

INDUSTRIAL TECHNOLOGY ADVANCES

A mobile world made of functions

HAOHONG WANG

We are currently living in a world dominated by mobile apps and connected devices. State-of-the-art mobile phones and tablets use apps to organize knowledge and information, control devices, and/or complete transactions via local, web, and cloud services. However, users are challenged to select a suite of apps, from the millions available today, that is right for them. Apps are increasingly differentiated only by the user experience and a few specialized functions; therefore, many apps are needed in order to cover all of the services a specific user needs, and the user is often required to frequently switch between apps to achieve a specific goal. User experience is further limited by the inability of apps to effectively interoperate, since relevant user data are often wholly contained within the app. This limitation significantly undermines the continuous (function) flow across apps to achieve a desired goal. The result is a disjointed user experience requiring app switching and replicating data among apps. With these limitations in mind, it appears as if the current mobile experience is nearing its full potential but failing to leverage the full power of modern mobile devices. In this paper, we present a vision of the future where apps are no longer the dominant customer interaction in the mobile world. The alternative that we propose would “orchestrate” the mobile experience by using a “moment-first” model that would leverage machine learning and data mining to bridge a user’s needs across app boundaries, matching context, and knowledge of the user with ideal services and interaction models between the user and device. In this way, apps would be employed at a function level, while the overall user experience would be optimized, by liberating user data outside of the app container and intelligently orchestrating the user experience, to fulfill the needs of the moment. We introduce the concept of a functional entry-point and apply the simple label “FUNN” to it (which was named “FUNC” in (Wang, 2014)). We further discuss how a number of learning models could be utilized in building this relationship between the user, FUNN, and context to enable search, recommendations and presentation of FUNNs through a multi-modal human-machine interface that would better fulfill users’ needs. Two examples are showcased to demonstrate how this vision is being implemented in home entertainment and driving scenarios. In conclusion, we envision moving forward into a FUNN-based mobile world with a much more intelligent user experience model. This in turn would offer the opportunity for new relationships and business models between software developers, OS providers, and device manufacturers.

Keywords: Mobile applications, Machine learning, Data mining, Recommender systems, Context awareness

Received 29 November 2016; Revised 16 February 2017

I. INTRODUCTION

Apps have become the primary means of user interaction on mobile devices such as phones, tablets, smart watches, and other wearable technologies. A recent report by Flurry [1] shows that time spent on mobile devices grew 117% between 2014 and 2015, while overall app usage has increased on average 58% year-over-year. However, according to Nielson [2], time spent with mobile apps is concentrated in a limited number of apps (26.7 apps/month/user), a number that has remained unchanged over the past several years for US mobile users. Compared with the number of total apps in the world (more than 5 million as of June 2016, as reported by Statista [3]), this small percentage ($\sim 0.0005\%$) suggests that people are significantly underutilizing the

broad offering of app services available and hence the overall power of their mobile devices. One of several causes often cited for this limited app usage is the inability of the user to discover the unique power or capability of a given app. A related cause is the difficulty of even discovering that an app exists to meet a specific need. The user is like a fisherman sitting on an iceberg catching a couple of fish everyday from an ocean containing millions of fish (Fig. 1).

Even once apps are discovered and learned, leveraging their full capability presents another impediment. While some needs are small enough to be resolved by a single app, others require multiple apps, since each one typically presents a constrained and targeted view that fails to match the broad, complex and often changeable environment that the user may find himself/herself in at any given moment. Consider, for example, the number of apps needed to attend a meeting in a different city: calendaring, multi-modal communication, finding and reserving meeting space, booking an airline ticket, deciding on and booking hotel and dining

TCL Research America, 2870 Zanker Road, San Jose, CA 95134, USA

Corresponding author:

H. Wang

Email: haohong.wang@tcl.com

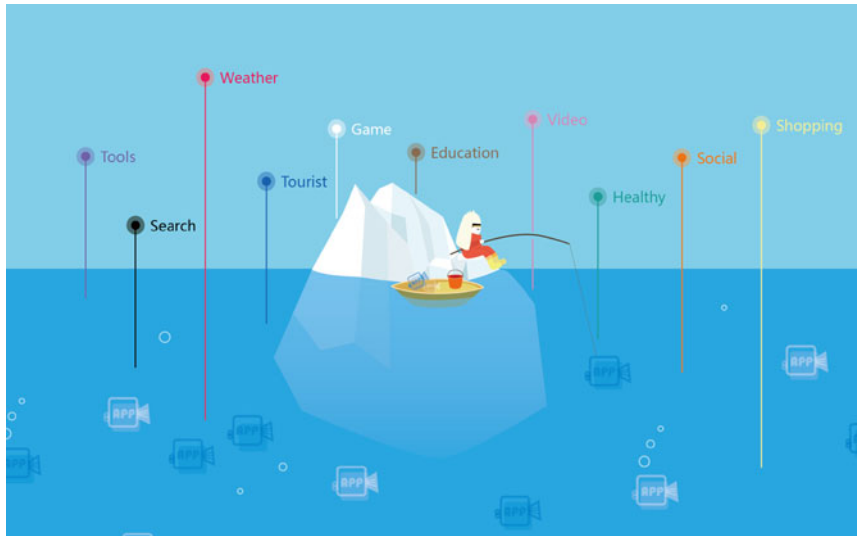


Fig. 1. An analogy to our current app-centric experiences on mobile devices.

reservations, finding a taxi to the airport of origin, arranging transportation upon arrival, and so on.

An intelligent personal assistant (IPA), or virtual assistant, is one way to help users experience the “intelligence” behind multiple apps. The best-known examples include Apple’s Siri, Google’s Google Assistant (previously Google Now), Amazon’s Alexa, Microsoft’s Cortana, and Facebook’s M. IPAs continue to evolve rapidly. For example, Google Assistant, announced in May 2016 at Google I/O, is an evolution of Google Now (originally unveiled in June 2012), a product utilizing Google Search with cards containing context-aware information obtained from relevant Google apps and system data. Google Now evolved into Google Now on Tap in May 2015 to enable users to take “action” through deep linking into a service contained within an app. Google Assistant adds two-way, voice-enabled dialog between the system and the user, leveraging Google’s Knowledge Graph to recognize repeated actions based on locations, calendar appointments, search queries, etc. to display more timely and relevant information for the device’s user. However, even Google Assistant fails to truly cross the boundaries of multiple apps to create a seamless experience. Instead, Google Assistant only optimizes the experience through more intelligent sequencing across discrete apps.

