Shopping Intent Recognition and Location Prediction from Cyber-Physical Activities via Wi-Fi Logs

Manpreet Kaur
RMIT University[1], Tableau Software[2]
[1]Australia, [2]United States
manpreet882@gmail.com

Flora D. Salim*
RMIT University
Melbourne, Australia
flora.salim@rmit.edu.au

Yongli Ren
RMIT University
Melbourne, Australia
yongli.ren@rmit.edu.au

Jeffrey Chan
RMIT University
Melbourne, Australia
jeffrey.chan@rmit.edu.au

Martin Tomko
Melbourne University
Melbourne, Australia
tomkom@unimelb.edu.au

Mark Sanderson
RMIT University
Melbourne, Australia
mark.sanderson@rmit.edu.au

ABSTRACT

This paper investigates the Cyber-Physical behavior of a user in a large indoor shopping center by leveraging anonymized (opt in) Wi-Fi association and browsing logs recorded by the center operators. Our analysis shows that many users exhibit high correlation between their cyber activities and physical context. To find this correlation, we propose a mechanism to semantically label a physical space with rich categorical information from Wikipedia concepts and compute a contextual similarity that represents a customer's activities with the mall context. We further show the use of cyber-physical contextual similarity in two different applications: user behavior classification and future location prediction. The experimental results demonstrate that the users' contextual similarity significantly improves the accuracy of such applications.

CCS CONCEPTS

• **Information systems → World Wide Web**; **Web log analysis**; *Content match advertising*;

KEYWORDS

Wi-Fi logs analysis, intent recognition, shopping behaviour analysis

ACM Reference Format:

Manpreet Kaur, Flora D. Salim, Yongli Ren, Jeffrey Chan, Martin Tomko, and Mark Sanderson. 2018. Shopping Intent Recognition and Location Prediction from Cyber-Physical Activities via Wi-Fi Logs. In *BuildSys '18: Conference on Systems for Built Environments, November 7–8, 2018, Shenzen, China.* ACM, New York, NY, USA, Article 4, 10 pages. https://doi.org/10.1145/3276774.3276786

*Corresponding author

BuildSys '18, November 7–8, 2018, Shenzen, China

ACM ISBN 978-1-4503-5951-1/18/11…$15.00
https://doi.org/10.1145/3276774.3276786

1 INTRODUCTION

Knowledge about consumer behavior is critical for retailers to make personalised recommendations in targeted marketing, improving services or do location prediction. The operators of large indoor shopping malls wish to better understand consumer's behaviors to compete with the competition of online retail. Currently, physical retailers primarily gather customer insights by analyzing point-of-sale data. The path a customer took when visiting a mall, how much time they spent at a particular location, or whether they looked for a specific item is information that is typically not available. In contrast, online retailers benefit from rich information about customer activities, including knowledge of Web interaction such as pages visits and dwell times. Combined with sales information, such data provides actionable insights that can help retailers improve the online shopping experience of customers. The activities inferred from the data can be exploited to recognise users' intent during their online shopping. This, however, has not been previously explored in indoor retail environments.

Malls, museums, galleries, and transport hubs are large heterogeneous environments offering a range of different services: retail, entertainment, information, catering, etc. Increasingly, Wi-Fi networks and Bluetooth© beacons are being introduced into these spaces allowing the logging of visitor indoor movement and information behavior. Coupled with an understanding of the functions of the different locations in a space (i.e. *physical contexts*) one can ground and classify user behaviors or can predict future movements. Such applications would be an important step towards creating and delivering services to visitors.

The user behavior within a physical space is exhibited by heterogeneous data that represent different domains: the cyber domain, and the physical trajectory domain. Cyber domain captures user's interest in the form of queries issued whereas the physical trajectory (association with Wi-Fi access points) captures the information related to area of interest to the user. We hypothesize that users with *contextual intent* exhibit similarity between their physical contexts and their cyber behavior (users issue queries related to the context of the physical space), their cyber-physical behavior reflects what they are interested in.

To illustrate this form of interest with some examples: consider user A who intends to buy a laptop and compares products online while in the vicinity of a store selling computers; user B, who enters the mall searching for a particular store and follows a trajectory that ends in the vicinity of that store; and user *C*, who checks an online footwear size chart while in a store selling shoes. From these examples, it is clear that user *A* is interested in computers, user *B* is interested in a specific store and user *C* is interested in footwear. These examples demonstrate user interests that can be inferred from the physical context and the combined cyber and physical activity of users.

Hence, in this work, we present an approach to formulate a correlation between user physical and cyber behavior from heterogeneous data i.e. the Web Query Logs(Cyber) and the Wi-Fi access point association logs(Physical) in order to identify users' interests specific to the physical location. There are number of challenges involved in order to identify such interests. The first one being *Semantic Labeling* of a physical space. In a shopping center setup, this can be done by assigning the category of shops. For example, Cosmetics, Footwear, Clothing etc. within the range of access point. However, these categories are very broad in order to correlate the query context. Hence, the second challenge is *Semantic Category Expansion.* i.e. to expand the shop categories such that they cover the range of sub categories and products. Third, to find *Contextual Similarity*, open text query also needs to be represented in the form of categories (and the relevant sub-categories) to discover the semantic similarity between queries and physical space.

The cyber physical contextual similarity shows the users' interests across different semantic categories related to the physical environment and can be used in various applications that involves understanding of user behavior. In our work, we show the use of cyber-physical similarity features in two different applications, Classification of User Behavior and Future Location Prediction. We hypothesize that contextual similarity is a strong indicator of what user is interested in out of the services offered and can hence be helpful in, identifying whether user exhibit high contextual intent with the physical space or is just browsing through the area; and, which places user will visit next.

For behavior classification, the aim is to identify shoppers with high contextual intent during their visit to a mall from those who cannot be correlated with their local physical context by using features from physical trajectory, Web logs and features extracted from cyber-physical correlation that are strong indicators of users' contextual intent. The example of users A, B and C exhibits high contextual intent as given above, whereas there are many visitors for whom their Web behavior and indoor context are *contextually intentless.* In such cases their shopping intent cannot be easily inferred from online and physical activities. Consider user *D*, who visits the mall searching the Web for information about a particular festival occurring in the city, and interleaving these searches with queries about "lost luggage" and "baggage claim". This user is likely a tourist more focused on the free Wi-Fi than the primary services provided by the mall. While such users clearly have an intent, from the point of view of the mall operator, their visit can be classed as intentless. We also place in this category 'window shoppers', or shoppers with a high-level shopping intent (e.g., 'I need to get a present for my brother') that cannot be tied to a particular retailer or category of retailers. While all such users in this second category are potentially of great interest to indoor retailers, we focus in this part of work on detecting contextually intentful customers, utilizing both the cyber and physical behaviours as signals in characterising their intent.

