# A Language Approach to Modeling Human Behavior

## Peng-Wen Chen, Snehal Kumar Chennuru, Ying Zhang

Carnegie Mellon Silicon Valley
NASA Research Park, Bldg. 23, Moffet Field, CA 94035
{pengwen.chen, snehal.chennuru, joy.zhang}@sv.cmu.edu

### Abstract

The modeling of human behavior becomes more and more important due to the increasing popularity of context-aware computing and people-centric applications. Inspired by the principle of action-as-language, we propose that human ambulation behavior share similar properties as natural languages. In this paper, we use a Life Logger system to build the behavior language corpus. The behavior corpus shows Zipf's distribution over the frequency of vocabularies which is aligned with our "Behavior as Language" assumption. Our preliminary results of using smoothed $n$-gram language model for activity recognition achieved an average accuracy rate of 94% in distinguishing among basic behaviors including walking, running, and cycling. This behavior-as-language corpus will enable researchers to study higher level human behavior based on the syntactic and semantic analysis of the data.

## 1. Introduction

The advancement of sensor technology makes context-aware and people-centric computing more promising than ever before (Campbell et al., 2008). However, to effectively extract user-behavior context from the generally equipped sensors of normal mobile devices (e.g. mobile phones) still poses challenges for researchers and industry professionals. Behavior models can typically be used to recognize anomalous behavior as well as deviation from routine, or skipped steps for elders with onset dementia (Wilson and Atkeson, 2005). The state-of-the-art approach in behavior modeling uses the Hidden Markov Model (HMM) to recognize sensed activities as one of the pre-defined activities. HMM assumes the first order Markov chain in the state space and usually does not consider the inherent "grammar" or "structure" of the ambulatory behavior. The activities that can be recognized by the HMM are limited to those pre-defined in the training data which also limits the application of behavior modeling in people-centric computing.

In this paper, we present anl approach to modeling human behavior as language and describe our works in building the "behavior language corpus". We begin the paper by introducing the principle of "behavior-as-language". We then present our Life Logger system[1] which is used to collect the data to build the corpus and the evaluation tool for behavior recognition task. Then we present our preliminary results on activity recognition using our initial behavior-language data. Finally, we present our findings and discuss future research in behavior language modeling.

## 2. Behavior as Language

The similarity between human behavior and language had been articulated by Burke (1966) and Wertsh (1998). Based on the "principle of language as action", natural language and human action are really the same thing. They are both "mediational means" or tools by which we achieve our ends. They exhibit structure and satisfy "grammars".

---

[1] http://www.lifelogger.info

Table 1 illustrates that ambulatory behavior shares a lot in common with natural languages at all levels. The anatomy of human bodies allows us to perform certain *atomic movements* such as "turn upper body left" where as "jump up at 10g acceleration" is not possible. Such atomic movements form the vocabulary of the behavior language. A sequence of atomic movements performed in meaningful order creates a *movement* such as an *action* of "standing up". *Actions* such as "climbing up stairs" are created by performing actions in a right order similar to create a "sentence". A sequence of actions builds up an *activity*. Higher level behavioral concept *event* is composed of a series of activities in a similar way as a *document*.

In this paper, we focus mainly on the ambulatory behavior such as "walking" and "climbing upstairs and walking into my office." We use 3-axis accelerometers to record users' motion. The accelerometer measures the acceleration at the X, Y, Z direction at the time of sampling. For the built-in accelerometers used in our experiments, the raw readings for each axis ranges from -360 to 360 which translates into 373,248,000 different $(a_x, a_y, a_z)$ combinations. Based on our assumption that we can only make certain atomic movements due to the anatomy, we quantize the raw accelerometer reading into $V$ groups using K-Means clustering algorithm. Once the K-means clustering algorithm converges, it results in $V$ cluster centroids and we give each cluster a label such as "D", "GC" and "DFR". We can then convert all the training and testing accelerometer data to their nearest cluster's label and thus convert the ambulatory behavior into "behavior text" (Figure 1). Figure 2 shows the process of modeling the ambulatory behavior as language.

