at which a definitive skill would typically be suggested, the assistant should recommend multiple (most-relevant) skills. This process accommodates less-relevant recommendations and meets other requirements of the recommendation task (such as showing the breadth of relevant skills available and supporting serendipitous skill discovery).

Leverage companion devices. Devices without displays may still have access to many screens through WiFi or Bluetooth connectivity, whether on a smartphone, tablet, or desktop PC. Signals from such devices that may not be available on smart speakers (such as recent smartphone apps used) would help enrich the context and assist in providing more-relevant skill suggestions. Given limitations in users' working memory,[13] evaluating result lists is considerably easier if a device has a display. If not, only the top few options can reasonably be vocalized by the assistant for consideration by users. Virtual assistants running on headless devices could use proximal devices with screens to better understand user tasks and display additional content to augment voice-only interaction.

Support continuous learning. Recommendations are needed when users are new to their virtual assistant. However, since skill volume grows silently and quickly over time (see Figure 1 for an example of such growth on Alexa), I foresee there will always be a requirement for assistants to offer suggestions to their users on how they can best help them with their current task. To help improve user understanding, virtual assistants could

occasionally suggest new skills based on their users' past skill usage. An appropriate format and time for such suggestions could be through an instructive tip at what may represent a teachable moment immediately following the use of a related skill. As mentioned, developing a notification strategy needs careful attention, given the need to balance the intrusiveness of alerting (especially audio alerting) vs. guiding users toward skills when they need them most.

Conclusion

Learning all that virtual assistants can do or relying on periodic skill-update email messages from their developers is insufficient for a user to make the most of such skills. Unlike apps, which are popular on smartphones and tablets, assistant skills are most likely to be invoked on headless devices that lack displays, increasing dependence on skill finding and limiting skill discovery. The limitations of browsing to discover new knowledge are well understood.[11] Even devices with screens, including Amazon's Echo Show, are limited in the number of recommendations they can present to users and would benefit from algorithms that leverage contextual and personal cues for skill recommendation. Looking ahead, the user-perceived utility of virtual assistants, especially as they manifest in smart speakers and other headless devices (such as personal audio), will depend largely on their ability to proactively identify and share skills that help their users at the moment they need that help the most. C

References
1. Adomavicius, G. and Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering 17*, 6 (June 2005), 734–749.
2. Duncker, K. and Lees, L.S. On problem solving. *Psychological Monographs 58*, 5 (1945), i.
3. Easwara Moorthy, A. and Vu, K.P.L. Privacy concerns for use of voice-activated personal assistant in the public space. *International Journal of Human-Computer Interaction 31*, 4 (2015), 307–335.
4. Friedman, J., Hastie, T., and Tibshirani, R. Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics 28*, 2 (2000), 337–407.
5. Horvitz, E. and Krumm, J. Some help on the way: Opportunistic routing under uncertainty. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, PA, Sept. 5–8). ACM Press, New York, 2012, 371–380.
6. Huang, P.S., He, X., Gao, J., Deng, L., Acero, A., and Heck, L. Learning deep-structured semantic models for Web search using clickthrough data. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management* (San Francisco, CA, Oct. 27–Nov. 1). ACM Press, New York, 2013, 2333–2338.
7. Horvitz, E. and Apacible, J. Learning and reasoning about interruption. In *Proceedings of the Fifth International Conference on Multimodal Interfaces* (Vancouver, BC, Canada, Nov. 5–7). ACM Press, New York, 2003, 20–27.
8. Lakshmanan, A. and Krishnan, H.S. The Aha! experience: Insight and discontinuous learning in product usage. *Journal of Marketing 75*, 6 (Nov. 2011), 105–123.
9. Kester, L., Kirschner, P.A., van Merriënboer, J.J., and Baumer, A. Just-in-time information presentation and the acquisition of complex cognitive skills. *Computers in Human Behavior 17*, 4 (July 2001), 373–391.
10. Krause, A. and Horvitz, E. A utility-theoretic approach to privacy in online services. *Journal of Artificial Intelligence Research 39* (Nov. 2010), 633–662.
11. Marchionini, G. and Shneiderman, B. Finding facts vs. browsing knowledge in hypertext systems. *Computer 21*, 1 (Jan. 1988), 70–80.
12. Mishra, N., White, R.W., Ieong, S., and Horvitz, E. Time-critical search. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Gold Coast, QLD, Australia, July 6–11). ACM Press, New York, 2014, 747–756.
13. Miyake, A. and Shah, P., Eds. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* Cambridge University Press, New York, 1999.
14. Norman, D.A. Affordance, conventions, and design. *Interactions 6*, 3 (May 1999), 38–43.
15. Reynolds, D.A., Quatieri, T.F., and Dunn, R.B. Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing 10*, 1-3 (Jan. 2000), 19–41.
16. Rhodes, B.J. and Maes, P. Just-in-time information retrieval agents. *IBM Systems Journal 39*, 3.4 (2000), 685–704.
17. Richardson, M., Dominowska, E., and Ragno, R. Predicting clicks: Estimating the click-through rate for new ads. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, AB, Canada, May 8–12). ACM Press, New York, 2007, 521–530.
18. Sawhney, N. and Schmandt, C. Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human interaction 7*, 3 (Sept. 2000), 353–383.
19. Schein, A.I., Popescul, A., Ungar, L.H., and Pennock, D.M. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Tampere, Finland, Aug. 11–15). ACM Press, New York, 2002, 253–260.
20. Schilling, M.A. A 'small-world' network model of cognitive insight. *Creativity Research Journal 17*, 2-3 (2005), 131–154.
21. Tur, G., Stolcke, A., Voss, L., Peters, S., Hakkani-Tur, D., Dowding, J., Favre, B., Fernández, R., Frampton, M., Frandsen, M., and Frederickson, C. The CALO meeting assistant system. *IEEE Transactions on Audio, Speech, and Language Processing 18*, 6 (Aug. 2010), 1601–1611.
22. Van Osselaer, Stijn M.J. and Janiszewski, C. Two ways of learning brand associations. *Journal of Consumer Research 28*, 2 (Sept. 2001), 202–223.
23. Wexelblat, A. and Maes, P. Footprints: History-rich tools for information foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada, Apr. 13–18). ACM Press, 1999, 270–277.
24. White, R.W., Richardson, M., Bilenko, M., and Heath, A.P. Enhancing Web search by promoting multiple search engine use. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, July 20–24). ACM Press, New York, 2008, 43–50.
25. Wölfel, M. and McDonough, J. *Distant Speech Recognition.* John Wiley & Sons, Inc., New York, 2009.

Ryen W. White (ryenw@microsoft.com) is a Partner Researcher and Research Manager at Microsoft Research AI, Redmond, WA, USA.