Previous intent recognition work relied on examining either physical behavior from Wi-Fi signals, mobile phone sensors, mobile proximity sensors [13, 24, 26], or exploiting cyber behavior from online Web browsing and searching logs [15]. To the best of our knowledge, this is the first time user's contextual intent in an indoor space is inferred from both physical and cyber behavior.

We also employed user's cyber-physical semantic similarity for future location prediction. [16] studied the effect of different features on location prediction from Location Based Social Network data. They reported that category of location visited by user has high impact on prediction accuracy. However, the same is not studied for an indoor setup where movements of a user are captured by Wi-Fi traces. Therefore, we did an experiment study to see if user's future locations can be predicted much accurately by using semantics of indoor locations visited by the user and query context.

The main contribution of this work is:

- Semantic Categorization, used to semantically label a physical space and find the correlation between open text query with the physical semantics;
- Cyber-Physical Contextual Similarity model, used to extract contextual features, including Physical and Cyber activities captured by Wi-Fi AP association and Web Query logs;
- Shopping Intent Recognition System (SIRS) for User intent recognition, used to classify user's intent into two broad categories (intentful or intentless).
- Effect of Semantic Context on Future Location Prediction

2 BACKGROUND

Semantic Labeling of Contexts Context is an influential factor in analysing both human behaviors [17] and the user intent behind mobile information needs [3]. Semantic labelling of a location context is an important step to identify users' intent and has been researched well. Krumm and Dany Rouhana proposed Placer, which treats the semantic labelling as a classification problem by using the following features from government diary data: the time of visits, the demographics of the user and nearby businesses [11]. They found that the demographic information and nearby businesses are helpful in semantic labelling places, e.g. school, home, and work. Elhamshary et. al. proposed CheckInside, a fine-grained indoor location-based social network, which utilizes the check-in data collected from crowd of people to associate a location with its name and semantic fingerprint. They claimed CheckInside provides more accurate localization and better coverage [6].

Indoor Behavior Analysis: To support the real-world, mobile-centric behavioral research, Misra and Balan presented LiveLabs, which is a large-scale mobile testbed for in-situ experimentation [14]. They also investigated the user behaviors when considering whether they are within a group or are alone, and found people behave significantly differently within these two scenarios in terms of mobility patterns, app usage and propensity to communicate over phones [10].

Shopping Behavior Recognition: Understanding users' shopping behavior is practical, and some existing studies have been conducted. For example, Zeng et. al. [26] studied how to determine shopper's physical behaviors based on Channel State Information (CSI) of Wi-Fi signals. Specifically, they focused on shopper's behaviors near the shop entrances or within the store, namely standing/walking, and found CSI of Wi-Fi signals provide a good source to classify these different physical behaviors [26]. Radhakrishnan et. al. presented how to use smartphone and smartwatch to segment users shopping behaviors in a more fine-grained level, e.g. putting an item in the cart [18]. However, these methods are either device-based or wearable-based in controlled experiments, and the proposed system in this paper is infrastructure-based and the data were collected in the wild. Ren et. al. analyzed how people use Wi-Fi to access the Web in indoor retail spaces while navigating through the mall. They found patterns in shoppers' weekly visits, repetition within such visits, Web usage while browsing the mall, and the physical contextual influence on individual's cyber behaviour [22]. Based on these findings, Ren et. al. developed a tripartite location-query-browse graph (LQB) for contextual recommendations of location, query, and Web content, inferred from searching, querying, and moving behaviors [20]. Further, using only moving, searching, querying, and social behaviours inferred from Wi-Fi logs, Ren et. al. also find strong correlations between these cyber, physical, social behaviours and user demography (e.g. age, gender, income, parental status, visitor types) [21]. The work in this paper is going to build on the contributions from the previous works by Ren et al. [20–22].

User Intent Recognition: Jansen et. al. studied the user intent of Web queries, but focused on determining the informational, navigational and transactional intents [9]. [3] investigated users' intent based on diary studies by focusing on their mobile information needs, and suggested two more intent: geographical and personal information management (PIM). Chuang-Wen You et. al. proposed a phone-based system to monitor shopping time in stores. Specifically, they transformed this into a problem of classifying user trajectories as shopping or non-shopping, by utilizing the spatial and temporal features that are extracted from both Wi-Fi signals and sensors in the phone, including accelerometer and digital compass [25]. Duan and Zhai studied the intent representation problem in the field of entity search, e.g. product retrieval, and proposed a coordinated intent representation by linking the query space and the entity space collectively. But, they achieved this by mainly utilizing the query terms and product attributes [5]. However, little research reveal whether shopping intent is hidden in individual's movement (physical) and query (cyber) behaviors recorded in his/her Web log, which is investigated in this paper.

3 SYSTEM OVERVIEW AND DATASET CHARACTERISTICS

These days, indoor spaces provide access to opt-in Wi-Fi that captures user movements and what they look online. They are also rich in semantic information, such that areas in shopping centers can be labeled based on shop categories and products they sell. The semantics, when combined with knowledge from user movements and queries, can help in determining interests of the user or where they go next. But the challenge is to find correlation between user behavior with the context of the shopping center. Our work tries to solve this problem and proves the applicability of contextual information in two different scenarios, *user behavior classification* (i.e., shopping intent recognition) and *future location prediction*.

In this paper, we propose a method to do *semantic categorization* of physical location (Wi-Fi access points) using the Semantic Web. After categorization, we find semantic similarity between user query with the physical locations (access points). Two experiments were performed to see the effect of semantic similarity. First, we build a classification model to classify user trajectory into two broad categories, *intentful* and *intentless* and second we study the effect of semantic or contextual intent on future location prediction. In both experiments, it turned out that using context from user cyber-physical activities significantly improves results from baseline models or features. Next we explain the dataset used for our experiments and its characteristics.

3.1 Data Acquisition

We study an anonymized dataset of Internet access, that was captured by an opt-in, free Wi-Fi network in a large inner-city shopping mall in Sydney, Australia. The dataset collected from the Wi-Fi network includes two kinds of logs: a Wi-Fi *access-point association log* (AL) and a *web query Log* (QL), collected between September 2012 and October 2013. The shared, hallway spaces of the mall are spread over six levels and are covered by around 70 Wi-Fi Access Points (AP). The mall contains over 200 stores spanning 29 shop categories, which are defined by the mall operator, e.g., Bakeries & Cafe and Cosmetics & Costume Jewellery. The locations of the stores and the APs are documented in 2D floorplans of the mall.

The AL contains the following parameters: 1) the association access point ID; 2) start timestamp of the association; 3) association duration; 4) data volume received/sent in this association; 3) encrypted persistent user device ID. The QL contains the following parameters: 1) query issued by user; 2) association access point ID at which the query was issued; 3) encrypted persistent user device ID. All user identifiable information (registration details and Wi-Fi MAC addresses) were replaced by a hash key in a non-reversible way. Moreover, the queries issued by users are quite dissimilar in terms of topics. To group the queries into a higher level of categorization, wedeployed BrightCloud service (http://www.brightcloud.com/).We found that they are spread over 68 BrightCloud categories. In order to understand the intent of shoppers, the category we are interested in is shopping. However, *shopping* category consists of only around 8% of queries, hence it is a challenge to recognize peoples' shopping intent from query text.

