Identifying the Activities Supported by Locations with Community-Authored Content

David Dearman and Khai N. Truong
University of Toronto
Department of Computer Science
Toronto, Ontario M5S 3G4, Canada
{dearman, khai}@cs.toronto.edu

ABSTRACT

Community-authored content, such as location specific reviews, offers a wealth of information about virtually every imaginable location today. In this work, we process Yelp's community-authored reviews to identify a set of potential activities that are supported by the location reviewed. Using 14 test locations we show that the majority of the 40 most common results per location (determined by verb-noun pair frequency) are actual activities supported by their respective locations, achieving a mean precision of up to 79.3%. Although the number of reviews authored for a location has a strong influence on precision, we are able to achieve a precision up to 29.5% when processing only the first 50 reviews, increasing to 45.7% and 57.3% for the first 100 and 200 reviews, respectively. In addition, we present two context-aware services that leverage location-based activity information on a city scale that is accessible through a Web service we developed supporting multiple cities in North America.

Author Keywords

activity, community-authored content, location, reviews

ACM Classification Keywords

H.3.m. Information storage and retrieval: Miscellaneous.

General Terms

Experimentation, Human Factors, Performance.

INTRODUCTION

Context-aware applications commonly require knowledge of a person's location and activity. Current methods for sensing activity rely on low cost sensors (*e.g.,* RFID, accelerometers) worn on the body [2, 5, 14, 15], integrated within an environment's infrastructure [8, 9, 18, 21, 22] and within everyday objects used by people within the environment [11, 24]. However, these methods for sensing and inferring context lack generality because they classify only a limited number of activities and are typically limited to a specific environment (*e.g.,* the home [8, 9, 18, 21, 22]) or domain (*e.g.,* weight lifting [5]). Furthermore, most of the places a person visits will support large and diverse ranges of activities, including activities that might be specific only to that location. The set of activities a person can perform at a location (*i.e.,* the activities supported by the location) is not static; rather, the activities can change depending on the situations that arise and the people who experience them [1, 7].

In this work, we propose a novel approach to identifying the set of potential activities supported by the location(s) a person may visit by processing community-authored reviews on Yelp. We exploit this rich source of data to identify the potential activities (as verb-noun pairs) that are supported by a reviewed location for a large number of reviewed locations in order to build a corpus of potential activities for each location. The potential activities we identify are based only on the respective location's reviews, not generalized activities based on the location's category (*e.g.,* restaurants). The activities we identify may not be immediately related to a user's current activity or the live context for an environment. However, they characterize a wide variety of *potential user activities* that are supported by a location, which is valuable for designing context-aware applications.

In this paper, we provide an overview of the Yelp community; the methods we used to identify activities (articulated as verb-noun pairs; *e.g.,* 'read book') from Yelp's review texts; and a user study to evaluate the precision, recall and validity of the identified verb-noun pairs. The results of the evaluations show that:

- Community-authored reviews, specifically Yelp reviews, are a diverse and comprehensive data source that can be processed to identify the activities supported by the reviewed location. With respect to the 40 most common verb-noun pairs we achieve a mean precision up to 79.3% and recall up to 55.9%.
- The number of reviews authored for a location has a significant impact on precision for the 40 most common verb-noun pairs. We achieved a mean precision up to 29.5% when processing only the first 50 reviews, increasing to 45.7% and 57.3% for the first 100 and 200 reviews, respectively.

UbiComp'10, September 26–29, 2010, Copenhagen, Denmark.

- There is a significant difference in the activities that participants describe performing at a location and the 40 most common verb-noun pairs identified for a location. The difference highlights the personal and individualized nature of the provided activities in contrast to the verb-noun pairs identified in the community-authored reviews.

These findings demonstrate that community-authored content can be successfully processed to identify valid activities supported by a location. This data source exists, is readily available, is constantly growing, and requires limited effort to harvest. In addition to evaluating our method, we have developed a Web-accessible activity service that provides city-wide activity information for multiple North American cities and demonstrate its use for research and development of context-aware applications.

RELATED WORK

The purpose of this work is not to infer *what* activity a person is *currently* performing but to identify a set of potential activities that are supported by a person's current location and other locations she can visit.

The activity a person is performing can be sensed and inferred by analyzing the physical manifestation of the base-level actions that are performed when executing the activity [2, 4, 6, 9, 11, 14, 17, 18, 24, 26]. These methods are typically constrained to a specific space (*e.g.,* the home [8, 9, 18]) and domain (*e.g.,* household activities [11, 18, 24], modes of transportation [2, 14, 15, 23]) and commonly employ low-fidelity sensors (*e.g.,* RFID tags, infra-red motion detectors, accelerometers) integrated within an environment's infrastructure [8, 9, 18, 21, 22] and usable objects [11, 24], or worn on the body [2, 5, 14, 15].

On-the-body sensing is challenging because it requires a tradeoff between simplicity and sensing a greater number of features [5, 15]. As a result, accelerometers have been explored heavily as a viable activity sensor in recent years. For example, Lester *et al.* instrumented people with a multi-sensor board (*e.g.,* accelerometers, audio, barometric pressure) showing that accelerometers produced the most useful features when classifying a small set of simple activities (*e.g.,* sitting, standing, walking, running, driving, biking) [15]. Similarly, Lee *et al.* combined the sensed features from an accelerometer and compass worn on the body to infer sitting, standing and walking behaviours within a controlled environment [14]. Bao *et al.* instrumented people with multiple accelerometers and asked them to perform whole body activities (*e.g.,* running) and sedentary activities (*e.g.,* lying down) [2]. Chang *et al.* focused specifically on free-weight exercise [5]; using a single 3-axis accelerometer placed on the user's hand and hip, they were able to accurately differentiate among nine kinds of exercises. Although these techniques are fairly robust, they lack generality. It is challenging to differentiate activities that produce similar features from the sensed data. As a result, many researchers limit their techniques to recognize only a small set of activities determined a priori for specific environments.

