Intelligent System for Locating, Labeling, and Logging ($\boldsymbol{ISL^3}$)

Yan Huang[1], Terry Griffin[2], and Salomon Lompo[3]

[1] University of North Texas 1155 Union Circle #311277 Denton, Texas 76203 USA
[2] Midwestern State University, 3410 Taft Blvd. Wichita Falls, TX 76308 USA
[3] Midwestern State University, 3410 Taft Blvd. Wichita Falls, TX 76308 USA

Abstract. Suspendisse erat justo, rutrum at, pulvinar non, tristique nec, tellus. Sed vestibulum interdum massa. Pellentesque facilisis arcu pharetra libero. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Cras dapibus, felis eget laoreet mollis, magna sapien lacinia neque, euismod sagittis nisi mi sit amet sem. Mauris metus urna, imperdiet sed, luctus ut, molestie id, est. Donec eu eros. Praesent vitae sem vel ipsum congue vulputate. Sed at diam. Cras convallis. Nulla non dui. Proin sagittis. Ut sapien. Suspendisse pede. Integer mauris. Phasellus tellus. Mauris sed arcu eu lorem laoreet venenatis.

1 Introduction

As mobile smart devices become ubiquitous in our society, users will need and depend more and more on location based information. A computer model that is able to predict a user's purpose for being at a specific location will make location based services more effective by providing personalized information to users. The implementation of such a location prediction model requires a classified and correct set of data locations collected from several users to tune the prediction model to an acceptable level of accuracy. Inaccurately labeled or incomplete data sets will badly affect the performance of a location prediction model. This paper proposes an application software solution for data collection and location labeling called: ISL^3 - *Intelligent System for Locating, Labeling, and Logging.*

A large collection of data will allow researchers to derive the parameters required to train most probabilistic models and also get a trend of users behaviors in general. In [4] the author was able to compute the likelihood of the user's next destination using the data collected from 200 users in the Seattle Washington area [5]. One key component of any prediction model is a users history. A users history holds trends and patterns necessary for the prediction models algorithm to learn from. And the veracity of a collected data set will significantly affect ones ability to ascertain a users history which in turn will impact the overall performance of a prediction model. In addition, researchers tend not to share these types of sensitive data sets, typically due to privacy issues. Therefore, it becomes necessary for researchers to have a convenient and correct tool to ease this burden of data collection. In recent history, many researchers have used

GPS devices to collect a user' history. But the labeling (or classification) of this history was done by hand in a highly interactive manner. The contribution of this tool is in that it allows researchers in many fields (location prediction, machine learning, data mining, etc.) the means to easily collect data that has a correct labeling of a users behavior in a manner that requires little interaction.

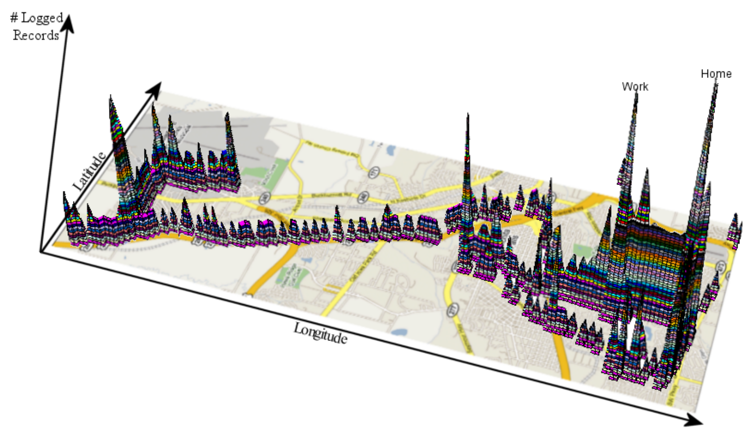

Fig. 1. Wichita Falls Area Visualization

2 Related Work

In spite of the important part data collection plays in the implementation process of any location prediction model, there have been few papers focusing on data collection and the different methods of labeling / classifying this data. In [3], data was collected for approximately 6 weeks for one user, to be used in a decision tree based prediction algorithm. Although this algorithm performed quite accurately, it's extensibility towards a general population comes into question. Ashbrook and Staner [1] collected data from one user for four months and six users for seven months in two different surveys. With those data sets the authors were able to extract significant locations to aid in the creation of location prediction algorithms. After each algorithm was trained and considered a viable prediction model, another round of data collection was conducted in a different area (Zurich) with more users but a much smaller training period. Another technical report [8] from the University of Helsinki gathered a dataset of mobile communication of a small number of people over a long period of time. The dataset obtained from this report was the central data used to test an algorithm developed in [7] and [6]. In [2] Hariharam and Toyama provide a data structure for analyzing and generating user histories from collected data. They also propose an algorithm that will extract interesting information from the raw data collected.