3.2 Access Point Association

The user movements that are captured by the AP associations represent the areas in the mall that were visited. However, the associations will also capture the situation where the user is just passing by a particular location and did not visit any stores in that region. In order to distinguish between the two types of association, we generate a CDF (cumulative distribution function). The AL has a sampling rate of 5 minutes and from the CDF around 30% of the associations are found to be less then 10 minutes. Therefore, we considered a user's association with an AP only if the duration of association exceeds 10 minutes (two sampling intervals), assuming users may only be passing through an area when association duration is below 10 minutes to an AP.

3.3 Web Content-Access Point Correlation

For each visit of a visitor, we extracted a trajectory (sequence of visited APs). We then analyze logs by extracting the top user trajectory sessions, as explained in Section 3.1. Next, we construct a sequence of cyber-physical term sequences that relate the change in information needs with the change in physical context.

Thus, we hypothesize that an individual's intent could be constructed by linking their physical behavior (trajectory in terms of shop keywords) and cyber behavior. The challenge is to automatically link the user query with the physical context. User intentful open query text can contain terms that might not be captured from the crowd sourced web applications. Hence, we need a better solution and here we propose a *context categorization system* (CCS) that uses the Semantic Web, explained in Section 4.

4 CYBER-PHYSICAL SEMANTIC CATEGORIZATION AND CONTEXTUAL SIMILARITY

We first define what are the *Physical Context* and *Cyber Context*:

Definition 4.1. Physical Context: is defined as the area in the mall served by APs, and characterized by the Semantic Categories of Table 1. These are the semantic categories proposed in this study.

Definition 4.2. Cyber Context: is defined as a document of entities/categories, extracted from users' queries issued in a single visit to the shopping mall.

In order to identify user intentions, we need to find if their physical trajectories and cyber activities have some correlation with the physical surroundings. We have shopping center context in terms of shop categories. The main challenge here is to use open text queries (a small set of terms) and find the correlation to another small set of terms (representing category) that have minimal or no lexical similarity with each other. For example, a user query may be product name or a specific brand, e.g. *Mascara* and *Ugg Shoes*, which are not in the mall-defined shop category list. Hence, we propose a system that uses structured information from Wikipedia to find intent signals from user queries with respect to physical context, by extending both queries and the categories. For extending the queries and categories we gather additional information from DBPedia, that are richer in textual content than Wikipedia as it is a Semantic Web application that makes the content of Wikipedia available in the Resource Description Framework (RDF). The extended content representations are then compared. Our approach comprises of three steps: 1) Modeling Physical space using extended categorical information from Wikipedia; 2) Extending user queries by mapping it to Wikipedia concepts/pages and extracting categories related to each concept/page; 3) Capturing intent signals by correlating extended user queries with predefined Categories.

4.1 Preliminaries

Documents are collections of semantic categories. *Terms* (e.g. shoes, boots) are nouns extracted from search logs and browsing logs. *Entities* are known concepts or resources in Wikipedia, which could include specific brand names. *Semantic categorization* is the method to extract *semantic categories* (and the related sub-categories) to represent the physical and query space. In this paper, we enrich the semantics of users' cyber and physical context by entity discovery with the DBPedia knowledge base [1].

4.2 Physical Context

This section explains how we enrich the Semantic Categorization in the shopping context and assign categories to Wi-Fi APs based on the region covered by each AP.

4.2.1 Enriching Semantic Categories. The semantics of the physical space are defined as shop categories by the mall operator. However, this categorization is too broad and limited in terms of user shopping intent recognition. Specifically, this study aims to find the correlation between user query with the physical semantics in order to discover the intent of the user. User queries can contain a broad set of terms that can be related to these categories. It is feasible to correlate the user query terms with these categories only using a rich corpus of terms, categories and products that represents shopping center context.

To generate the corpus that contains a vast range of terms related to shopping context, we use structured information from Wikipedia. Information on Wikipedia is organized by categories and each category has further subcategories forming a tree like structures for the aid of navigation. We exploit this categorical information to enhance semantics and generate a rich corpus of categories. Our hypothesis is, user queries to some extent can be related to Wikipedia categories in order to get an understanding of query intent.

To start with, we first manually map each semantic category defined by the mall operator to a Wikipedia category and get 18 broad Wikipedia categories, as given in Table 1. We then input each category to our *content categorization system*, which iterates through sub-categories using depth-first search of up to λ levels and create a document of all the categories/subcategories iterated in the process. We tested category collection up to $\lambda = 7$ and realized that increase in the number of levels lead to increase in noise in the category documents. Due to this noise we found that there is no clear distinction between different semantic category document related to shopping center. Increasing the λ results in high overlap in category documents with irrelevant categories. Hence, we set $\lambda = 5$ which was optimal and can vary depending on the context for which the corpus of category document is being generated. This collection of 18 documents corresponding to each Wiki category given in Table 1 along with number of sub-categories act as the training corpus or shopping center context.

4.2.2 Wi-Fi Access Point Semantics Assignment. In order to correlate user physical movements captured by the access points, it is needed to label each AP with semantics corresponding to its location in the shopping center.As each AP covers a certain area with signal in the mall, the service area is approximated by a so called *Voronoi cell*, in which any location is closest to its seed location (the AP) than to any other seed location [2]. Then, we have manually rectified them to match the shop frontages and thus better represent

Table 1: Semantic Categories. The number in the brackets denotes the number of sub-categories, product names, and related terms.

1. Bags (104)	2. Bakeries (48)
3. Clothing (183)	4. Coffee (74)
5. Consumer Electronics (381)	6. Cosmetics (173)
7. Decor (188)	8 . Fashion Accessories (203)
9. Food Retail (91)	10. Footwear (94)
11. Home Appliances (174)	12. Jewelery (153)
13. Mobile Phones (229)	14. Restaurants (127)
15. Retail (214)	16. Sports (141)
17. Watches (123)	18. Fashion (292)

physical contexts [22]. Once we get these Voronoi cells, we know which shop falls under an AP from the shopping center floor map (on average, there are 3.67 shops in each Voronoi cell). We then assign a list of semantic categories to an AP corresponding to each shop in the region covered by an AP.

4.3 Cyber Context

As defined earlier, *cyber context* is a document of Wikipedia categories for the entities/concepts extracted from user queries. The generation of this document is explained in the next section.