Sensing user activities within the home has been the focus of many projects [8, 9, 11, 18, 21, 22, 24]. Philipose *et al.* instrumented common household objects with RFID tags to detect activities within the home (*e.g.,* making tea) in order to show that the specific actions performed with these objects and the sequence of these actions is an effective method to infer high-level activities [24]. Logan *et al.* instrumented a single dwelling with over 900 low-fidelity sensors to evaluate the effectiveness of different sensing modalities for recognizing the breadth of high-level activities that a person may perform inside the home [18]. They found that the type of sensor significantly impacted performance, citing the best performance from multiple infrared motion detectors because of a logical mapping between an activity and a location. Building upon this concept, Hodges *et al.* showed that when an activity is known a priori, an object-use fingerprint that models the activity can be developed to sense actions using RFID tags and infer activities such as "making coffee" [11]. This work was extended by Huynh *et al.* by showing that the object-use fingerprint used to model an activity can be sensed rather than defined *a priori* [12].

In contrast to the more complex distributed sensing techniques described above, an alternative approach is to use a small number [8] or single [9, 21, 22] sensor deployments at critical points in a home's infrastructure. Fogarty *et al*. deployed a small number of sensors at critical points in a home's water distribution infrastructure [8]. Using simple microphone sensors, they were able to efficiently classify and identify activities associated with specific water fixtures (*e.g.,* use of the clothes washer, dishwasher, shower and toilet). Froehlich *et al.* simplified the sensing needs by detecting similar activities using a single sensing unit that can be installed at any point along the water distribution infrastructure [9]. Patel *et al.* took a different approach by implementing a single plug-in sensor that detects unique electrical events (*e.g.,* light switches, electric stove, TV) [22]. Similarly, they instrumented the HVAC air handler within a home to accurately detect unique transition events between rooms [21].

The techniques described above highlight the simplicity of sensing required to infer certain context-constrained events. However, as Logan previously noted [18], in real world deployments it is difficult to characterize the specific actions and sequence of actions that are performed during a high-level activity. The activities people perform are commonly intertwined, and it is not always evident which action is associated with which activity [18]. The applicability of these sensing platforms to fully account for the complexities introduced in different unconstrained environments (*e.g.,* public locations) remains to be explored. The underlying assumption that the set of activities for a location can be constrained *a priori* does not hold. A location, particularly a public one, is a dynamic space where the activities that take place are dependent upon the situations that arise [1, 7]. That said, it is important to take a step back and first develop methods to determine the set of potential activities supported by a location (*e.g.,* “purchase book”) or may occur as a result from being in the place (*e.g.,* “get inspired”). The purpose for identifying and collecting these activities is to bootstrap and support a breadth of context-aware applications that require contextual knowledge of the activities that can be performed at a specific location.

A LOCATION-BASED ACTIVITY SERVICE

Community-generated content offers a wealth of information about a location that can be processed to reveal greater context about a location [19, 28]. In this work, we explore the use of community-authored reviews (*e.g.,* Yelp.com) as a rich data source that can be processed to identify the potential activities supported by a location. The ubiquity and continuous growth of community-authored content for diverse and numerous locations make it possible to determine location-specific attributes on a city scale. We leverage this data source to build an on-line service that applications can query for location-specific activity information. Our service currently supports city wide activity information for San Francisco and the South Bay Area, Los Angeles, Seattle, New York, and Toronto.

Yelp Community Authored Reviews

Yelp (Yelp.com) is an online community whose members author reviews about specific public and commercial locations. As of August 2009, “Yelpers” have authored over 7 million reviews about their personal experiences with restaurants, shopping, beauty, arts, entertainment, and events [29]. A review can range from a well-constructed narrative, to a short and simple “it’s great” comment.

Yelp is generally recognized as the premier location review website in North America. The technique described in this paper is not limited to Yelp. It is general enough that it can be adapted to operate with other location-based review communities such as City Search, Kudzu, Insider Pages and Google Pages. Twitter provides the infrastructure for an alternative type of community that is rapidly gaining popularity. However, until recently Twitter did not offer a method to easily geo-locate content. Moreover, the geographic position associated with a “tweet” is based on the poster’s location, not necessarily the location described in the “tweet”.

Yelp reviews are associated with a specific location and are therefore less ambiguous. In addition, a Yelp review is specific to a single instance of a business or entity. For example, New York’s Central Park and San Francisco’s Dolores Park are categorically similar, but the activities these locations support are significantly different. The activity extraction process we discuss in the next section is able to identify activities that distinguish the two apart (*e.g.,* “check zoo”, “take rowboat”, “play chess” for Central Park; and “play tennis”, “sit hill”, “drink beer” for Dolores Park), along with activities that are more commonly supported by parks in general (*e.g.,* “ride bike”, “walk dog”, “walk paths”).

Deriving Activities from Community-Authored Content

The process of deriving the activities supported by a unique location from Yelp’s reviews is accomplished by the *activity service* (Figure 1) in four steps; (1) h*arvest* the review texts and related attributes (*e.g.,* date authored) for each unique location; (2) *parse* the review texts to identify each sentence; (3) *tag* each word of a sentence with its part-of-speech and *extract* local verb-noun pairs to form activities; and, (4) *populate* and *update* the activity database with the identified verb-noun pairs. In this section, we discuss each step in greater detail.

Harvest

First, the *review harvester* retrieves the attributes (*e.g.,* name, URL, latitude, longitude, number of reviews) that describe a unique location using the Yelp Phone API; these attributes are then archived in the activity attribute database. The URL attribute identifies the unique location’s Yelp webpage, which contains the review texts. The URL is accessed and the review texts are extracted using XPath. Given that reviews are continually being authored, the *review harvester* can be configured to extract new reviews on a predefined schedule. The retrieved review texts are then parsed by the *sentence tokenizer*.