To empirically evaluate the similarity between the ambulatory behavior and natural languages, we check if the behavior language corpus follows the Zipf's law. Zipf's law states that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table. In other words,

| Natural Language | Behavior Language | Example |
|---|---|---|
| Word | Atomic Movement | Turn upper body left |
| Phrase | Movement | Stand up |
| Sentence | Action | Climb up stairs |
| Paragraph | Activity | Enter building, climb up stairs and walk into office |
| Document | Event | Left home and ride bicycle to campus arrived at my office at 2nd floor |

Table 1: Behavior as language at different levels.

[31,271,37] [37,281,42] [37,276,47] [42,271,47] [42,266,53] [58,271,47] [53,271,47] [74,271,42] …

⇓

CZ DG GI FK C BI CS DC HQ BX FI FI BX FI O …
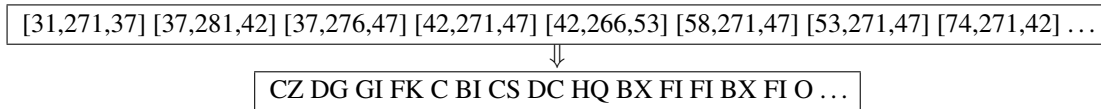
Figure 1: An example of converting accelerometer readings to behavior language.
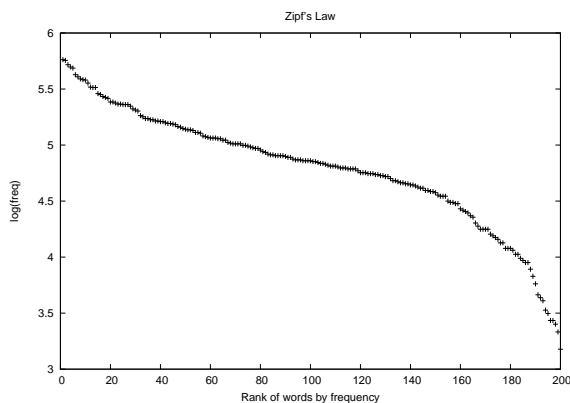


Figure 3: log(freq) vs. rank of frequency of word types in behavior language corpus.



Figure 4: Sensor helmet for collecting behavior data.

the logarithm of a word's frequency is linear to its rank in a natural language corpus. Figure 3 plots the logarithm frequency of word types in the behavior language corpus for word type ranked 1, 2, 3, etc. Though not exactly linear, it does plot a line similar to Zipf's distribution.

## 3. Building Behavior Language Corpus

Recording accelerometer data from a user's activity is trivial. The challenge is to annotate the accelerometer data for follow-up applications. Traditional experiment designs either require the presence of observers who annotate user's activity on the side or ask volunteers to annotate the collected data by themselves. The former has a drawback of higher man-cost whereas the latter faces the risks of inaccurate annotations.

To address these issues, we developed a Life Logger system which eliminates the observer role and helps users better annotate their collected data. The main idea behind this system is that we record as many types of sensory data as possible including GPS coordinates for outdoor locations, gyroscope for rotation, microphone for sound, camera view finder for pictures, and WiFi signal strength

for indoor locations. These sensory data are all recorded with their corresponding timestamps.

We mount a Nokia N95 mobile phone on a helmet (Figure 4) to collect and transmit sensory data to the Life Logger server. When displayed, all sensory data is aligned according to their timestamps. Only data comes from accelerometers and gyroscopes would be used to build the behavior language models and all the other sensory data serve as complimentary information which helps volunteer recall what he / she was doing in the past and therefore increases the accuracy of annotation.