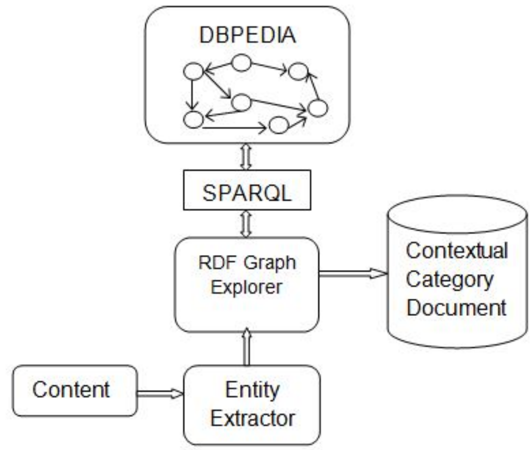
User Query Extension using CCS: Content Categorization System. Given a query content, CCS extracts entities from the query, then gather contextual Wikipedia categories for each entity from DBPedia [1]. DBPedia provides structured information from Wikipedia in the Resource Description Framework (RDF) format, which can be queried using SPARQL, a query language to retrieve and manipulate data stored in RDF format. The CCS system is shown in Figure 1a, and we describe the components by taking the query, "*The Face Shop clear mascara review*", as an example in the following sections. The process of query categorization is quite similar to [12]

Entity Extraction. Entity extraction in our system can be considered as a component that retrieves DBpedia resources from the query text. We use Targeted Hypernym Discovery(THD) [4], an advanced unsupervised entity discovery and classification system for entity extraction. THD discovered two entities from the example query given above, the Face Shop and Mascara

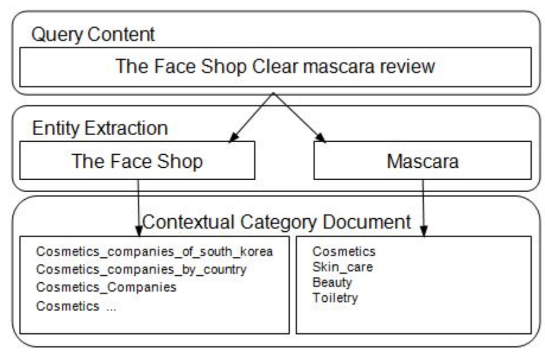
Graph Explorer. Graph explorer is a depth first search algorithm written to explore DBPedia graph. This algorithm takes, as its input, a list of entities annotated from input query text by entity extraction and looks for resources connected to it via the Simple Knowledge Organization System (SKOS) properties skos:subject and skos:broader. The algorithm iterates through each entity discovered and performs a depth first search on the DBPedia graph. The subject is retrieved only for the main entity discovered and for the broader property. Then the graph is iterated recursively for *n* hops to form a contextual category list and return this list as a document.

4.4 Cyber-Physical Contextual Similarity

In this section, we define the contextual similarity between user's physical movements with what they are looking for online.



(a) Content Categorization System Architecture



(b) Example

Figure 1: Content Categorization System

4.4.1 Physical Context. We consequently define the physical context as the area in the shop served by a single AP, characterized in terms of represented latent semantic categories from Wikipedia as described in Section Semantic Categorization, denoted as $C = \{c_1, c_2, ..., c_h\}$, where h is the number of categories.

The categories, generated by the proposed CCS system (as shown in Fig. 1) are represented as *documents* $D = \{d_1, d_2, ..., d_h\}$ of subcategories and broader categories for $c_i \epsilon C$. Then, we denote the physical context for each AP as $P_a : \{p_{a,1}, p_{a,2}, ..., p_{a,l}\}$, where $p_{a,i} \in C$ for all shops that are located in the Voronoi regions of AP a_i.

4.4.2 Physical Activities. We define physical activity of a user in terms of a trajectory $T = ((a_1, t_1), \ldots, (a_n, t_n))$ which is a list of tuples of visited AP IDs and the cumulative time of association. We use $a = \{a_1, a_2 ..., a_n\}$ to represent AP, where n is the number of APs user connected to during a single visit to the mall and t to represent time association where t_k is the duration user spent connected to a_k during the visit: $t = \{t_1, ..., t_k, ..., t_n\}$. If a user was associated with an AP multiple times in a visit, the total duration of time spent at this AP is stored.

Annotated Category	Identified Category	Query
Cosmetics	Cosmetics	the face shop clear mascara reviews, Muk Hair Wax
Clothing	Clothing	Superdry Sale, Emporio Aramani
Fashion	Fashion	TopShop Sydney
Footwear	Footwear	Ugg Shoes
Mobile Phones	Mobile Phones	Nokia Lumia 520 reviews

Table 2: Semantic Categories with Max Cosine Similarity for sample queries

4.4.3 Cyber Context. We define cyber context in terms of queries extracted from query logs(terms extracted from click through URLs). During a single session of a user, we extract queries from all URLs accessed by user and represent as $q = \{q1, q2, ..., qj\}$, j is the total number of queries extracted from the URLs visited by user. We use our Content Categorization system, defined in Section 4.3, denoted as $C_{q_i} = \{c_{q_1}, c_{q_2}, ..., c_{q_m}\}$ for each $q_i \in q$. Cyber context will then be presented as

$$Q_c = \bigcup_{i=1}^{m} c_{q_i} = c_{q_1} \cup c_{q_2} \cup \cdots \cup c_{q_m}. \quad (1)$$

4.4.4 Similarity Between Physical and Cyber Contexts. The similarity between the physical context P_c and Cyber Context Q_c is calculated in 2 steps. First, we represent the TF-IDF (Term Frequency - Inverse Document Frequency [23]) between each document $d_i \in d$ and Q_c as $V(d_i)$ and $V(Q_c)$, then we compute the cosine similarity with

$$cos(d_i, Q_c) = \frac{V(d_i).V(Q_c)}{|V(d_i)||V(Q_c)|}. \quad (2)$$

The contextual similarity with Semantic Category c_i represented as $CS(c_i)$ is $cos(d_i, Q_c)$ boosted with Physical Context similarity i.e. time spent at each category denoted as t_{c_i}.

$$CS(c_i, Q_c) = t_{c_i} * cos(d_i, Q_c) \quad (3)$$

where $t_{c_i} > 0$ and $cos(d_i, Q_c) > 0$.

4.5 Analysis

We examined cosine similarity results for user issued queries with Semantic Categories. We first annotated each query issued by a user manually with 18 Semantic Categories. The annotation was done by 3 participants who were given a list of queries and a list of semantic categories. They performed the task independent of the contextual similarity model and were asked to label queries with the semantic categories out of the given list. We then compare the distribution of count of manually annotated query categories with the Top-3 categories retrieved by max cosine similarity of query document with Semantic Category documents Figure 2. To see the similarity in distribution across the two sets of categories labels, we calculated Pearson Correlation Coefficient at significance level of $p < 0.05$ and found to be significant ($R = 0.6084$ and p-value $= 0.0073$).