Parse

The *sentence tokenizer* then receives the review texts and parses the text into its individual sentences using the Stanford Part-Of-Speech Tagger [25] implementation of the English maximum entropy sentence tokenizer. The review texts are pre-processed and modified to ensure that sentences terminate with appropriate use of punctuation. The words for each sentence are then tagged with their *part-of-speech*.

Tag and Extract

The *part-of-speech tagger* evaluates each sentence using the Stanford Part-Of-Speech Tagger to identify the part-of-speech use (*e.g.,* verb, noun and pronoun) of each word. The *activity finder* then traverses the sequence of tagged words in a sentence to identify the verbs. A verb is ignored if it is less than three characters in length or matches a stop-list word. For each valid verb, the *word-pair finder* identifies all the local nouns (nouns within 5 words following the verb) and pairs the verb with the nearest noun; a sequence of nouns is concatenated to form a single collocation. If a local noun is not found, the verb is ignored. Although the proper noun for a verb is not always found, this method allows for quick pairing of logically local verb-noun pairs. These verb-noun pairs represent the potential activities supported by the location. The *word-pair finder*

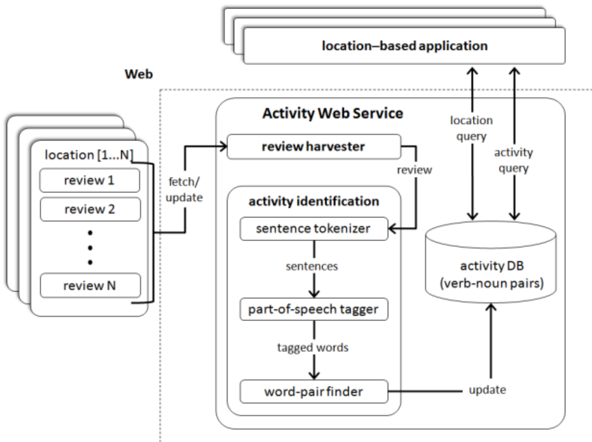

Figure 1. The system architecture and process of extracting the verb-noun pairs from the community authored reviews.

also records if the verb-noun pair is written in 1st (*i.e.,* I, we), 2nd (*i.e.,* you) or 3rd (*i.e.,* he, she) grammatical person.

Populate and Update

Finally, the *activity finder* converts each verb and noun to their respective base-word (*e.g.,* "purchases" is converted to "purchase") using WordNet [28] and archives the original verb-noun pair, the base-word verb-noun pair and the grammatical person in the *word-pair database*. The base-word allows for commonly identified verb-noun pairs to be quickly associated with one another. We also maintain a frequency count of repeated verb-noun pairs.

EXPERIMENTAL VALIDATION

The evaluation of the verb-noun pair dataset was accomplished through two formative studies where participants (with personal experience with the locations selected) are asked to:

- Provide activities that they have personally performed or experienced at one of the locations.
- Validate each of the 40 most common verb-noun pairs for each of 14 locations as actual activities supported by the respective location.

The combined list of activities provided by participants and the verb-noun pairs validated by participants serve as a list of *known* activities that are supported by a specific location. We acknowledge that this list of activities does not represent all of the activities supported by a location. However, identifying an exhaustive set of potential activities is a non-trivial task. Each new person may articulate a new or nuanced activity that has not been previously considered. Instead, our approach enables us to determine the rate of the true positive, false positive and false negative activities with respect to a list of the most common verb-noun pairs for a location. We define each as:

- True positive – a verb-noun pair that is validated (by a participant) as an activity supported by a location and is correctly identified as an activity for the location.
- False positive – a verb-noun pair that is rejected (by a participant) as an activity supported by a location and is incorrectly identified as an activity for the location.
- False negative – an activity that is supported by a location and not included in the list of most common verb-noun pairs for the location.

In turn, we are then able to determine the precision (defined as the number true positives divided by the number of true positives and false positives) and recall (defined as the number of true positives divided by the number of true positives and false negatives) for our system as it pertains to the 40 most common verb-noun pairs and the list of *known* activities as described by 59 users for 14 different locations. As previously mentioned, an evaluation of these rates across the entire data set is not possible because an exhaustive list of activities for a location is intractable.

Location Selection

The fourteen locations we selected for the evaluation are presented in Table 1. We selected these locations because they were the most heavily reviewed locations in the three most commonly reviewed Yelp categories [29]: restaurants (L1-4), arts & entertainment (L5-8) and retail (L9-14) combined represent 61% of Yelp reviews. In addition, we divided the shopping category into books/music (L9-11) and clothing (L12-14). The three geographical regions (New York, San Francisco and Los Angeles) were not selected intentionally, but rather occurred as an effect of selecting heavily reviewed locations for the chosen categories.

Procedure

We conducted two formative studies to collect and validate known activities for the 14 locations using separate online questionnaires. This approach allowed us to reach a large number of geographically distributed participants and meet our specific recruitment requirements (*i.e.,* a participant must have personally been to the location they are providing feedback on). Each participant completed only one questionnaire and provided responses for only one location.

Provide Activities Questionnaire

The purpose of this questionnaire is to identify activities that people familiar with the test locations have previously performed at these locations. In addition, the activities are articulated in their own terms. Participants completed an online questionnaire asking them to "*list at most ten activities–written as verb-noun pairs—that [they] have personally performed at the location*" with respect to one location. Each participant completed one questionnaire only. We recruited the participants for this evaluation by contacting members of Yelp who had written the most recent reviews for our fourteen locations. In total, we contacted 294 potential participants—21 participants for each location. Because members' email addresses are not publicly listed we contacted each member using the mail