In this paper, we envision a new mobile world made of functions and propose a new term, FUNN (was named FUNC in [4]) to represent a subset of app “intelligence” that addresses particular user needs. This smaller form factor (as opposed to app) would be the foundation of a new mobile ecosystem. With similarities to mobile deep linking [5], FUNN represents the spectrum of access points to functions in mobile apps or connected services such as

- shortcuts to a page of a native mobile app (this is deep linking),
- entrances to web app functions, and
- customized functions on mobile devices based on cloud API services.

This concept may sound familiar to fans of Amazon’s Alexa, which uses voice interaction to meet a user’s simple need such as playing back music, making a to-do list, or accessing information such as weather, traffic, etc. In a FUNN world, we consider users’ needs not just at a simple task level, but within the broader context of what the user is trying to accomplish through the joining of a series of discrete tasks. This awareness of context and control over the orchestration of FUNNs further deploys a variety of multimodal interactions inherent to the mobile device as opposed to the limited voice interface employed by Alexa (as elegant as that may be). In other words, the experience of a mobile device can be tailored based on a user’s current context and intentions by employing interaction interface(s) accordingly.

Consider also that for similar needs, users may use different FUNNs inherent in different apps, which can be selected more precisely based on context. For example, a given user may use Facebook to share a picture with a particular friend group while using Instagram to accommodate a different group, according to their habits, preferences, subscriptions, and contextual objective. By better understanding a user’s context and intentions, the system is capable of selecting the appropriate FUNN to meet his or her needs. In this way, the user is relieved of having to remember specific app names or having to navigate within an app to a specific action page or having to manually switch among multiple apps (or even engage in new app discovery) to meet the demands of the current moment. All the user needs to do is to effectively express his/her current intention to the system via gesture, finger application, voice or other means and trust the system to provide an appropriately matched FUNN solution and contextually relevant user experience.

Let us consider a concrete example by imagining that you invite a friend for lunch, in the current app-centric mobile world, an app (e.g., Yelp or Foursquare) maybe used to find a desired restaurant with a convenient location for both of you, then another app (e.g., OpenTable) for table reservation

maybe used before you confirm the place to your friend with a messaging app (e.g., SMS or WhatsApp). You may need a navigation app (e.g., Google Map) during driving and then you may need a parking app (e.g., ParkMe) when you arrive. Clearly missing any app, due to user inability or unwillingness in app discovery, in this sequence may cause the whole experience broken. In a FUNN world, selected services are linked in automatically that matches the best for the current context. To be more specific, when you and your friend are scheduling the lunch via phone, the restaurant recommendation FUNN is pushed to you with suggested candidates, once a desired one is chosen, the table reservation FUNN is available for you to confirm the time and seats requirements, once confirmed, both you and your friend will receive a confirmation in your calendar, and the location will be automatically stored for navigation FUNN, which will be triggered after you get into your car. The parking reservation will be taken care of either before you start the journey or when you are on the way and close to the destiny. It is indeed the awareness of context and control over the orchestration of FUNNs that accomplishes the seamless mobile experiences by joining discrete tasks into natural sequences and appears in desired timing.

In this paper, we demonstrate how such a mobile world can be built, and we showcase several examples of experiences using such a system. To the best of our knowledge, this is the first systematic effort to propose a mobile future made of functions. The paper is organized as follows: Section II demonstrates the future ecosystem in a FUNN-based mobile world, Section III introduces moment-first user experiences that fit the FUNN-based world best, Section IV explains how machine learning and data mining would be applied to make the system perform, and Section V showcases two examples of the future of mobile experiences. Challenges and future directions are discussed in the last section.

II. FUNN-BASED MOBILE WORLD

In the manufacturing of mobile devices, software and applications are seamlessly integrated with the hardware components. To complete a shippable device, the following contributors are involved:

- Chip maker: supplies the chipset enabling computing and communication capability of the mobile device;
- OS provider: provides and maintains releases of mobile Operation System (OS), e.g. Android, Firefox, or iOS;
- Mobile device maker: develops and integrates the hardware and software components to form a completed device at both the physical and software levels, often porting and/or customizing certain OS features, integrating cloud services, and preinstalling selected in-house and third-party apps;
- App developer: takes charge of the user experience of a particular app product;
- Cloud service provider: provides intelligence or service via cloud, which can be a part of an app or a dedicated service for various clients.

Consumers today buy mobile devices from any one of multiple outlets: carriers such as AT&T or Verizon, open distribution channels such as BestBuy or Amazon.com, or device manufacturers online or in retail outlets such as an Apple store. Depending on which means they use, there are forces that influence the overall experience, and who should take full responsibility of the user's experience of a mobile device has long been debated. In the past 10 years, while OS providers have had very strong influence on the mobile-device user experience, a few extraordinarily successful apps have had an especially pronounced impact on users' daily mobile experience (consider Facebook or Instagram). Meanwhile, the mobile-device maker, who directly delivers the products and enables hardware components, can also make certain customizations to refine user experiences. Players who control more roles in the value chain – for example, Apple, which both provide the OS and makes the device – have had the advantage of being able to optimize hardware and software performance more fully than others. This is a trend that more companies are following: Google is bringing to market their own pixel phone, of which it is both OS provider and device maker, while Samsung has long distinguished itself in this fashion. Meanwhile, Apple, Google, and Samsung are all investing in IPAs to exert greater control over third-party apps and thereby optimize the user experience through more intelligent sequencing.