The Life Logger server is responsible for storing, preprocessing, and providing user interface for annotating the collected data. Figure 5 shows the architecture of the Life Logger server. The server was implemented as a Ruby on Rails web application. The user can access the Life Logger system easily through their web browsers without the hassle of installing additional desktop applications. To help volunteers recall their past activities, we fit three types of sensory data - images, audio, and GPS locations - into one screen (Figure 6) to let users easily combine these memory clues. Moreover, since all sensory data is associated with synchronized timestamps, users can

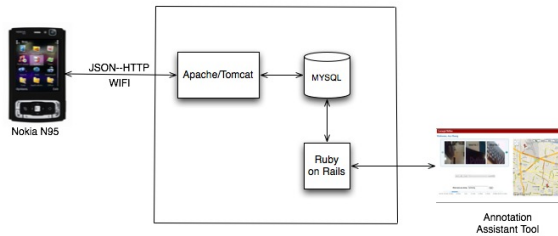Figure 2: Generating behavior language data.



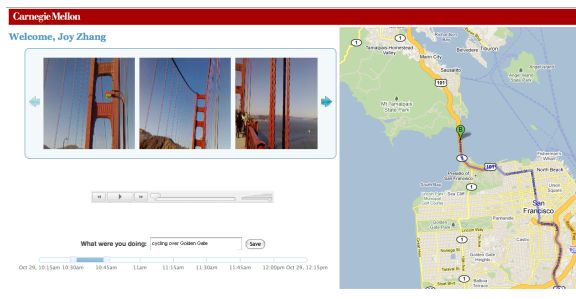Figure 5: Life Logging system architecture.



Figure 6: Annotation interface of behavior corpus.

| Configuration | Value |
|---|---|
| Activity Type | Walking, jogging, cycling |
| Phone position | Head, waist, both |
| Vocabulary size | 100, 150, 200 |
| Max. Sent. Length | 5, 15, 25 |
| $n$-gram order | 2, 3, 4, 5, 6 |
| Smoothing | Good Turing, Witten Bell |

Table 2: Configuration of the experiments.

| | Predicted Activity | | |
|---|---|---|---|
| | walking | running | cycling |
| walking | 94% | 3% | 3% |
| running | 6% | 92% | 2% |
| cycling | 8% | 0% | 92% |

Table 3: Classification accuracy on corpus with vocabulary=100.

of $t$ being generated by activity $a_i$ and predicts the activity of the testing data to be $i^*$ such that

$$i^* = \arg\max_i P(t|a_i) \qquad (1)$$

The pit-fall of using language model for activity recognition is that language model probabilities are not directly comparable if their training data has different vocabulary size. To by-pass this problem, each training data is augmented with the vocabulary list from all training data. Thus, all language models have the same vocabulary size and the probabilities are comparable.

The experiments were conducted using various configurations to better understand the strengths and weaknesses of our approach. Table 2 lists configuration of different features used in experiments.

We built the behavior language models for each possible combination of these settings to answer the following questions:

- How much does the phone position affect the quality of training data?

- Is there a significant difference in recognition accuracy between general or user-specific language models?

- What is the set of parameters that generate language models with the highest recognition accuracy?

Table 3 and 4 compare the recognition accuracy of language models trained over a corpus of vocabulary size 100 vs. the one with 200 word types. With larger

navigate through the data set by simply drag-n-drop the timeline at the left-bottom corner and all the three types of data would be updated simultaneously. Finally, volunteers annotate the data by selecting a range on the timeline and provide a text description.

We have collected 10 hours of data from two volunteers by the time of this paper submission. We will keep improving the volume and quality of the behavior language corpus by collecting more data using our specialized sensor helmet.

## 4. Behavior Recognition by Language Modeling

With the collected and annotated behavior language data, we validate the assumption that behavior can be modeled as language through the behavior recognition experiments. Most of the gesture and activity modeling methods we see today are based on Hidden Markov Models (HMM). HMM classifies the input sensing information into one of the predefined activities such as walking, running, and standing. HMM approaches are limited to recognizing predefined activities using first-order Markov models. In our approach, we view the labeled data for each activity $a_i$ as the training corpus and train a smoothed $n$-gram language model over the converted behavior language text. For each testing data $t$, we use each language model to calculate the probability

|         | Predicted Activity | | |
|---------|---------|---------|---------|
|         | walking | running | cycling |
| walking | 95%     | 1%      | 4%      |
| running | 4%      | 94%     | 2%      |
| cycling | 2%      | 0%      | 98%     |