Id	Location Name	City	# Reviews (n)	Yelp Categories
L1	Burma SuperStar	San Francisco	1797	Restaurant; Burmese
L2	Gary Danko	San Francisco	1616	Restaurant; American (New)
L3	Shake Shack	New York	1019	Restaurant; Burgers; Food Stands
L4	Pink's Hot Dogs	Los Angeles	1612	Restaurant; Hot Dogs
L5	California Academy of Science	San Francisco	937	Arts & Entertainment; Museums
L6	De Young Museum	San Francisco	666	Arts & Entertainment; Museums
L7	San Francisco Museum of Modern Art	San Francisco	567	Arts & Entertainment; Museums
L8	The Getty Center	Los Angeles	549	Arts & Entertainment; Museums
L9	Green Apple Books	San Francisco	708	Shopping; Music & DVD; Bookstores
L10	Amoeba Music	San Francisco	684	Shopping; Music & DVD
L11	Strand Book Store	New York	406	Shopping; Bookstores
L12	UNIQLO	New York	468	Shopping; Men's Clothing; Women's Clothing
L13	Century 21	New York	358	Shopping; Department Stores; Thrift Stores
L14	H&M	San Francisco	475	Shopping; Men's Clothing; Women's Clothing

Table 1. The 14 unique locations selected from Yelp to be evaluated. The locations span four different categories: restaurants (L1-4), museums (L5-8), shopping book/music (L9-11) and clothing (L12-14). The reviews were extracted October 18, 2009.

functionality built into Yelp's community interface. Restrictions imposed by Yelp limited us to contacting three members per day. Each email contained text inviting the person to participate in the study and a personalized URL to the survey. The personalized URL ensured that multiple submissions from a single participant could be identified and removed. The reviews from the participants who responded to the survey were excluded from our analysis. Of the 294 potential participants, 59 completed the survey, a 20% response rate: 34 female, 24 male, and 1 unknown. The age of respondents varied, with the majority between 18-29 years of age: 18-29 (38), 30-39 (12), 40-49 (3), 50-59 (4) and unreported (2).

Validate Verb-Noun Pairs Questionnaire

The purpose of this questionnaire is to determine which of the 40 most common verb-noun pairs for a location are true positives (valid activities) and false positives (verb-noun pairs the system returns as valid activities, but are in actuality not valid). Participants completed an online questionnaire asking them a single question – *"... indicate if the listed activities can be performed at [location name]"* The questionnaire was unique for each location and contained a list of the 40 most commonly identified verb-noun pairs (in decreasing order by instance count) with respect to the location. Participants were instructed to indicate if each verb-noun pair is a valid activity that *can* or *cannot* be performed at the respective location with the following instructions:

- can perform – *"You have performed, observed someone performing or believe the activity can be performed here"*
- cannot perform – *"The activity cannot be performed here or the activity as written does not make sense"*

We employed snowball sampling to recruit participants for this evaluation, choosing not to recruit participants from Yelp because of the complexity and length of the recruitment process. A recruitment email was sent to friends and colleagues who live in the three respective cities. The email invited them to participate by completing one or more of the questionnaires, but only for locations they have personal visited. We encouraged them to forward the email to other people they believed may be able to participate. In total, we directly emailed 20 individuals and 67 people eventually completed the survey: 44 female, 21 male, and 2 unknown. The age of respondents varied, with the majority between 18-29 years of age: 18-29 (33), 30-39 (23), 40-49 (6), 50-59 (2) and unreported (3).

Method of Analysis

As mentioned previously, the primary metrics of our analysis revolve around evaluating the true positives, false positives and false negatives with respect to the most commonly identified verb-noun pairs for a location, across the fourteen test locations. The process for identifying verb-noun pairs is liberal by design, but ensures there is significant diversity in the verb-noun pairs that are identified. However, the process also results in a large number of false positives. To increase precision and reduce the number of overall false positives within the verb-noun data set, we evaluated alternate methods of filtering the verb-noun pairs {*1^{st} person, f>1*} against no filtering {*no filter*}:

- *no filter* – the verb-noun pairs are not filtered; all of the verb-noun pairs are considered.
- *1^{st} person* – the verb-noun pairs are filtered to include only those written in the 1^{st} grammatical person.
- *$f>1$* – the verb-noun pairs are filtered to include only those with an instance count greater than one.

We emphasize verb-noun pairs written in the first grammatical person (*1^{st} person*) because they are potentially activities the author has performed. Third grammatical person implies the author is describing another person's activity and the subject is ambiguous for a second

Id	Location	pairs	Filter		
			no filter	1st person	f>1
L1	Burma SuperStar	16831	10197	5137	1871
L2	Gary Danko	20865	12767	6608	2322
L3	Shake Shack	9534	6413	3075	987
L4	Pink's Hot Dogs	14003	8688	4026	1392
L5	Cali. Academy of Sci.	14748	9997	4433	1525
L6	De Young Museum	7907	6197	2744	738
L7	San Francisco MOMA	5632	4492	2188	500
L8	The Getty Center	5173	3848	1382	503
L9	Green Apple Books	5670	3852	1773	472
L10	Amoeba Music	5373	4116	1968	470
L11	Strand Book Store	3628	2658	1194	308
L12	UNIQLO	3520	2788	1325	345
L13	Century 21	3516	2871	1160	306
L14	H&M	3788	3019	1510	354
	Avg.	8585	5850	2752	864

Table 2. The number of verb-noun pairs extracted from the Yelp review text. The values are for unique verb-noun pairs.

grammatical person. The verb-noun pairs for the data sets are sorted in descending order according to their frequency.

In addition to filtering the results, we also manipulate the method of comparing the known activities against the identified verb-noun pairs when searching for a match. When we compare the known activities provided by the participants in the *provide activities questionnaire* we first convert the verb and noun to their base word. Converting the activity to its base word ensured a fair comparison when matching the terms with the identified verb-noun pairs because the ambiguity of tenses and modifiers is removed. We use three different methods to compare the known activities with the identified verb-noun pairs:

- *Exact terms* – compare verb-nouns pairs using the base word for the terms.
- *Similar terms* – compare verb-noun pairs using statistically similar permutations of the base words [16].
- *Synonyms* – compare verb-noun pairs using synonymous permutations of the base words provided by WordNet.