In Table 2, we show the category with max Cosine Similarity for a given query. For example, given a query *Ugg Boots* which is a footwear brand we get maximum cosine similarity for *Footwear* Semantic Category as shown in Figure 3. Figure 4 shows the Cosine

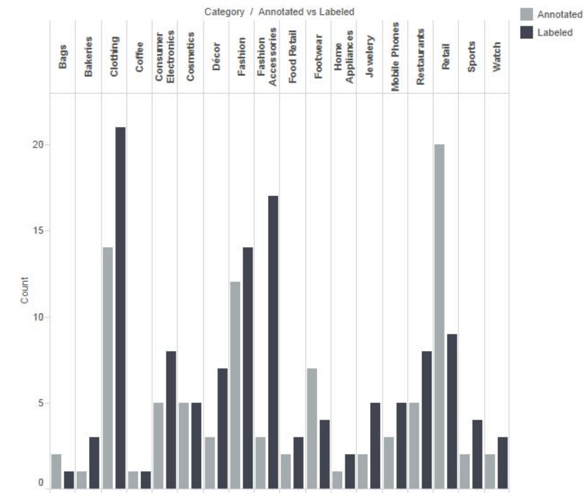


Figure 2: Distribution of manually annotated and labeled categories

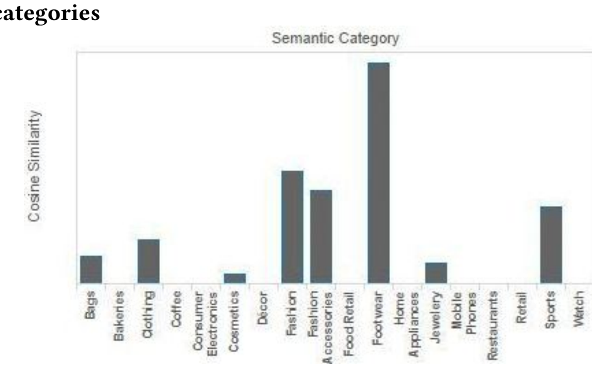


Figure 3: Cosine Similarity for Cyber Query: Ugg Shoes

Similarity for each Semantic Categories for query *the face shop clear mascara reviews*, and it is observed the max similarity is for *Cosmetics*, which is clearly correct. We evaluated query categorization using Accuracy@3 for 217 manually annotated queries. We were able to categorize 49.2% of the queries correctly which shows successful mapping of the query text with the Semantic Categories using the proposed approach. In our work, Shopping Intent recognition explained in next section, we did not use the specific interest category from query but the similarity distribution across all categories for given query set per user trajectory as a feature vector that showed to improve the accuracy of Shopper classification into 2 categories intentful and intentless.

5 SHOPPING INTENT RECOGNITION SYSTEM

Given AL, QL, Shop Categories and Voronoi Regions we create an Intent Recognition Model as shown in Figure 5. The first step is to enrich indoor semantics by using the shop categories as given by mall operator. This step is explained in detail in Section 4 where we create a corpus of 18 documents containing terms and subcategories for Semantic categories shown in Table 1 along with the count of sub-categories in each category document. Once we get the semantic categories for the shops in the shopping center, we

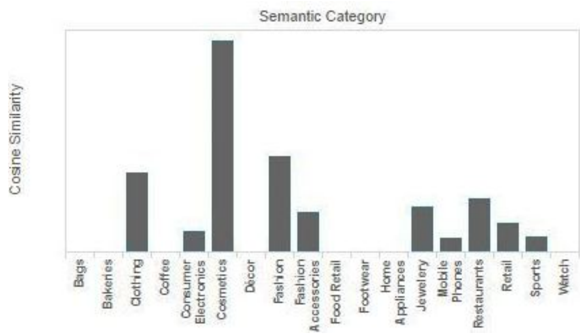


Figure 4: Cosine Similarity for Cyber Query: the face shop clear mascara reviews

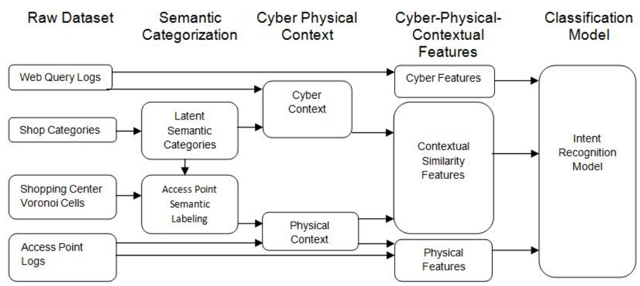


Figure 5: Shopping Intent Recognition System

then need to label each access point with these categories based on the range covered by the AP determined by voronoi cells as explained in Section 4.2.2 of Section 4. After semantic labeling of each AP, we then extract Physical Activity and Cyber Context for user sessions i.e. user visit to the mall from AP and Web query logs. We define *Physical Activity* for a user session as total time spent by user in each semantic category for a visit to the mall and *Cyber Context* as a document of categories/entities extracted from users' queries issued in a single visit. We then calculate the Cyber-Physical Contextual similarity from physical activity and cyber context of the user. This will give us Contextual Similarity features that act as input to our Intent Recognition Model along with Physical and Cyber features derived from AL and QL.

5.1 Cyber-Physical-Contextual Features

We investigate an approach for recognizing in-store shopping behavior from an individual's physical movements from Wi-Fi traces and cyber activity from Web queries that users issue. Our approach rests on the belief that user intent can be identified by correlating their movements with the content they look online. During a typical visit to a shopping center, a shopper uses Wi-Fi either for browsing or when they are looking for some shop or an item they are interested in. If the user who is using Wi-Fi, has a shopping intention, then there is high possibility that they visit some specific category shops and look for related items/category online either to compare the price or for reviews. For example, we show part of the user trajectory in Table 3 where user looked for "nest au homeware" online and an association of more that 10 minutes was found with an AP wap032 listed under category "Homeware". We try to correlate this behavior using 3 feature set , Physical, Query

Wi-fi AP	AP Semantics	Query
wap030	Restaurant Cafe Groceries	nest au homeware
wap032	Homeware Clothing Footwear	
wap009	Clothing Footwear	
...	...	...

Table 3: User Trajectory

and Contextual as given below where we use Trajectory-based Cyber-Physical contextual Similarity for contextual features. These features are then used to build a binary classifier for labeling the user trajectory as Intentful(IF) or Intentless(IL):

1) *Physical Activity vs Intent*:

- F1: Trajectory length: is defined as the number of APs in a user's trajectory;
- F2: Total duration: how long they spend in the mall in seconds;
- F3-F20: Time spent per shop category: means the distribution of the total duration over shop categories.

2) *Cyber-Physical activity vs Intent*: F1-F20; F21: # of queries.

3) *Contextual Features vs Intent*:

- F22-F39: $CS(c_1)$-$CS(c_{18})$ - Contextual Similarity of User's Cyber-Physical activity with Semantic Category documents i.e. d_1-d_{18};
- F40: Max Contextual Similarity - is $max(CS(c_1) : CS(c_{18}))$;
- F41: Sum of $CS(c_1)$:$CS(c_{18})$;
- F42: Cosine Similarity of query with shop list - is the Cosine Similarity of categories extracted from user issued queries in a single visit with the list of over 200 stores in the shopping center;
- F43: Cosine Similarity between query and keyword list from Foursquare, Yelp and Google Places - is the Cosine Similarity of categories extracted from user issued queries in a single visit against the list of keywords/categories extracted from crowdsourced Web applications including Foursquare, Yelp and Google places for stores in the shopping center.