By contrast, in a FUNN-based mobile world, the key contributors can be narrowed down to two roles:

- FUNN provider: provides FUNN service and intelligence either on cloud or at the device;
- Mobile device maker: takes charge of user experiences by putting FUNNs together to handle user needs based on intelligently understanding context and the user.

Clearly FUNN provider role covers a huge landscape of players, which includes all the services and functions that human beings are consuming or to be consuming. Some FUNNs are functional tools to make our daily life more efficient, for example, photo-processing tools, speech recognition tools, and so on. Some are services in a specific field, as an example, Fig. 6 in Section A indicates a few entertainment FUNNs such as ABC live TV channel, or certain video title at Netflix. As a tangible example, Fig. 7 in Section B lists a number of FUNN providers for the car driving moment, such as Starbucks (coffee order service), Parkme (parking reservation service), Spotify (music service), Nuance (speech recognition and language understanding functions), Look4 (map and navigation service), and so on.

In this model, the mobile-device maker gains more control over the user experience by deciding which FUNNs are to be included and when they will be triggered in response to contextual conditions. Imagine a modern native app today as a house with only one entrance door (i.e.,

the launching page of the app); the FUNN concept would enable the house to have multiple doors or entry points. Orchestrating multiple FUNNs would optimize the user experience, demanding fewer interactions as opposed to requiring the user to invoke an app and then navigate to a specific app page and then switch and do the same with other apps. In this optimized fashion, the functional flow of users' actions can be built up much more smoothly, without the constraints of the app-centric model. That is, we use FUNN-level operations; instead, to cross multiple apps and/or employ additional intelligence via connected services.

Imagine that a human being's daily life consists of a list of needs and solutions. For every FUNN that a user requires, the system can often predict the next FUNN that the user may need and thus prepare the FUNN ahead of time. Transferring data between FUNNs are also supported, that is, some information used in the previous FUNN will be arranged and in place for the current FUNN and propagated to the next FUNN, obviating the need to input the same information again. The mobile-device maker can use a corresponding list of FUNNs and deliver an optimized experience with the least touch interactions to accomplish these needs.

This FUNN-based model presents a compelling vision for the future, especially as computation capability on mobile devices increases. Consider the further case where some FUNNs may be dependent on a certain OS and apply the situation where a user may rely on different FUNNs at different times. For example, a user may want to leverage Windows at work and use Android and iOS at home. An ideal mobile device would have more than one OS running on the same device through virtualization technology called "just needed device virtualizing," enabling mobile devices to support scene-adaptive experiences [6]. In other words, the system could be enabled – and intelligently orchestrated – to support multiple user environments running in parallel and having different OSs. In this fashion, mobile devices can make good use of FUNNs, while the OS is transparent to the user.

A user's daily life consists of a series of needs and digital solutions. Using a FUNN approach, mobile-device makers can better leverage an extensive list of FUNNs across apps, OSs, cloud, and the hardware/software stack to orchestrate a smooth presentation of solutions with fewer interactions. We elaborate on how the system makes good usage of FUNNs and how to use machine learning and data mining technology to realize the mobile world comprised FUNNs in the coming sections.

III. MOMENT-FIRST USER EXPERIENCES

In April 2015, Google published research in which they concluded that the mobile user experience consists of a series of "Micro Moments": intent-rich moments corresponding

to the statements I-want-to-know, I-want-to-go, I-want-to-do, and I-want-to-buy. The research suggests that it is in these moments that decisions are made and preferences are shaped [7]. In our daily lives, we have many such moments, although they occur across a range of contexts such as sitting in front of the TV, in the car, in a meeting, etc. The context of these moments can be perceived by mobile sensors and algorithms, which recognize these moments automatically or semi-automatically [6, 8]. Given the close relationship between contextualized moments and user intention, would not a moment-first user experience of mobile devices fit better into users' natural behavior model? In contrast to the app-centric mobile user experience, in which a user has to select (in some cases search for and discover) an app from a wide variety of options, FUNN-enabled solutions better cater to a moment-first user experience. In this way, users can directly access a FUNN that fulfills their immediate needs in the current moment in a contextually appropriate way.

The high-level architecture of moment-first mobile systems is illustrated in Fig. 2. A moment-recognition engine supported by mobile-sensing modules determines the moment or context "category" first and then triggers a certain user experience specifically designed for such a

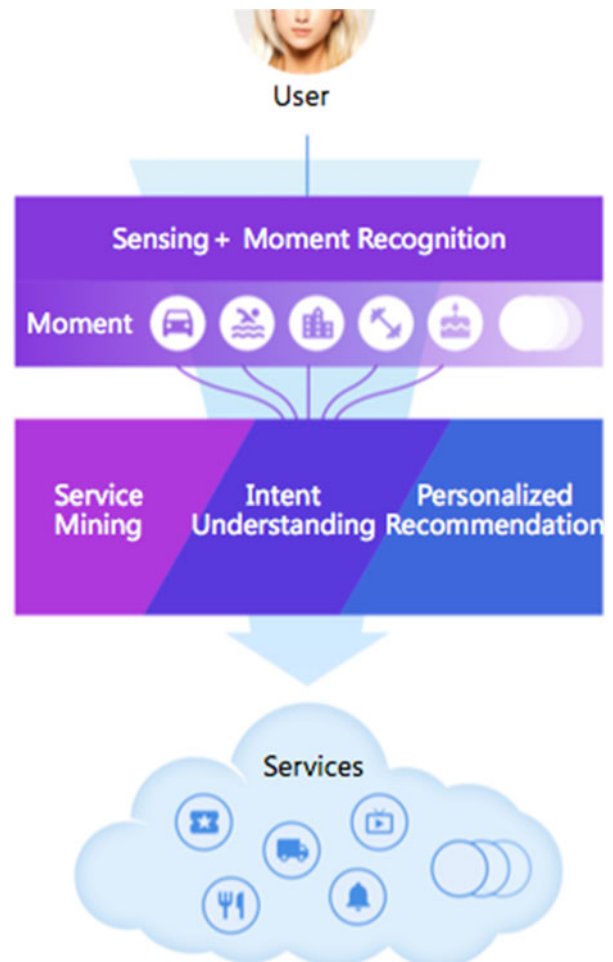


Fig. 2. High-level architecture of moment-first systems.

moment. The user intention, determined and shaped by the context of the current moment, allows for personalized FUNN recommendations to be triggered, linking the user directly to the available services that can fulfill his/her current need. The better the understanding of the user intention, the better the service recommendation and thus the better the user experience achieved. Where the average user-interaction count (number of touches or gestures on the device) is used as a measure of user-experience success on a mobile device, the lower the count the better the user experience [4]. In other words, if the user behavior and preferences have been learned and the system is able to predict the user's next immediate need and thus can prepare the appropriate FUNN to satisfy this need, then the user experience is optimized.