The similar word lists are derived from Dekang Lin's online thesaurus where similarity is based on statistical

		Precision				Average Precision	
		no filter		1st person		no filter	1st person
		n	%	n	%	%	%
Validated		444	79.3	438	78.2	88.3	88.9
Provided	exact	32	5.7	29	5.2	23.2	24.7
	similar	66	11.8	62	11.1	26.7	26.7
	synonym	73	13.0	73	13.0	34.6	31.6

Table 3. The precision and average precision; averaged across the 14 locations.

dependency [17]. For each verb and noun, we selected at most 10 synonymous and similar terms.

Event-count measures (*e.g.,* number of identified verb-noun pairs) are analyzed using the nonparametric Kruskal-Wallis test to identify significant differences. Post-hoc pairwise comparisons are conducted using the Mann-Whitney U, controlling for Type 1 error using Holm-Bonferroni.

RESULTS

The process of extracting the verb-noun pairs from the Yelp review texts generated 120,188 verb-noun pairs for the 14 locations combined. The number of verb-noun pairs for each of the 14 locations is presented in Table 2. The participants of the *provide activities questionnaire* identified 446 activities for the 14 locations, an average of 31 activities per location.

Number of Community-Derived Verb-Noun Pairs

A Kruskal-Wallis test revealed a significant difference in the size of the verb-noun dataset for filter type, $\chi^2_{(2, N=42)} = 26.44$, $p<0.001$. As expected, the *f>1* and *1st person* filters significantly reduced the number of verb-noun pairs in the verb-noun dataset in comparison to *no filter*, both at $p<0.001$. In addition, the *1st person* filter significantly reduced the number of verb-noun pairs in the verb-noun dataset in comparison to the *f>1* filter, $p<0.001$.

Precision of the Most Common Verb-Noun Pairs

Participants of the *validate verb-noun pairs questionnaire* accepted, when using *no filter* and the *1st person* filter, the majority of the 40 most commonly identified verb-noun pairs as valid activities with respect to their locations. The precision averaged across the 14 locations is 79.3% (444/560) when using *no filter* and 78.2% (438/560) when using the *1st person* filter. The average precision measure which is sensitive to the ranked ordering of the 40 most common verb-noun pairs (Table 3) is 88.3% and 88.9% respectively. Close examination of the precision for each of the 40 verb-noun pairs—according to their ranked order—shows that (Figure 2) precision is highest for the 10 most commonly identified verb-noun pairs and decreases as we delve deeper into the dataset. No significant difference was observed for precision with respect to filter type.

Recall of the Most Common Verb-Noun Pairs

As we mentioned previously, the participant-provided activities and the validated verb-noun pairs can be merged to form a complete list of *known* activities for our locations. On average, there are 57 *known* activities for each location. We utilize this number of known activities as the denominator in our calculation of recall. However, given we limit our evaluation to the 40 most common verb-noun pairs and we have on average 57 *known* activities for a location, the maximum recall value we can achieve is 70.2%. The recall is 55.9% and 55.5% for no filter and 1st person filter respectively (Figure 2). No significant difference was observed for recall based on filter type.

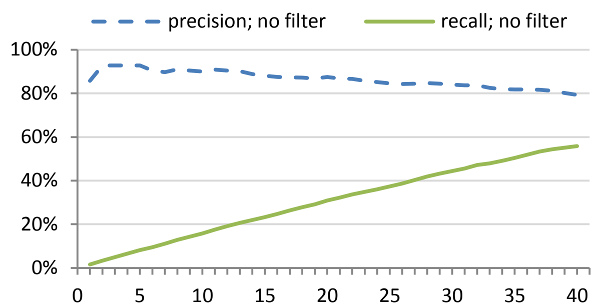

Figure 2. The mean precision and recall calculated at each position of the 40 most common verb-noun pairs, averaged across the 14 locations. We present only no filter because the results overlap significantly with the 1st person filter.

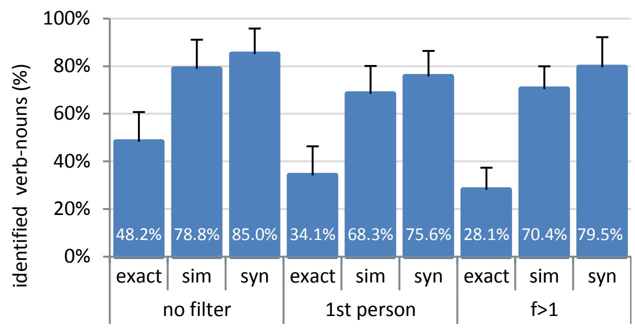

Figure 3. The percentage of participant-provided activities that are identified in the set of verb-noun pairs grouped by method of comparison and filter, averaged across the 14 locations. The error bars represent one standard deviation.

Identifying Participant-Provided Activities

Although we already know that the majority of the common verb-noun pairs are valid activities, we wanted to evaluate these verb-noun pairs with respect to their coverage of the activities that participants report having performed at the locations. As presented in Table 3, the precision when using the participant-provided activities (independent of filter type and method of comparison) is low. Of the 446 participant-provided activities, we are able to identify at most 73 (16.4%) when the comparison is conducted using synonymous terms; the number of false negatives is high at 373 (83.6%). However, it does not mean the activities do not exist with the complete set of verb-noun pairs.

When evaluating the dataset more deeply we are able to identify the majority of the participant-provided activities (Figure 3). A Kruskal-Wallis test revealed that the type of filter and the method used to compare terms has a significant impact on the number of participant-provided activities we can identify within the verb-noun dataset, $\chi^2_{(2,N=126)} = 8.40$, $p<0.05$ and $\chi^2_{(2,N=126)} = 79.12$, $p<0.001$ respectively. The number of participant-provided activities that we can identify deeper in the dataset is significantly higher when synonymous terms are used to compare terms, rather than similar ($p<0.005$) and exact terms ($p<0.001$). Additionally, similar terms resulted in a significantly higher match percentage than exact terms ($p<0.001$). The number of participant-provided activities we can identify deeper in the dataset is significantly higher when *no filter* is applied to the dataset than with the *1st person* ($p<0.01$) or *f>1* ($p<0.05$) filters. No significant difference was observed between the *1st person* and *f>1* ($p=0.886$) filters.