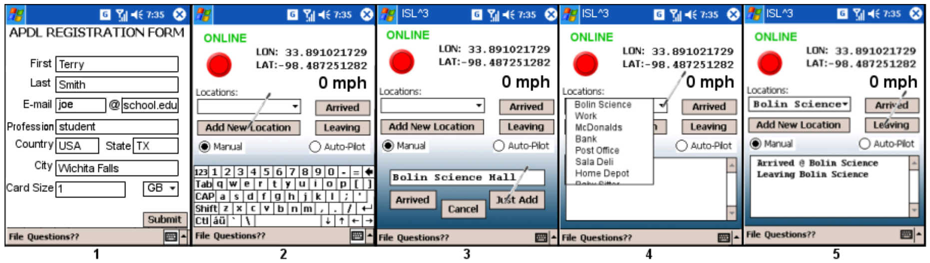

Fig. 2. ISL^3 Graphical User Interface

3 Software Development Process Components

Notation We will be using the following notation when referring to GPS location data. Raw GPS location data consist of a Lat, Lon pair, ℓ and a timestamp, τ .

The Graphical User Interface. ISL^3 displays an easy to use and intuitive interface for users. This plays into the overall purpose of the software itself, which is to give the everyday user the ability to easily collect a complete and fully labeled travel diary. To start the program the user clicks on the "play" button. As soon as the GPS enters in contact with the array of GPS satellites, the longitude and latitude of the current location are displayed at the right upper corner (see figure 2). The concept is simple: 1)Add locations, 2)Record arrivals and departures.

To add a location, the user simply presses the "Add New Location" button (see figure 2-2), bringing the keyboard into focus allowing the user to type a short descriptive name for a location. This adds the location to a "Locations" file that contains a unique id for the location and user, along with the coordinates of that location. This file populates a drop down list that is accessible on the main screen (see figure 2-4). This drop down list is continuously sorted, where the top of the list is the location closest to a users current location.

When the user needs to record arrivals and departures, they simply press the *Arrived* or *Leaving* button depending on the circumstance. This action is recorded in the Labels file along with a specified location. Specifying a location consists of selecting it from the drop down list, or by default, the top item. The logged record includes: Action (arrived or leaving), Location ID, User ID, Longitude and Latitude, as well as visual assurance in the form of a log entry on the users screen (see figure 2-5).

Initially the interaction by the user is relatively low, but will continuously diminish as the users typical locations are entered. Typical location entry consists of adding said location as it is encountered (meaning a record would be logged in the Labels file), but the software does have an option to "Just Add" a location. This allows a user to populate the list with their typical locations before they are visited. The one drawback to this method of location entry is that the location will have no Lat-Lon pair associated with it until the first time it is physically visited. Consequentially that location would not bubble to the top of the list as the user approaches it (for reasons explained earlier). ISL^3 has two modes of operation: 1)Standard and 2)Autopilot. Standard mode operates as previously discussed. Autopilot mode requires slightly less interaction.

Autopilot mode works much like a GPS logger in the fact that it senses motion. If the "Autopilot" senses that the user is not moving

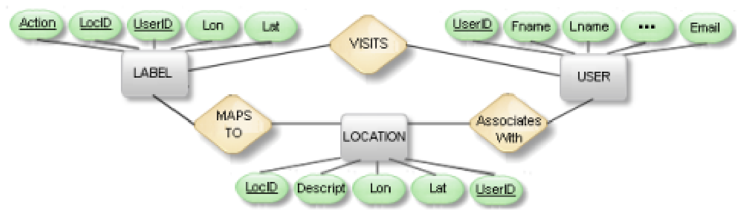

Fig. 3. ISL^3 ER-Diagram

Through testing it was decided to add a speedometer to give users additional reassurance that the device was functional and logging. In addition, the closest recorded location is always displayed on top of the location list at show in 2.

3.1 Architecture of the ISL^3 Software

The architecture of the ISL^3 software is shown in figure 4. When the handheld device receives a GPS stream from one of the communication ports, the APDL application scans through the ports of the handheld device to get the correct port receiving the GPS stream. To scan those ports the software uses the .Net framework library in order to remain in managed mode. The managed mode option guaranties a stable system from windows mobile operating prospective. After ISL^3 detects the correct port, the software sends the GPS stream to a GPS library. The main job of the GPS library is to parse the GPS stream. ISL^3 finally selects the needed information from the GPS library, creates the necessary files for classification and finally saves the information in the appropriate file. For ISL^3 to work properly it requires a configuration file to start with.

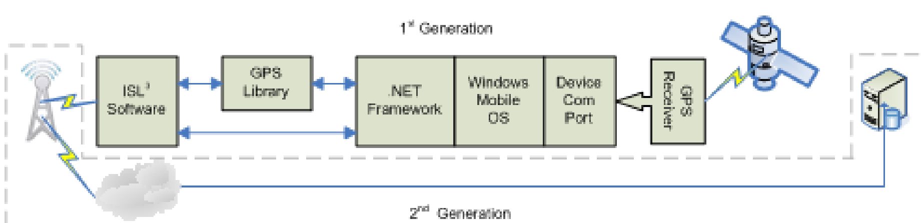

Fig. 4. ISL^3 Architecture

The Configuration File. The ISL^3 software allows the user to setup some key parameters such as the frequency of recorded data or the maximum number of locations allowed. When the software starts, it will load those parameters from a configuration file. A sample of the ISL^3 configuration file is showed in figure 2. The configuration file allows ISL^3 to be more flexible and robust.