5.2 Intent Recognition Model

As most of our cyber-physical-contextual features are independent of each other, we deploy Decision Table/Naive Bayes(DTNB) hybrid classification method [7] to perform the Intentful and Intentless classification. DTNB selects the deterministic features for recognizing intent of a user's visit to a shopping center from a range of input features, and we show how Decision Table, Naive Bayes classifier, and the Hybrid model work with the proposed features as follows.

Decision Table. Given a set of labeled instances as a training sample, e.g. the labelled intentful/intentless sessions, an induction algorithm creates a decision table with default rule mapping to the majority class, which we abbreviate as Decision Table Model (DTM), including two main components:

- Schema: set of features selected by maximizing cross-validated performance using forward search.
- Body: multiset of labeled instances.

Each instance consists of a value for each of the features in the schema and a value for the class.

For label assignment to an unlabeled instance I by a DTM classifier, let L be the set of labeled instances in the DTM matching given instance I. There is a match between 2 instances if the features in the schema are same. If $L = 0$, DTM returns the majority class, otherwise return the majority class in L.

Naive Bayes. Naive Bayes classifier is a widely used Machine Learning technique based on the well-known Bayes Theorem, which states:

$$p(l_i|f_i) = \frac{p(f_i|l_i)p(l_i)}{p(f_i)}, \tag{4}$$

where l_i is a class label and f_i is a feature from the set of contextual features described in Section Cyber-Physical Contextual Similarity; $p(l_i, f_i)$ is probability of f in l_i; $p(f_i|l_i)$ is the probability of f_i given class l_i; $p(l_i)$ is the probability of occurrence of class l_i and $p(f_i)$ is the probability of occurrence of feature f_i.

Considering the features are defined from physical and cyber perspectives, we assume that they have independent distribution and thereby Eq. 4 becomes:

$$p(f|l_i) = p(f_1|l_i) * p(f_2|l_i) * ... * p(f_n|l_i). \tag{5}$$

In a classification task, given feature set $f = \{f_1, f_2, ..., f_n\}$ for binary classification of $\{l_i, l_j\}$, Bayes Classifier labels an instance as class l_i if its posterior probability is higher than the other class, namely $p(f|l_i) > p(f|l_j)$.

Decision Table Naive Bayes(DTNB) Hybrid model. DTNB hybrid model is a simple Bayesian network in which the decision table (DT) represents a conditional probability table [7]. The algorithm for learning the combined model (DTNB) works in a similar way as that of stand-alone DTs. It basically splits the feature set into two disjoint subsets: one for the DT, the other for NB. Then, it uses forward selection, where, at each step, selected attributes are modeled by NB and the remainder by the DT.

The class probability of the DT and NB are then combined to generate overall class probability estimates. Assuming f_{DT} is the set of features in the DT and f_{NB} the one in NB, the overall class probability is computed as

$$P(l_i|f) = a * P_{DT}(l_i|f_{DT}) * P_{NB}(l_i|f_{NB}/P(l_i)) \tag{6}$$

where $P_{DT}(l_i|f_{DT})$ and $P_{NB}(l_i|f_{NB})$ are the class probability estimates from the DT and NB respectively, a is a normalization constant, and $P(l_i)$ is the prior probability of the class label l_i.

5.3 Future Location Prediction

Here, we deploy the proposed findings in this study to further investigate: Given users' physical and cyber activities in terms of Wi-Fi AP association and Web query logs, is the semantic content of queries are indicative for location prediction?

Specifically, we set up the following configurations: Provided with a list of m user trajectories $T = \{t_1, t_2, ..., t_m\}$ and a list of n access points $A = \{a_1, a_2, ..., a_n\}$. Each user trajectory t_i has a list of AP A_{t_i} where $A_{t_i} \subset A$, which the user has connected to in the order of association time and a set of n queries $Q = \{q_1, q_2, ..., q_n\}$. Given that, can we calculate the likelihood of an unvisited ap $a_j \notin A_{t_i}$ that the user will visit in future for a target trajectory $t_j \in T$. For example, given a user's current physical and cyber features, we can predict the likelihood he will visit other APs. This is reasonable for indoor environment due to its structured lay-out.

For the recommendation algorithms, we deploy both Item-Based Collaborative Filtering [19] with Contextual Similarity defined as follows.

Contextual Similarity. The similarity between the physical and Cyber Context Q_c is calculated in 2 steps. First cosine similarity between each document $d_i \in d$ and Q_c TF-IDF vector represented as

$$cos(d_i, Q_c) = \frac{V(d_i).V(Q_c)}{|V(d_i)||V(Q_c)|} \tag{7}$$

generating a similarity vector CS of size h where S_i is the similarity of query document Q_c with category document d_i.

The second step is to generate a dot product of physical context vector P_{a_i} for $a_i \in A$ with the similarity vector CS that represents user query context corresponding to each AP a_i as follows

$$SS = P_{a_i} \cdot CS = \sum_{j=1}^{h} P_{a_i,j} \cdot CS_j = P_{a_i,1} \cdot CS_1 + P_{a_i,2} \cdot CS_2 + ...+ P_{a_i,h} \cdot CS_h \tag{8}$$

where h is the no. of categories.

The semantic similarity vector SS where SS_i is the semantic similarity of AP a_i , used to weigh the similarity score calculated in Item-Item similarity by taking the product of JS_i and SS_i where i is an AP a_i:

$$weightedSimilarity(JS_i, SS_i) = JS_i * SS_i. \tag{9}$$

Finally, prediction is then given by sorting the weighted similarity score and extracting top k items.

6 EXPERIMENTS

We evaluate the performance of DT, NB and DTNB hybrid approach for shopper's intent recognition by using different sets of Physical (Phy), Cyber(Cyb), Contextual(Cont) features (see above). We report the final results of a 10-fold cross-validation using Weka's implementation of DT, NB and DTNB. The classification accuracy for identifying shopper's intent are shown in Table 4 on the dataset described in next section.

6.1 Experiment Configuration

We focus on a subset of *complete* user trajectories where users issued Web queries. A complete trajectory is a trajectory where the user enters from one entry point of the mall and exits from the same or other exit point, thus connecting at least 3 AP's. Out of 6784 total trajectories, we have identified only 176 such complete trajectories in our dataset. Four annotators without in-depth knowledge of the experiment then manually categorised the 176 trajectories into *intentful* (48 trajectories) and *intentless* (128 trajectories), with 100% inter-annotator agreement. The annotators inspected only the queries and marked them as *relevant* if deemed related to the environment of the shopping centre. A session was labelled as *intentful* if at least one of the queries issued in this session is *relevant*, and *intentless* otherwise.