Although the term "Context Awareness" is not explicitly spelled out in Fig. 2, it is associated with almost every component of the system. The purpose of context awareness is to recognize a situation by using sensors and known attributes of a user's behavior to detect location, identity, activity, time, etc., and thereby trigger actions or experiences based on the context. For example, when the mobile device detects a workplace WiFi access point SSID, the processed sensing data may indicate the user is currently in a workplace whereas when the home WiFi access point SSID is detected they indicate the user is at home [6]. This approach was adopted by *Aviate* [9] (acquired by Yahoo! in 2014), which follows a moment-first user-experience model by showing the apps it thinks are useful depending on where the user is and the time of day. We will get to more details related to context awareness in Sections IV and V with more concrete examples and technology specifics. But this shift from the app world to a FUNN world is a major step toward achieving moment-first user experiences. By the same token, a moment-first user-experience model can only be realized in a mobile world comprised FUNNs.

IV. LEARNING AND UNDERSTANDING

As indicated in Fig. 2, machine learning and data mining can bridge a user's immediate needs and the ideal services to address those needs. In this section, we examine the relationship between FUNN, user, and context and explain how to enable a contextually relevant, moment-first user experience using the technical capabilities of machine learning, service mining, and intent understanding.

A) Service mining

Building a knowledge base by mining the available FUNNs (or services) is one of the most critical components of the envisioned system. Doing so relies heavily on the description document of each FUNN or the apps containing them. User reviews and user-generated content can also be good sources of knowledge of the FUNNs within an app and of aid in retrieving them from the App Store [10], and the same

work is easily extended to a FUNN-based system. Leveraging both FUNN descriptions and user reviews is challenging because these two types of unstructured texts are written by different authors from different perspectives. To exploit user reviews as well as app descriptions, they need to be combined. This can be accomplished using a novel topic model called *AppLDA* [10].

When a user writes a review, the user decides the topic descriptor – such as "installation problem" – whereas the topics in app descriptions are expected to be about FUNN features. *AppLDA* makes use of user reviews by comparing topics between the two different types of text and discarding parts of reviews that do not share topics with the app description. To accomplish this, two different topic clusters are built: shared topics and review-only topics. This allows us to define the relationship between users and FUNNs. Even though different users may use different FUNNs to achieve the same goal – for example, some people send messages via SMS, while others may use a messenger app such as Whatsapp – *AppLDA* can easily associate the respective meaning with the corresponding FUNN and thus deploy it appropriately in a specific context.

In addition, the correlation between sequentially used apps is likely to be a strong indicator of the relationship between apps [11]. A heterogamous network called R-Knowledge [12] can be systematically built to reflect the relationship between user and app in order to understand the roles and relationships of users and the connections among apps in the existing digital ecosystem. In the R-Knowledge framework, each user has a profile that includes a digital representation. For example, this profile can combine a user's social network identity with his/her activity stream app data (represented as text snippets). With a novel user-relationship topic model, the information about which topics and words the user typically writes is secured and sent to a classifier to understand the hidden relationships between the sender and receiver represented in the text information.

The relationships between FUNNs and the user can be similarly learned. If a FUNN is represented as a word and a sequence of FUNNs employed by the same user as a sentence, the relationship can be defined using a new-language-model architecture called Skip-gram [13]. For a text sentence, the Skip-gram model is able to generate high-quality vector representations of words from large amounts of unstructured text data, which can capture many hidden linguistic regularities and patterns. By applying the same model in this scenario, the relationship between a pair of FUNNs can be calculated as the distance between them after the data model is trained.

B) User intent understanding

Understanding the intention behind queries is not a trivial task. Traditionally, Web search engines take user queries through a single text input box from which the system must derive the search intent by analyzing a short list of keywords or name entities [14]. It has been reported that

nearly 70% of query logs contain single named entities (e.g., “Gone girl trailer”) [15]. These named entities might range across movies, music, books, autos, electronic products, etc. The training of named-entity-recognition models requires as input carefully labeled datasets, which are now easily obtainable. Intent can be further refined by leveraging such limited data types as location, person’s first and last name, address, product names, etc. Despite these refinements, users’ intentions can surpass the scope of such analysis since different users may look for different aspects of a named entity and it is difficult for the search engine to tell users’ exact search intent from such limited data.

A novel scalable framework [16] has been proposed that would learn from both the huge amounts of public query logs and an individual’s own query activity in order to learn a user’s intent more accurately. Even without a personal query history, a reasonable understanding of intent may be achieved by leveraging a model learned from the public at large, since the model also assumes, adapts and reflects the fact that user interests can be changeable. Using a large query log dataset as the input of the model, each sample of the dataset is represented as query log data containing the user query and the FUNN the user selected. When given a new query, the combination of learning from this dataset and the user’s query history, the framework will return several highly probable FUNNs that likely reflect the user’s intent.

Further, a Restricted Boltzmann Machine (RBM) can be employed [16] to learn the correlation between different user inputs and user behaviors. Given the need to process data of different types, for example, query text represented as word count vectors, or user interested URLs represented as binary vectors, this multimodal learning framework can be trained in an unsupervised manner over a pair of hidden layers, such as $h1, h2$ as shown in Fig. 3 that contains representation vectors, for each input modality. The output-hidden layer pair for each input modality is used as an input layer to train the higher output layer.