Number of Reviews Authored for a Location

The number of reviews authored for a location has a strong influence on the precision of the location's 40 most common verb-noun pairs (Figure 4) and the number of participant-provided activities that can be identified within the verb-noun dataset (Figure 5).

Using the complete set of *known* activities (the participant-provided activities plus the validated 40 most common verb-noun pairs) for the locations, we are able to achieve a mean precision of 29.5% for the 40 most common verb-noun pairs when processing only the first 50 reviews, increasing to 45.7% and 57.3% for the first 100 and 200 reviews, respectively. Similarly, we are able to identify up to 43.2% of the participant-provided activities (using synonyms to perform the term comparison) within the complete set of verb-noun pairs when processing only the first 50 reviews, increasing up to 60.1% and 69.8% for the first 100 and 200 reviews, respectively.

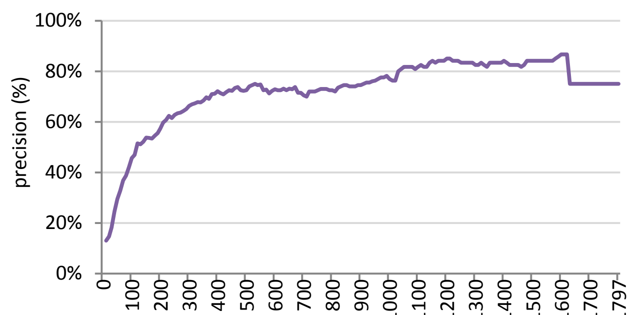

Figure 4. The mean precision of the 40 most common verb-noun pairs when varying the number of reviews processed, averaged across the 14 locations. X-axis is the number of processed reviews. The values are for no filter.

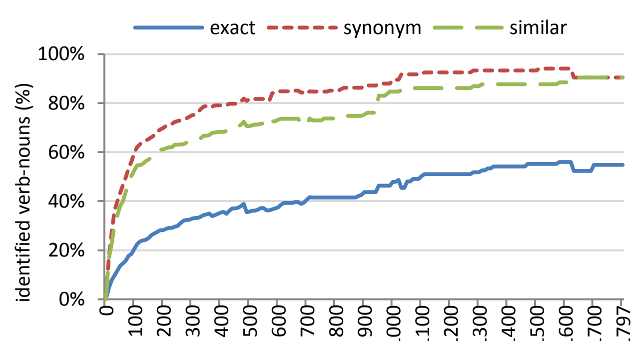

Figure 5. The percentage of participant-provided activities that are identified in the set of verb-noun pairs when varying the number of reviews processed, averaged across the 14 locations. X-axis is the number of review processed. The values are for no filter.

Regression analysis reveals that the growth of precision (R-sq: 0.886), exact (R-sq: 0.955), synonyms (R-sq: 0.956) and similar (R-sq: 0.969) can be modeled as a logarithmic function (Figure 4). With precision, we begin to see the logarithmic growth level out at approximately 400 reviews. Although sustained reviewing does little to improve precision and percentage of identified activities, it accentuates the frequency of common verb-noun pairs which helps to differentiate them from seemingly random verb-noun pairings.

Difference between Activities for Similar Locations

The fourteen locations we sampled (Table 1) are categorically different, encompassing restaurants (L1-4), museums (L5-8), and retail stores for books/music (L9-11) and clothing (L12-14). The majority (*n*=567; 60.1%) of the 40 most common verb-noun pairs that are valid for a location are different between locations within a category, $p<0.001$. For example, "see Picasso" is an activity that is identified only for the San Francisco Museum of Modern art, while "buy membership" is an activity that is identified for all three of the museums we sampled.

Types of Community-Derived Activities

An analysis of the 883 valid community-derived activities was completed using grounded theory affinity clustering [3, 10]. Two researchers performed an affinity analysis clustering similar responses. The process of clustering responses was repeated for three iterations until a consensus was reached on the three high-level activity categories (see Table 4). Each of the researchers then independently re-categorized all the activities using the three categories. The re-categorization resulted in an overall high level of observed agreement (0.98) and inter-rater reliability (Cohen's Kappa: 0.947, $p<0.001$).

The majority of activities (77.5%) we are able to identify are *physical activities* (Table 4: C1), characterized as a physical action or a purposeful interaction with a person or object. For example, when at a restaurant a person may "eat [a] hamburger" or when in a book store she may "buy [a] book". Although research on activity inference traditionally focuses on sensing the physical manifestations of an action, many of the activities we identified are *not* physically performed. A *cognitive activity* (C2: 11.4%) is entirely mental, involving an awareness and reaction to the environment or stimulus. For example, a person can go to the park to "enjoy life" or a museum to "appreciate art". Similarly, a *perceptual activity* (C3: 9.2%) pertains the person's senses. For example, a person might go to a museum to "watch people" or "view art". Using the community-authored reviews we can identify some of these alternative activities that are difficult to sense.

SUMMARY & DISCUSSION

Yelp's community-authored reviews offer a rich data source that can be processed to identify activities (articulated as verb-noun pairs) that are supported by a location, for a large number of locations. Participants confirmed that the majority of the 40 most commonly identified verb-noun pairs extracted from the community-authored reviews are valid activities. A mean precision of up to 79.3% and recall up to 55.9% was achieved across the 14 locations we evaluated. Although we employ a simple frequency analysis to identify the 40 most common verb-noun pairs for a location, the simplicity of our approach still highlights verb-noun pairs which are valid activities that characterize a location. In the future, we aim to employ more sophisticated methods of natural language processing to help us reduce the number of false positives within the complete set of verb-noun pairs and identify activities that are unique to a specific location in contrast to activities that are globally unique to a category of locations.