For intent recognition, we used Accuracy%, F-Score, Precision and Recall to evaluate classification model: *Accuracy%* denotes what percent of classification was correct; *Precision* denotes what percent of positive classification were correct; *Recall* denotes what percent of positive instances were correctly classified; $F - Score$ is a weighted harmonic mean of Precision and Recall.

For location prediction, we used Accuracy@k and Mean Reciprocal Rank(MRR) to evaluate prediction results. *Accuracy@k* is number of correct locations predicted over k, which is the total no. of locations predicted. *MRR* is used to evaluate the ranking of first correct location predicted and can be calculated as $MRR = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{R_i}$, where n is the no. of prediction results and R_i is the rank of first correct predicted location for trajectory i.

6.2 Results on Intent Recognition

Features	Method	Accuracy %	F-Score	Precision	Recall
Phy	NB	63.06	0.59	0.56	0.63
	DT	72.73	0.61	0.53	0.73
	DTNB	72.73	0.61	0.53	0.73
Phy + Cyb	NB	63.06	0.59	0.56	0.63
	DT	78.41	0.73	0.81	0.78
	DTNB	78.41	0.73	0.81	0.78
Cont	NB	73.29	0.68	0.69	0.73
	DT	76.13	0.73	0.74	0.76
	DTNB	76.7	0.75	0.75	0.77
Phy + Cont	NB	69.32	0.66	0.65	0.69
	DT	76.14	0.73	0.74	0.76
	DTNB	76.14	0.74	0.74	0.76
Phy + Cyb + Cont	NB	69.32	0.66	0.65	0.69
	DT	78.41	0.75	0.79	0.78
	DTNB	**81.25**	**0.8**	**0.8**	**0.81**

Table 4: Intent Recognition Results

The DTNB hybrid classifier always performs comparably or better than DT and NB. The best accuracy of 81.25% is achieved with DTNB on all Cyber-Physical-Contextual features, and demonstrated how accurately shoppers' intent can be identified. Moreover, the Intent Recognition results (Table 4) show how contextual features improve the accuracy of classification. To verify whether the contribution of contextual features is statistically significant, we tested pairs of Physical with Physical+Contextual, and Cyber+Physical with Cyber+Physical+Contextual features (two-tailed paired t-test [8], with a 95% confidence level). Each group in a pair contains 12 distinct values corresponding to accuracy, f-score, precision and recall for 3 different methods DT, NB, DTNB giving a df of 11. The results show that the increase in performance with contextual features is statistically significant.

6.3 Results on Future Location Prediction

We performed prediction experiment on 994 full and partial trajectories where at least one query was issued. Full trajectory is a trajectory where user entered from one entry point and exited from the same or the other whereas partial trajectory is a trajectory where one connects somewhere in the middle but exits from either of the exit points. We then partitioned 325 trajectories into train and test trajectories. The partition point is the access point where user issued its first query and the rest of the trajectory (access points in the other half) are used for evaluating prediction results to see if the semantic context of queries with respect to physical locations helps in improving prediction results. We then used 669 full trajectories and 325 partitioned train trajectories to generate collaborative filtering matrix and get Top-10 prediction results for 325 partitioned test trajectories using simple Item-Item method and Item-Item-Weighted using Semantic Similarity weights.

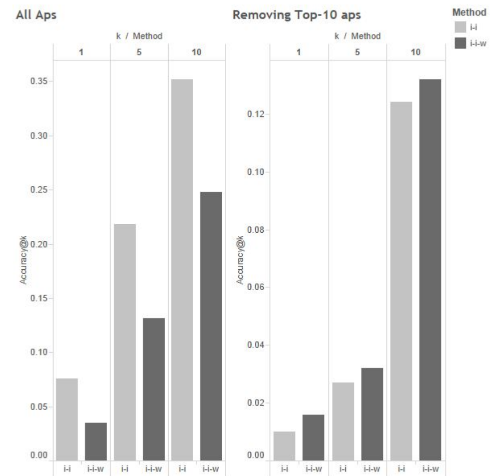


Figure 6: Prediction Results

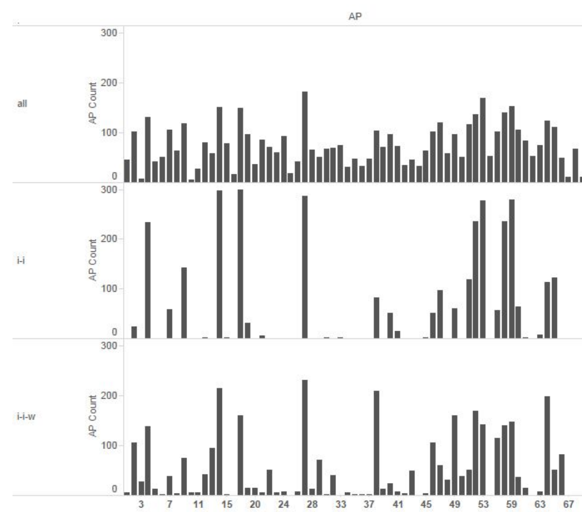


Figure 7: Prediction Results

Bar chart on the left in Figure 6 show results with no improvement in accuracy using contextual similarity weight. We then generated a chart of predicted APs on x-axis and count of AP's on y axis in the test set(all) along with predicted AP's using *i-i* and *i-i-w* methods as shown in Figure 7. From the visualization, we see that *i-i-w* method (third bar from top) works well at predicting less popular APs. This can be because less popular locations might be semantically similar.

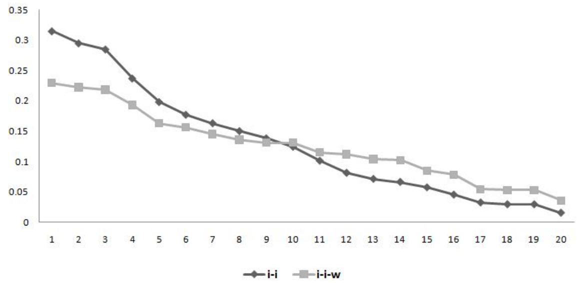


Figure 8: Sensitivity Analysis of Accuracy@10 by removed top-n APs, n ranging from 1 to 20

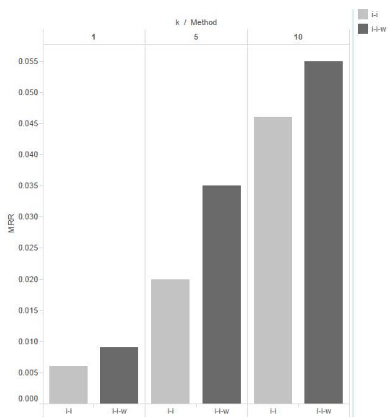


Figure 9: Mean Reciprocal rank for k predictions, k = 1,5,10

To assess the correctness of our assumptions based on the chart, we did sensitivity analysis on Accuracy@10 by removing top 1-20 APs. Figure 8 shows that *i-i-w* consistently outperforms *i-i* on removing some of the popular APs. We then checked Accuracy@k for $k = \{1, 5, 10\}$ after removing Top-10 APs from the test set (Figure 6). We see an improvement in accuracy for item-item weighted(i-i-w) compared to item-item(i-i) where accuracy increases with increase in no. of predictions(k). The improvement in accuracy is statistically significant between *i-i* and *i-i-w* ($p = 0.0188$, two-tailed paired t-test [8]). We also used *MRR* to evaluate the ranking of first correct location predicted using *i-i* and *i-i-w* for top-k prediction, $k = 1, 5, 10$. As shown in Figure 9, MRR for *i-i-w* is better then *i-i*. We thus conclude that contextual similarity improves prediction of less popular locations with better ranking as well.