By learning from the entire query-log dataset, the model could learn the inner relation of query and FUNN for a majority of users. And the model could be used for FUNN recommendation.

Disambiguating these explicit queries is difficult enough. There are also implicit queries where an intention is not clearly expressed. For example, statements such as “I’m hungry,” “I’m so tired,” and “I have no one to talk to” do not

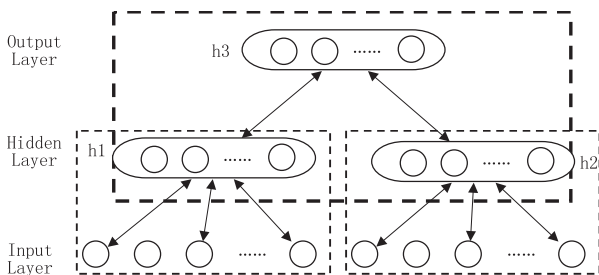


Fig. 3. Multimodal RBM.

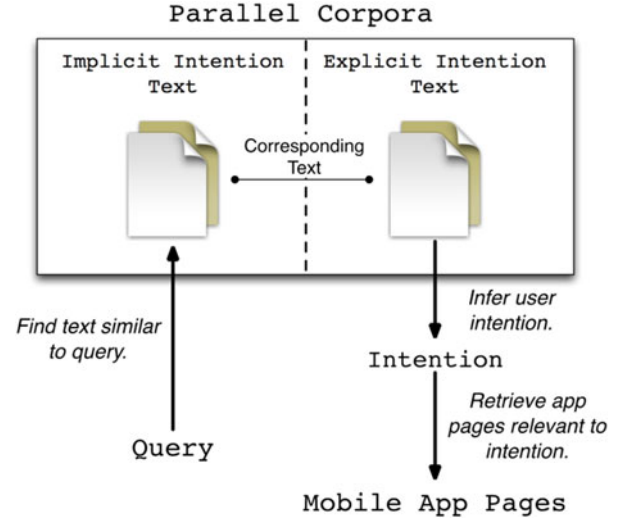


Fig. 4. Using parallel corpora to convert an implicit query to an explicit intention.

reveal an intention explicitly but certainly imply the need for something. We classify these expressions as implicit queries and might match these through intelligent mapping with explicit queries such as “food available,” “relaxing music,” and “dating app,” respectively. Implicit queries are typically converted to explicit intention [17].

One idea is to leverage social media to build parallel corpora that contain implicit intention text and corresponding explicit intention text. Specifically, we can model various intentions in social media text using topic models and then predict a user’s intention given a query that contains an implicit intention. As shown in Fig. 4, when a user implicitly expresses needs in a query, we search for similar text in explicit intention text in the parallel corpora using the Query Likelihood retrieval model. Then, an intention-topic-modeling approach would help us understand different intentions by removing noisy topics. Using intention inference via Intention LDA and an intention language model, the parallel corpora would help “translate” implicit query text into explicit user intention.

C) Context-aware personalized recommendations

In the area of recommender systems for mobile devices, much work is focused on using structured context information such as location, time, ambient light level, accelerometers, gyroscope, etc., to predict the launching of apps [18, 19]. Recent proposals [20] deliver recommendations based on FUNN (or app page). They also consider unstructured contextual data from app content pages or from user inputs for better recommendation performance. Specifically, the clues are represented by structured features captured by the mobile system, such as time, latitude, longitude, speed, etc., and analyzed alongside unstructured features such as bag-of-words representations from user inputs and FUNN pages. In order to combine clues from two separate feature channels, a multimodal Deep Boltzmann Machine

(DBM) [21] is applied to generate a joint representation of the multimodal data by training the historical user data. Then the outcome representations are used to build a logistic regression model and thereby enable prediction. When a FUNN recommendation is needed, the mobile clues representing the unstructured text data are used as a query for searching for relevant FUNNs in real-time, and then the offline trained logistic regression model just mentioned is used, first, to predict how likely a user is to launch a FUNN, and, thus, to re-rank the FUNNs based on prediction scores to complete the recommendation process.

When we consider more advanced context-awareness scenarios for mobile devices, Focal Point enhanced Conditional Random Field (FPCRF) [22] is another promising method since it enables the system to leverage both textual display information and the interaction between user and mobile device. For example, once the user intention is identified as “finding a restaurant,” associated entities and services could be preloaded and prepared for the user in advance. In this way, once a Point of Interest (POI) is recognized in the back-end system, a query can return entities such as “Kokkari Estiatorio” and “Gary Danko” (two local restaurant names) along with preprocessing navigation. Additional refinement could be based on how the user further interacts with the mobile device, such as using a pointer to hover over an entity or otherwise focusing on the mobile screen. Figure 5 represents how, based on the user interaction, the POI “Kokkari Estiatorio” is awarded the highest probability recommendation via FPCRF. Such a prediction can then be used not only in the entity recommendation process, but also to queue the next logical FUNNs.

Specifically, FPCRF is a conditional random field model for modeling and automatically inferring current interest through user context, user profile, and POI information. This output can complement the prediction as to whether a user would be interested in the POI appearing on a mobile screen. According to the predictions, the mobile

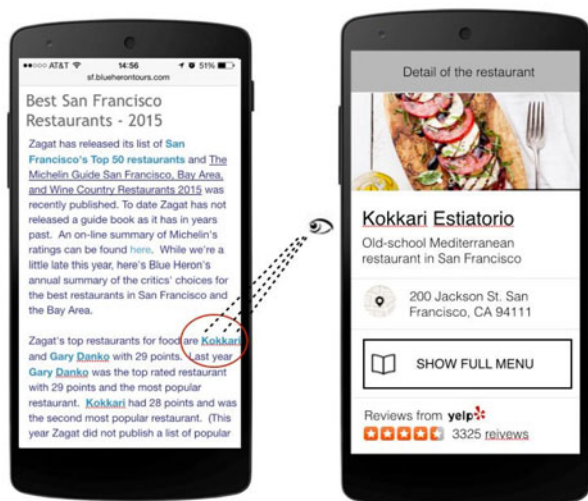


Fig. 5. Example of focal point awareness.

system could dynamically adjust the POI display strategy by predicting which POI corresponds to a user’s actual interests. The focal point on a mobile screen, contextualized by POI information, could provide information about the user’s actual interest. During the process, POIs are constantly detected from user text data and passed on to FPCRF with their coordinating information. When the user interacts with the mobile device via operations such as pointing, gesturing, grasping, shaking, tapping, or gazing, the focal point location on the screen is identified and sent to FPCRF as well. Then FPCRF re-ranks POIs based on the distance between POIs and focal points, device context, and the user’s profile.