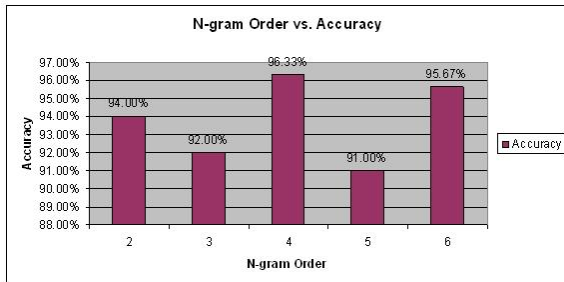Table 4: Classification accuracy on corpus with vocabulary=200.



Figure 7: Recognition accuracy vs. $n$-gram order.

vocabulary size, i.e., more atomic movement types, the behavior language text has more discriminative power to differentiate activities.

Figure 7 shows the average activity recognition accuracy vs. the order of $n$ in language model training. Overall, for this simple activity recognition task, the order of history does not play a significant role here.

## 5. Related Work

Several approaches had been used to distinguish basic human behavior. They can be categorized into two flavors: heuristic threshold-based classifiers and pattern recognition techniques such as decision trees, nearest neighbor, Naive Bayes, support vector machines (SVM), neural networks, and Gaussian mixture models (Nguyen et al., 2007). For distinguishing high-level human behavior, several attempts had been made in (Aipperspach et al., 2006; Patterson et al., 2003).

The MyLifeBits system (Gemmell et al., 2002) is designed to store and manage a lifetime's worth of everything that can be digitized. MyLifeBits supports capture, storage, management and retrieval of many media types, and logs as much usage data as possible. The missing technology in MyLifeBits work is *content analysis*. The system can only search text information that has been added by users. For other media types such as video and audio, MyLifeBits does not know how to represent the "meaning" of the video and audio data. As discussed in this paper, such representation is key for content analysis.

## 6. Conclusion and Future Work

In this paper, we present an approach of modeling ambulatory behavior as language. We verify the similarity between behavior and language by demonstrating Zipf's distribution over our behavior language corpus. The experimental

results presented in Section 4. demonstrate high accuracy of using language models for human behavior recognition. Unlike the traditional HMM approach which is limited to behavior recognition task, modeling behavior as language enables many other applications such as behavior clustering, daily events summarizing, and behavior semantic analysis. Our future work includes 1) continuing extending our behavior language corpus database, 2) using our approach to recognizing high-level human behavior, and 3) finding more evidence on the similarity between behavior and language including unsupervised grammar induction from the behavior language.

## 7. References

Ryan Aipperspach, Elliot Cohen, and John Canny. 2006. Modeling human behavior from simple sensors in the home. In *Proceedings of IEEE Conf. on Pervasive Computing*, pages 337–348, Dublin, Ireland, April.

Kenneth Burke. 1966. *Language as Symbolic Action*. University of California Press.

Andrew T. Campbell, Shane B. Eisenman, Nicholas D. Lane, Emiliano Miluzzo, Ronald A. Peterson, Hong Lu, Xiao Zheng, Mirco Musolesi, Kristóf Fodor, and Gahng-Seop Ahn. 2008. The rise of people-centric sensing. *IEEE Internet Computing*, 12(4):12–21.

Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong. 2002. Mylifebits: fulfilling the memex vision. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 235–238, New York, NY, USA. ACM.

A. Nguyen, D. Moore, and I. McCowan. 2007. Unsupervised clustering of free-living human activities using ambulatory accelerometry. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 4895–4898, Lyon, France, Aug 22-26.

Donald Patterson, Lin Liao, Dieter Fox, and Henry Kautz. 2003. Inferring high-level behavior from low-level sensors. In *Proceedings of the Fifth International Conference on Ubiquitous Computing (UbiComp 2003)*, pages 73–89, Seattle, Washington, October 12-15.

James V. Wertsch. 1998. *Mind As Action*. Oxford University Press, USA.

D. Wilson and C. Atkeson. 2005. Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In *Proceedings of he 3rd International Conference on Pervasive Computing*, Germany, May 8-13.