ID	Activity Category	Example	n	%
C1	Physical	*buy book*	685	77.6%
C2	Cognitive	*appreciate art*	101	11.4%
C3	Perceptual	*watch people*	81	9.2%
	Unclassified		16	1.8%

Table 4. Three categories of activities we identified when evaluating the 40 most common verb-noun pairs.

Our comparison of the participant-provided activities against the 40 most common verb-noun pairs resulted in a low mean precision across the test locations. At most, 73 of the 446 provided activities were identified. However, the disparity between the precision of the validated activities and the participant-provided activities (which we know are all actual activities) highlights the personal and individualized nature of these activities in contrast to the common verb-noun pairs that can be identified through the community authored reviews. A deeper examination of the complete set of verb-noun pairs revealed that on average up to 85.0% of the participant-provided activities can be identified with the complete set of verb-noun pairs, the majority of which (79.5%) have a frequency greater than 1.

The number of reviews authored for a location has a strong influence on precision. We were able to achieve precision up to 29.5% when averaged across the 14 locations when processing the first 50 reviews, increasing to 45.7% and 57.3% for the first 100 and 200 reviews respectively. In addition, we are able to identify on average up to 43.2% of the participant-provided activities when processing the first 50 reviews, increasing up to 60.1% and 69.8% for the first 100 and 200 reviews respectively. These results indicate that we can obtain significant precision within 200 reviews and that thousands of reviews do not necessarily diversify the set of valid verb-noun pairs, but would reinforce the validity of a valid verb-noun pair by increasing its relative frequency. Two-hundred reviews may seem significant, however reviews can be combined from multiple sources (*e.g.,* City Search, Kudzu, Insider Pages, Google Pages) to ensure the greatest number of reviews are considered for a location.

Figure 6. Activity Compass – a mobile application that characterizes the activities available in the user's vicinity.

The timeliness of an activity is an important feature that will influence the activities validity—we do not address this issue currently. The activities a location supports can change over time, meaning that some activities operate on a set schedule while others are available ad-hoc. Given that reviews are continuously being authored—more frequently for some locations than others—it may be possible to identify the temporal features or when an activity is no longer valid. For example, if an activity is mentioned frequently and contiguously across the reviews for a location, but suddenly ceases to be mentioned, this change may indicate a change in the validity of the activity for this location.

CONTEXT-AWARE SERVICES ENABLED

The activities we are able to identify for a location can support a breadth of context-aware applications. In this section, we present two such applications (Figure 6 & 7).

Activity Compass: Characterizing Supported Activities

Localized searches services (*e.g.,* Google Mobile [13]) assist users to discover nearby locations to perform their activities, but these services require the user know what activity she wants to locate. However, in some instances the user may first need to discover what activities her environment can support before she can identify a specific location. This is true for people exploring a new city or neighborhood, or the visually impaired who do not have visual cues to identify features offered by their environment. In support of this need, we are developing and evaluating a mobile application called Activity Compass (Figure 6) which characterizes the activities supported by the user's environment.

Activity Compass presents the user with common activities that characterize the forward direction of the device at three distance measures. The text size of the activity emphasizes the frequency with which the activity is mentioned within a fixed distance—larger means more locations support it. The presentation of activity, distance and direction can offer a new perspective on the affordances of a user's environment (new or familiar).

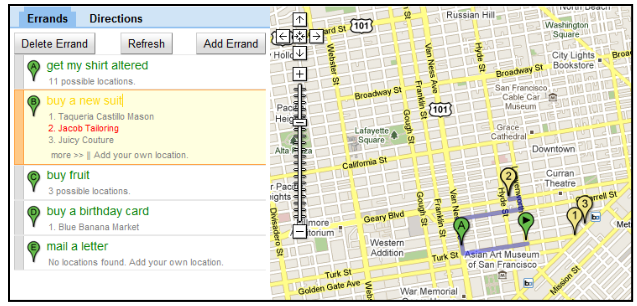

Figure 7. Better Errands – a web application that helps users discover nearby locations to perform their tasks and errands.

BetterErrands: Context-Aware Support for Running Errands

It is common for people to maintain a list of tasks or errands. Numerous web services (*e.g.,* Google Tasks, Remember the Milk) have been developed to assist users manage and maintain their lists. However, these services do not offer the ability to opportunistically discover locations where the user's tasks or errands can be performed. To address this limitation, we have developed Better Errands (Figure 7), a web-based service that synchronizes with Google Tasks. Better Errands queries the activity service for locations that support the activities in the user's Google Tasks list and presents these locations to the user. The locations are then reflected back in Google Tasks. A mobile interface to the errands service allows the user to opportunistically identify new plans for completing her errands based on her location.

CONCLUSION

In this paper, we validate the use of community-authored content, specifically location-based review authored on Yelp as a source to identify activities that are supported by the respective location. We show that the community-authored reviews provide a diverse and comprehensive data source and that the 40 most common verb-noun pairs identified from the reviews for a location achieve a mean precision of up to 79.3% and recall of up to 55.9%. Although we do not evaluate alternate location-based review communities, the method we propose can support alternate types of communities and can be used to aggregate verb-noun pairs identified across communities. Finally, we present two context-aware services that leverage location-based activity information on a city scale accessible through a Web service we developed supporting multiple cities in North America.

ACKNOWLEDGEMENTS

We thank Elaine Huang and Gillian Hayes for their feedback on early versions of this paper.

REFERENCES

1. Agre, P.E. 2001. Changing places: contexts of awareness in computing. *Human-Computer Interaction* 16, 2 (Dec. 2001), 177-192.