V. CASE STUDIES

In this section, we showcase two analogies to the user experience in a FUNN-based mobile world: a future TV watching experience named CloserTVTM and an in-car driving experience. Through these examples, our goal is to demonstrate how FUNN-based experiences can be easily expanded to any number of moments in our lives.

A) CloserTVTM

For many people, TV watching moments are common in their daily lives. However, even with a single person the experience can vary widely depending on the broader contexts of time, available programming, location, etc. Of particular difficulty in today’s world is how to identify programming across both digital stream and a traditional cable set-top box. Figure 6 illustrates how using a FUNN-based system can deliver a more intelligent and informed experience for the user.

When a user sits in front of a TV, a his or her mobile device is able to discover TV and set-top-box devices via WiFi, Bluetooth pairing and other methods. Proximity and other clues can be used to determine that a TV watching moment is desired. In such a moment, information about programming and details regarding the content are needed to fulfill the most immediate user needs. The CloserTVTM experience is able to bridge the live-broadcasting world with the over-the-top (OTT) content world of streaming video choices (Fig. 6) to present the user with a single integrated experience. To do so, the system leverages the following resources.

- The metadata of both the live-broadcasting channels, for example, ABC, NBC, etc., and the OTT service providers, for example, Netflix, Hulu, YouTube, etc. These are obtained through a data-mining engine to enable user queries via keywords.
- Deep-linking URIs of every content title from the OTT services are accessible to the mobile device to quickly access corresponding title content.
- The mobile device assumes full control of the content to be displayed on the TV screen by controlling the TV and

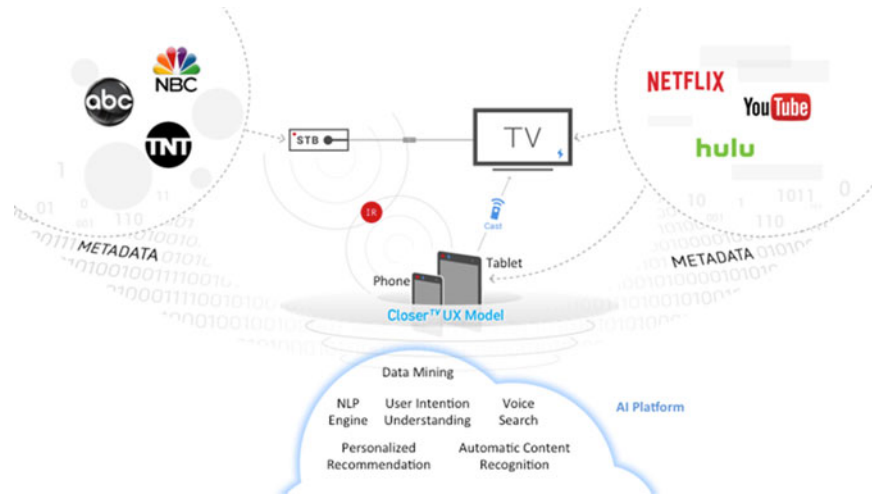


Fig. 6. *CloserTVTM* user experiences.

set-top-box with IR technology and by using Google Cast to display OTT content to the TV.

- The AI platform with a set of intelligent FUNNs enables understanding of the user’s immediate needs and recommending of personalized content and services to him/her.

When a user wants to interact with content programming, contextual information such as what content is currently playing and what is available; deep data resources about the actors, characters and other metadata; details regarding background music; information about brands or other objects that appear within the content; and so on become important data that can be acted upon. With *CloserTV^{TN}*, the user does not have to remember, which mobile app (or remote-control key sequence) he or she needs to use to access certain channels or content. This FUNN-based platform understands the user’s intention and provides the best prediction and display of most relevant services from which to select. The system is then able to quickly access the content either via set-top-box channel or via OTT service deep-links and bring the content to the TV screen.

Since user needs are frequently changing, the system is able to extend the viewing experience by leveraging rich data to power a related shopping moment – either when the programming is displaying a commercial or the user requests details about an object within the program itself. A FUNN can be utilized, such as a generic multi-cue product-detection framework that understands visible, topological, and spatial-temporal cues, to detect object class and select the best path that occurrences of the target product class can follow in the video [23, 24]. In this way, a pizza can be ordered or a garment purchased in response to, and in concert with, TV content.

It is worth emphasizing again that the learning and understanding capability in the AI platform can play a significant role in matching a user’s immediate needs with personalized services in real time. In this fashion, *CloserTVTM* successfully dissolves many boundaries, including those between the live-broadcasting world and the OTT world,

between OTT streaming apps, between user, TV, and set-top box, and between customer needs. Since the user experience is controlled by the system, additional resources can augment the largely visual experience using expanded techniques such as FPCRE, voice, and touch interactions.