7 CONCLUSION

We proposed a semantic enrichment and contextual similarity model to deal with one major challenge, mapping semantic similarity across two different domains: cyber and physical behaviours. It is obtained that using this contextual similarity can further improve the accuracy of both applications (intent recognition and location prediction) with respect to just using cyber/physical features only. We also show that the proposed contextual features significantly improve the accuracy of intent recognition and future location prediction. A validation of this approach on a more comprehensive is envisaged, using crowdsourcing for the manual labelling task.

REFERENCES

[1] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. *Dbpedia: A nucleus for a web of open data.* Springer.

[2] Y. B. Bai, S. Wu, G. Retscher, A. Kealy, L. Holden, M. Tomko, A. Borriak, B. Hu, M. Sanderson, H. R. Wu, and K. Zhang. 2014. *A New Method for Improving Wi-Fi Based In-door Positioning Accuracy.* Springer Verlag, Berlin.

[3] Karen Church and Barry Smyth. 2009. Understanding the Intent Behind Mobile Information Needs. In *IUI (09).* ACM, New York, NY, USA, 247–256.

[4] Milan Dojchinovski and Tomas Kliegr. 2013. Entityclassifier.eu: Real-time Classification of Entities in Text with Wikipedia. In *ECMLPKDD (2013).* 1–1.

[5] Huizhong Duan and ChengXiang Zhai. 2015. Mining Coordinated Intent Representation for Entity Search and Recommendation. In *CIKM (2015).* ACM, New York, NY, USA, 333–342.

[6] Moustafa Elhamshary and Moustafa Youssef. 2014. CheckInside: A Fine-grained Indoor Location-based Social Network. In *UbiComp (14).* ACM, New York, NY, USA, 607–618.

[7] Mark A Hall and Eibe Frank. 2008. Combining Naive Bayes and Decision Tables.. In *FLAIRS Conference*, Vol. 2118. 318–319.

[8] Henry Hsu and Peter A Lachenbruch. 2008. Paired t test. *Wiley Encyclopedia of Clinical Trials* (2008).

[9] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. 2008. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management* 44, 3 (May 2008), 1251–1266.

[10] Kasthuri Jayarajah, Youngki Lee, Archan Misra, and Rajesh Krishna Balan. 2015. Need Accurate User Behaviour?: Pay Attention to Groups!. In *UbiComp (15).* ACM, New York, NY, USA, 855–866.

[11] John Krumm and Dany Rouhana. 2013. Placer: Semantic Place Labels from Diary Data. In *UbiComp (13).* ACM, New York, NY, USA, 163–172.

[12] Michal Laclavík, Marek Ciglan, Sam Steingold, Martin Seleng, Alex Dorman, and Stefan Dlugolinsky. 2015. Search Query Categorization at Scale. In *WWW.* International World Wide Web Conferences Steering Committee, 1281–1286.

[13] Claudio Martella, Armando Miraglia, Marco Cattani, and Maarten van Steen. 2016. Leveraging Proximity Sensing to Mine the Behavior of Museum Visitors. In *PerCom 2016.* IEEE.

[14] Archan Misra and Rajesh Krishna Balan. 2013. LiveLabs: Initial Reflections on Building a Large-scale Mobile Behavioral Experimentation Testbed. *SIGMOBILE Mob. Comput. Commun. Rev.* 17, 4 (Dec. 2013), 47–59.

[15] Wendy W Moe. 2003. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of consumer psychology* 13, 1 (2003), 29–39.

[16] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. 2012. Mining user mobility features for next place prediction in location-based services. In *ICDM.* IEEE, 1038–1043.

[17] Veljko Pejovic, Neal Lathia, Cecilia Mascolo, and Mirco Musolesi. 2015. Mobile-Based Experience Sampling for Behaviour Research. *arXiv preprint arXiv:1508.03725* (2015).

[18] Meeralakshmi Radhakrishnan, Sharanya Eswaran, Archan Misra, Deepthi Chander, and Koustuv Dasgupta. 2016. IRIS: Tapping Wearable Sensing to Capture In-Store Retail Insights on Shoppers. In *PerCom 2016.* IEEE.

[19] Yongli Ren, Gang Li, and Wanlei Zhou. 2015. A survey of recommendation techniques based on offline data processing. *Concurrency and Computation: Practice and Experience* 27, 15 (2015), 3915–3942.

[20] Y. Ren, M. Tomko, F. D. Salim, J. Chan, C. L. A. Clarke, and M. Sanderson. 2018. A Location-Query-Browse Graph for Contextual Recommendation. *IEEE TKDE* 30, 2 (Feb 2018), 204–218.

[21] Yongli Ren, Martin Tomko, Flora D. Salim, Jeffrey Chan, and Mark Sanderson. 2018. Understanding the predictability of user demographics from cyber-physical-social behaviours in indoor retail spaces. *EPJ Data Science* 7, 1 (2018), 1.

[22] Yongli Ren, Martin Tomko, Flora Dilys Salim, Kevin Ong, and Mark Sanderson. 2017. Analyzing Web behavior in indoor retail spaces. *JASIST* 68, 1 (2017), 62–76.

[23] Gerard Salton and Christopher Buckley. 1988. Term-weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manage.* 24, 5 (Aug. 1988), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

[24] C. W. You, H. L. C. Kao, B. J. Ho, Y. H. T. Chen, W. F. Wang, L. T. Bei, H. H. Chu, and M. S. Chen. 2014. ConvenienceProbe: A Phone-Based System for Retail Trade-Area Analysis. *IEEE Pervasive Computing* 13, 1 (Jan. 2014), 64–71.

[25] C. W. You, C. C. Wei, Y. L. Chen, H. h Chu, and M. S. Chen. 2011. Using Mobile Phones to Monitor Shopping Time at Physical Stores. *IEEE Pervasive Computing* 10, 2 (April 2011), 37–43.

[26] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. 2015. Analyzing Shopper's Behavior Through WiFi Signals. In *Proceedings of the 2Nd Workshop on Workshop on Physical Analytics (WPA '15).* ACM, New York, NY, USA, 13–18.