2. Bao, L. and Intille, S.S. 2004. Activity recognition from user-annotated acceleration data. In *Proc. of the Second International Conference on Pervasive Computing*. Pervasive '04. Springer-Verlag, Berlin, Heidelberg, 1-17.
3. Beyer, H. and Holtzblatt, K.. 1998. Contextual design: defining customer-centered systems. San Francisco, CA, USA: Morgan Kaufmann.
4. Buettner, M., Prasad, R., Phillipose, M, and Wetherall, D. 2009. Recognizing daily activities with RFID-based sensors. In *Proc. of the 11th International Conference of Ubiquitous Computing.* Ubicomp '09. ACM Press, New York, NY, 51-61.
5. Chang, K., Chen, M.Y. and Canny, J. 2007. Tracking free-weight exercises. In *Proc. of the 9th International Conference on Ubiquitous Computing.* Ubicomp '07. Springer-Verlag, Berlin, Heidelberg, 19-37.
6. Dey, A.K. 2001. Understanding and using context. *Personal and Ubiquitous Computing* 5, 1 (Feb. 2001), 4-7.
7. Dourish, P. Seeking a foundation for context-aware computing. *Human-Computer Interaction* 16, 2 (Dec. 2001), 229-241.
8. Fogarty, J., Au, C. and Hudson, S.E. 2006. Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. In *Proc. Of the 19th Annual ACM Symposium on User Interface Software Technology.* UIST '06, ACM Press, New York, NY, 91-100.
9. Froehlich, J., Larson, E., Campbell, T., Haggerty, C., Fogarty, J. and Patel, S.N. 2009. HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In *Proc. of the 11th International Conference of Ubiquitous Computing.* Ubicomp '09. ACM Press, New York, NY, 235-244.
10. Glaser, B.G. 1992. Basics of grounded theory analysis: emergence vs. forcing. Mill Valley, CA, USA: Sociology Press.
11. Hodges, M. and Pollack, M.E. 2007. An 'object-use fingerprint': The use of electronic sensors for human identification. In *Proc. of the 9th International Conference on Ubiquitous Computing.* Ubicomp '07. Springer-Verlag, Berlin, Heidelberg, 289-303.
12. Huynh, T., Fritz, M. and Schiele, B. 2008. Discovery of activity patterns using topic models. In *Proc. of the Tenth International Conference on Ubiquitous Computing.* Ubicomp '08. ACM Press, New York, NY, 10-19.
13. Kamvar, M., and Baluja, S. 2009. A large scale study of wireless search behavior: Google Mobile search. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '06. ACM Press, New York, NY, 701-709.
14. Lee, S-W. and Mase, K. 2002. Activity and location recognition using wearable sensors. *IEEE Pervasive Computing* 1, 3 (July 2002), 24-32.
15. Lester, J., Choudhury, T., Kern, N., Borriello, G. and Hannaford, B. 2005. A hybrid discriminative/generic approach for modeling human activities. In *Proc. of the International Joint Conference on Artificial Intelligent.* IJCAI '05, 766-772.
16. Lin, D. Automatic retrieval and clustering of similar words. In *Proc. of the 17th International Conference on Computational Linguistics.* COLING-ACL '98. ACL / Morgan Kaufmann, 768-774.
17. Lin, D. Dependency-based thesaurus. http://webdocs.cs.ualberta.ca/~lindek/demos.htm. Retrieved June 27, 2010.
18. Logan, B., Healey, J., Philipose, M., Tapia, E.M. and Intille, S. 2007. A long-term evaluation of sensing modalities for activity recognition. In *Proc. of the 9th International Conference on Ubiquitous Computing.* Ubicomp '07. Springer-Verlag, Berlin, Heidelberg, 483-500.
19. Mummidi, L.N. and Krumm, J. 2008. Discovering point of interest from users' map annotations. *GeoJournal* 72 (3-4), 215-227.
20. OpenNLP. http://opennlp.sourceforge.net. Retrieved June 27, 2010.
21. Patel, S.N., Reynolds, M.S. and Abowd, G.D. 2008. Detecting human movement by differential air pressure sensing in HVAC system ductwork: an exploration in infrastructure mediated sensing. In *Proc. of the Sixth International Conference on Pervasive Computing.* Pervasive '08. Springer-Verlag, Berlin, Heidelberg, 1-18.
22. Patel, S.N., Robertson, T., Kientz, J.A., Reynolds, M.S., Abowd, G.D. 2007. Flick of a switch: detecting and classifying unique electrical events on the residential power line. In *Proc. of the 9th International Conference on Ubiquitous Computing.* Ubicomp '07. Springer-Verlag, Berlin, Heidelberg, 271-288.
23. Patterson, D.J., Liao, L., Fox, D., Kautz, H. 2003. Inferring high-level behavior from low-level sensors. In *Proc. of the 5th International Conference on Ubiquitous Computing.* Ubicomp '03. Springer-Verlag, Berlin, Heidelberg, 73-89.
24. Philipose, M., Fishkin, K.P., Perkowitz, M., Patterson, D.J., Fox, D., Kautz, H. and Hahnel, D. 2004. Inferring activities from interactions with objects. *IEEE Pervasive Computing* 3, 4 (Oct. 2004), 50-57.
25. Stanford log-linear part-of-speech tagger. http://nlp.stanford.edu/software/tagger.shtml. Retrieved June 27, 2010.
26. Tapia, E.M., Intille, S.S. and Larkson, K. 2004. Activity recognition in the home using simple and ubiquitous sensors. In *Proc. of the 2nd International Conference on Pervasive Computing.* Pervasive '04. Springer-Verlag, Berlin, Heidelberg, 158-175.
27. WordNet. http://wordnet.princeton.edu/. Retrieved June 27, 2010.
28. Yahoo! Research Berkeley: Tag Maps. http://tagmaps.research.yahoo.com. Retrieved June 27, 2010.
29. Yelp. An Introduction to Yelp: metrics as of June 2009. http://www.yelp.com/html/pdf/press_yelp_snapshot_june_09.pdf