B) Drive

In our second example, we discuss how a similar system can be used to power in-car moments. Consider a scenario wherein a user’s attention is initially focused on driving while being aided in navigation by a mobile device. Then a phone call comes in or the user needs to find parking or a gas station. In such a moment, the driver must interrupt the navigation aid and switch apps, a difficult task since it requires his or her eyes and hands, while in motion and thus significantly impairs his or her safety. In a FUNN-based mobile world, this conflict is more easily managed since tasks can be running in parallel and handled through a contextually-aware interaction model [25] able to leverage intelligent priority and timing attributes. Since the user experience is orchestrated by the system, a voice interface FUNN can enable hands-free and eyes-free interaction when the car is in motion and automate mundane tasks such as app switching. This experience would represent a significant improvement over today’s app-first user-experience model, which does not support voice interaction to cross multiple apps. By promoting the voice control FUNN from that of the app domain to that of the system-managed user experience, the appropriate user interaction method is selected to optimize the human-machine interaction, freeing both eyes and hands from the need to perform multiple tasks when on the move.

As indicated in Fig. 7, the *Drive* FUNN model brings all kinds of services to the on-road user experience by using context awareness and multi-tasking enabled by an AI platform. Context understanding can be aided by signal inputs from the mobile device such as Bluetooth and GPS and possibly even from the car via OBD-II interface. The system can use these data points to understand context cues

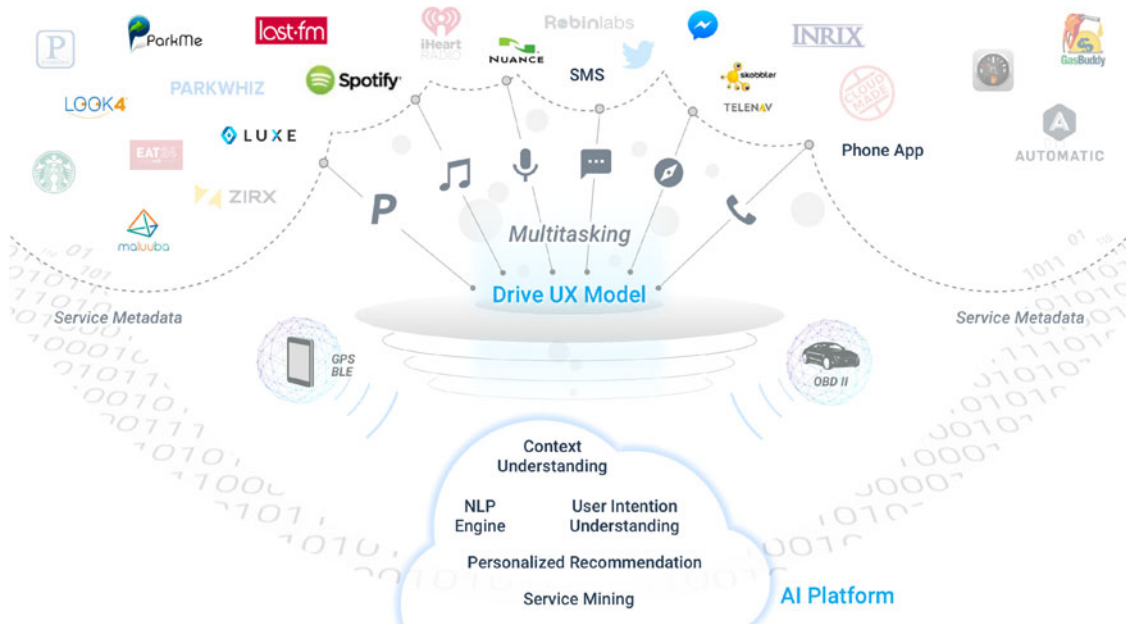


Fig. 7. Drive user experiences.

and enhance context awareness, improving user-intention understanding, service mining and personalized recommendations and NLP services. In this scenario, the mobile device can be more completely utilized by also analyzing data from the front-end camera to assist in detecting high-risk factors such as pedestrians and other vehicles and to handle traffic sign recognition by utilizing deep learning algorithms [26, 27].

VI. CONCLUSION

In this paper, we envision the future mobile experience powered by FUNNs, the function-level component units of apps and the OS. Recent efforts by Google, with their Google Assistant, and Amazon, with their Alexa-enabled devices (such as Echo) support our vision. One could equate each skill developed through the Alexa Skills Kit as a FUNN. When a complete set of FUNNs are available to serve all moments of our daily lives, we are able to move to moment-first user experiences to better serve user needs. In order to accomplish this, the main challenges rest with the development of algorithms to understand the context and user intentions that can lead to accurate personalized recommendations of the FUNNs needed to meet a user's needs and expectations. As Schadler et al. wrote in their book *The Mobile Mind Shift* [28], the ideal expectation is that for any moment, when someone pulls out a mobile device, he (or she) can get what he (or she) wants immediately in context. Specifically, wherever he (or she) is, the service is available; whenever he (or she) chooses, the service is at his (or her) fingertips; whatever his (or her) next step is, the system has anticipated his (or her) needs; whatever his (or her) action is, the system is ready to respond. With the booming of IoT devices and the growth of sensing capability, we foresee

a significant improvement in the context understanding capability and accuracy in near future, hence, we predict the future human mobile experiences will transform gradually from the current app-centric pull model to a moment-first push model, where context understanding plays a significant role in driving the dynamic user interfaces to push the user expected FUNNs to his (or her) fingertips.

It is also worth noting that a FUNN-based ecosystem is even better aligned with a world of ubiquitous IoT devices, many of which lack UI displays and are very restricted in the modality of user interactions. Given the rapid growth in computational capability and the heavy investment in AI technology, we predict increasing maturity of technology and services embrace a new mobile world made of functions.

ACKNOWLEDGEMENTS

The author would like to acknowledge Matt Berardo, Colleen Hamilton, Mengwen Liu, Yue Shang, and Freya Robles, for proofreading and helping refine this paper, and the team of *OneTouch* project of TCL Research America for achieving the products demonstrated in the paper. The author would like to especially thank Lifan Guo and Wei Gu for their great contributions during the initial effort of FUNN concept creation.