Hardware Used. ISL^3 is designed to work on .Net Compact framework 2.0 architecture or above. The application was tested using two type of Pocket PC handheld device. The first handheld was the T610 model from the AIRIS manufacturer. The AIRIS T610 has an integrated GPS chip embedded in the device. The second Pocket PC is a WAYPOINT PDA coupled with an external GPS device via a Bluetooth connection. Both devices were equipped with windows mobile 2005 operating system. The WAYPOINT PDA had shown a better battery life saving and a quick response in getting satellite connection. However after a long period of inactivity, the Bluetooth connection was frequently lost which force the user to restart the ISL^3 program. This can be annoying to the user when set in auto-pilot mode. The Integrated GPS enabled device is suitable for the auto-pilot mode. During the testing phase, we setup the AIRIS T610 for a week in a car on auto-pilot mode without touching anything and the data collection was done.

3.2 Files Generated

ISL^3 produces four different files. Three of these files are used for the classification functions itself and one file is used to gather information from the user. The division of stored information was designed in such a way as to ease the extraction of important information from the collected data set, along with reducing storage requirements for the device in use.

The User Info File. The *User Info* file contains information specific to the volunteer. Most of the information requested is optional except for the user's e-mail address. The purpose of this file is simply to allow a particular data set to be paired with an individual. The email address is used as a a unique identifier when the volunteer uploads data to the central server. Some of the questions asked, if answered, will be used in future work in developing a real time location prediction model based on heuristic reasoning (discussed further in the future works section). The *User Info* file is a one time cost during the ISL^3 setup.

The Location File. The *Location File* holds data in the following format: (int ID, string Description, float Lon, float Lat). The ID is simply an auto-incremented unique identifier for a location and the Description is limited to 64 characters. The latitude and longitude stored are used by the software to determine the users current distance from any saved location, and continuously update a list of locations keeping the closest saved location at the top of the list. This allows the user to easily choose .... This files purpose was simply to avoid data redundancy and save memory, by not repeating location information with each logged entry.

The Activity File. The activity file records the departure and arrivals of the user during the day. An activity file is generated every day. For every trip the activity file saves the identification number of the departure location with a departure time stamp. When the user arrives at destination, ISL^3 saves also the ID number of the arrival location and a time stamp associated with it. The main purpose of the activity file is to help researchers validate the result of their prediction model during training.

The Log File. The log file contains the raw GPS stream needed to test a location prediction model. The log file does not save the all GPS stream but only little information is needed to derive most of the other data. The key data are the longitude, latitude, time stamp and the number of satellites. The number of satellite is used to measure the strength of the signal which may help explain certain gaps between recordings. The frequency at which ISL^3 saves GPS stream in the log file is set in the configuration file. Only one log file is generated per day.

4 Future Work

In the near future we plan to extend ISL^3 to most handheld device operating systems available and move it toward the Smartphone area. Our aim is to target more volunteered user to record data of their city and a setup a huge database of classified location data. Another expectation of the ISL^3 software is to reduce to the maximum possible the participation of user in the recording process. For that some predefined location name could be available for the user to download while setting up ISL^3. Therefore the users could already select places they usually go and saves it in the location file. Finally the user privacy is a serious issue. In order to protect our users some type of cryptography can be used. For instance instead of recording the user's house coordinate particularly, we assign the same GPS coordinate to a group of house within the same area.

References

1. Ashbrook, D., T. Staner: Using GPS to Learn Significant Location and Predict Movement Across Multiple Users. Personal and Ubiquitous Computing, 7(5): pp. 275–286 (2003)
2. Hariharam R., Toyama K. Project Lachesis: parsing and Modeling Location Histories. M.J. Egenhofer, C. Freksa, and H.J. Miller (Eds.): GIScience,LNCS 3234, pp. 106-124 (2004)

3. Huang Y., Griffin T.: A Decision Tree Classification Model to Automate Trip Purpose Derivation. Computer Applications in Industry and Engineering (Proceedings), pp. 44–49 (2005)
4. Krumm, J.: Real Time Destination Prediction Based On Efficient Routes. Microsoft Research (2006)
5. Krumm, J., E. Horvitz: The Microsoft Multiperson Location Survey (MSR -TR-2005-103). Microsoft Research (2005)
6. Laasonen K.: Clustering and Prediction of Mobile User Routes from Cellular Data. PKDD,LNCS 3721, pp. 569–576 (2005)
7. Laasonen, K. Raenato, M. Toivonen, H.: On-device Adaptive Location Recognition. A. Ferscha and F. Mattern (Eds.): Pervasive 2004, LNCS 3001, pp. 287-304 (2004)
8. Raenato, M.: Mobile Communication and Context Dataset. Technical report. University of Helsinki (2004)
9. Cheng, C., R. Jain, and E.v.d. Berg: Location Prediction Algorithms for Mobile Wireless Systems. Wireless Internet Handbook: Technologies, Standards, and Applications. CRC Press: Boca Raton, FL, USA. p. 245–263 (2003)