REFERENCES

- [1] Flurry Report: Media, Productivity & Emojis Give Mobile Another Stunning Growth Year, <http://flurrymobile.tumblr.com/post/136677391508/stateofmobile2015>, January 2016.
- [2] Nielsen Report: *SO MANY APPS, SO MUCH MORE TIME FOR ENTERTAINMENT*, <http://www.nielsen.com/us/en/insights/news/>

- [2015/so-many-apps-so-much-more-time-for-entertainment.html](#), June 2015.
- [3] Statista Report: *Number of apps available in leading app stores as of June 2016*, <https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/>, June 2016.
- [4] Wang, H.: Least Touch Mobile Device, US Patent Application Filed December 2014.
- [5] Mobile deep linking, http://en.wikipedia.org/wiki/Mobile_deep_linking
- [6] Tang, J.; Wang, H.: *System and Method for Mobile Platform Virtualization*, US Patent 9,063,770, June 2015.
- [7] Llewellyn, G.: “Micro-moments: What are they and how do marketers need to respond?,” <http://www.smartinsights.com/digital-marketing-platforms/google-marketing/google-micro-moments/>, July 2015.
- [8] Perera, C.; Zaslavsky, A.; Christen, P.; Georgakopoulos, D.: Context aware computing for the internet of things: a survey. *IEEE Commun. Surv. Tutorials*, 16 (1) (2014), 414–454.
- [9] Aviate, <http://aviate.yahoo.com>
- [10] Park, D.H.; Liu, M.; Zhai, C.; Wang, H.: Leveraging user reviews to improve accuracy for mobile app retrieval, in *Proc. ACM SIGIR 2015*, 2015, 533–542.
- [11] Huang, K.: Predicting mobile application usage using contextual information, in *Proc. 2012 ACM Conf. on Ubiquitous Computing*, 2012.
- [12] Guo, L.; Wang, H.: R-Knowledge: Brige Users and apps via Relationship Learning, US Patent Application Filed on December 2015.
- [13] Mikolov, T.; Chen, K.; Corrado, G.; Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781*, 2013.
- [14] Yi, X.; Allan, J.: A comparative study of utilizing topic models for information retrieval, in *European Conference on Information Retrieval*. Springer, Berlin, Heidelberg, April 2009, 29–41.
- [15] Xu, G.; Yang, S.-H.; Li, H.: Named entity mining from click-through data using weakly supervised latent dirichlet allocation, in *Proc. 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2009, 1365–1374.
- [16] Shang, Y. *et al.* Scalable user intent mining using a multimodal Restricted Boltzmann Machine, in *Proc. IEEE ICNC 2015*, 2015, 618–624.
- [17] Park, D.H.; Liu, M.: Method and System for app Page Recommendation via Inference of Implicit Intent in a User Query, US Patent Filed on June 2016.
- [18] Yan, T.; Chu, D.; Ganesan, D.; Kansal, A.; Liu, J.: Fast app launching for mobile devices using predictive user context, in *Proc. of the 10th Int. Conf. on Mobile Systems, Applications, and Services*, 2012, 113–126.
- [19] Natarajan, N.; Shin, D.; Dhillon, I.S.: Which app will you use next?: collaborative filtering with interactional context, in *Proc. 7th ACM Conf. on Recommender Systems*, 2013, 201–208.
- [20] Liu, M.; Shang, Y.; Guo, L.; Wang, H.: Multimodal Clue Based Personalized app Function Recommendation, US Patent Application Filed August 2015.
- [21] Srivastava, N.; Salakhutdinov, R.R.: Multimodal learning with deep Boltzmann machines, in *Advances in Neural Information Processing Systems*, 2012, 2222–2230.
- [22] Guo, L.; Li, G.; Wang, H.: Focal point based recommender System for Mobile Device, US Patent Application Filed August 2015.
- [23] Fleites, F.C.; Wang, H.; Chen, S.: Enhancing product detection with multicue optimization for TV shopping applications. *IEEE Trans. Emerging Top. Comput.* 3 (2) (2015), 161–171.
- [24] Fleites, F.C.; Wang, H.; Chen, S.: Enabling enriched TV shopping experience via computational and temporal aware view-centric multimedia abstraction. *IEEE Trans. Multimed.*, 17 (7) (2015), 1068–1080.
- [25] Meng, D.: Mobile Design for Multitasking with Priority and Layered Structure, US Patent Application Filed December 2016.
- [26] Ning, G.; Zhang, Z.; Huang, C.; He, Z.; Ren, X.; Wang, H.: Spatially Supervised Recurrent Convolutional Neural Networks for Visual Object Tracking. *arXiv preprint arXiv: 1607.05781*, 2016. <https://arxiv.org/pdf/1607.05781.pdf>
- [27] Ning, G.; Ren, X.; Wang, H.: Deep Learning based Road Situation Analysis, US Patent Application Filed March 2016.
- [28] Schadler, T.; Bernoff, J.; Ask, J.: The Mobile Mind Shift, Forrester Research, 2014.

Haohong Wang is the General Manager of TCL Research America at TCL Corporation. He leads the technology innovation and new business creation based on a vision called *OneTouch – The Right Service through a Single Touch*. He oversees the R&D activities in North America in the areas of mobile Internet, home entertainment, media processing and interaction, and AI-based services and applications. Before joining TCL, he held technical and management positions at AT&T, Catapult, Qualcomm, Marvell, and Cisco. Dr. Wang’s current research areas include multimedia computing and mobile systems and services. He has over 70 publications, five books (e.g., *3D Visual Communications*, *4G Wireless Video Communications*), and over 100 granted and pending patents. He has been the Editor-in-Chief of the *Journal of Communications*, the Steering Committee Chair of *IEEE ICME*, the Chair of *IEEE Multimedia Communications Technical Committee*, the Co-Chair of the *IEEE Technical Committee on Human Perception and Multimedia Computing*, the General Chair of *IEEE ICME 2011* and *ACM Multimedia 2017*, and the TPC Chair of *IEEE GLOBECOM 2010*. He received the *Distinguished Service Award* from IEEE MMTTC in 2013, the *Gold Innovation Award* from TCL in 2014, and the *Industrial Distinguished Leader Award* from the Asia-Pacific Signal and Information Processing Association (APSIPA) in 2016. He received his Ph.D. degree from Northwestern University, Evanston, IL, USA.