Behaviour Detection Using Bluetooth Proximity Data

Muhammad Awais Azam [1, 2], Laurissa Tokarchuk [1]
[1]Queen Mary University of London, School of Electronic Engineering and Computer Science, UK
[2] University of Engineering and Technology Taxila, Pakistan
muhammad.azam@elec.qmul.ac.uk, laurissa.tokarchuk@elec.qmul.ac.uk

Abstract

The abundance of Bluetooth enabled devices used in daily life has created new ways to analyze and model the behaviour of individuals. Bluetooth integrated into mobile handsets can be used as an efficient short range sensor. In this paper we use Bluetooth proximity data to detect the repeated patterns and behaviour of an individual by using n-gram technique. The primary purpose of this study is to determine what kinds of repeated behaviours can be detected in Bluetooth data. These repeated behaviours can show the daily, monthly or yearly patterns in an individual's life that can help us in determining the complex and unusual routines of human behaviour.

1. Introduction

Human behaviours are complex in nature and it is a challenging task to predict and learn from daily life activities. Modelling human behaviour such as individual routines from proximity data and relations with gathered data of daily life activity patterns is an emerging realm in ubiquitous computing [5, 6, 12]. Different sensing technologies such as Radio Frequency Identification (RFID), motion sensors and tracking devices are being used to get the real time proximity and daily life data. In particular, devices like mobile phones provide a rich platform for diverse data gathering. They are able to sense both geographic location and proximity information using GPS transceivers, digital cameras, motion sensors, microphones, Bluetooth ID, wireless access points and cell tower ID. Bluetooth radios are frequently incorporated in to the mobile devices [9]. Nearly 2.4 billion Bluetooth enabled equipment will be dispatched worldwide by 2013 [3].

Researchers have analyzed complex behaviours and activities by using different indoor sensor devices such as accelerometers, digital cameras and microphones. Frameworks to identify close proximity social behaviours [18], group actions in meetings [10] and audio visual perception of a lecture in smart environment [17] are discussed in these studies. Problems with the indoor sensor devices are that most of these devices are fixed and thus only useful when activities take place in their proximity.

The large penetration of Bluetooth enabled devices enables the use of these devices as personal identifiers. Various researchers have explored this hypothesis by using the mobile phone as a sensing device. Mobile phone, now a days is an essential part of society and almost everyone carries it. The advantage of mobile phone is that people always try to keep it with them and there is no need to add extra hardware with it. They have explored social proximity sensing [2, 15], tracking and positioning [8, 7], social behavioural modelling and routine classification [5, 6, 9, 12]. The significance of these studies is that they have identified new techniques and ways to detect an individual's behavioural patterns and daily life activities. In [5] [6], hierarchical Bayesian topic models like Latent Dirichlet Analysis (LDA) and Author Topic Model (ATM) are used for routine classification, they also present a framework for daily life activity recognition based on the user's location and group affiliation. Other researchers use Bluetooth proximity data with addition of cell tower ID's or other sensor data to classify the groups of users or simply to detect proximate devices and time spent with other devices [9].

The basic aim of this research is to detect the daily life activity patterns and individual behaviours in order to aid in the detection of abnormal behaviour such as wandering behaviour, a behavioural disorder in dementia patients. This behaviour will be detected by only considering the record of Bluetooth proximities because of easy and economical availability of Bluetooth enabled devices such as mobile phone as sensing device. Repeatable behaviour can be detected on the basis of repeated patterns of Bluetooth

device proximities. The repeating data is modelled as n-grams, technique to get the repeated patterns in the data.

The results presented here use the reality mining dataset [12] collected at MIT for the year 2004-2005. Nokia 6600 cell phones were used to record the data of 100 users over the duration of 9 months. This research uses one month Bluetooth proximity data of one user in order to show that the repeated behaviours can be detected using daily Bluetooth proximities. This paper presents the techniques and analysis of data that shows that there can be different repeated sequences depending on an individuals work routine. Also the time duration and specific hours of Bluetooth interactions is also analysed. This gives the percentage of repeated patterns with respect to time through out the whole month.
The rest of the paper is as follows: Section 2 contains the related work on daily life activity recognition and using Bluetooth as a sensing device; Section 3 explains the methodology of our analysis that we adopted to get the results. Section 4 discusses the results and Section 5 contains the conclusions of this study and direction planned for our future work.

2. Related Work

Human behaviour is very complex and challenging task for researchers since ages. Recently, with the improvements in network systems and information technology, people have done different work on detecting the behaviours and activities of humans. Most research has been about the activities inside the home. This is because most the majority of work has used sensing devices that either have short range of detection, less battery power and storage, or not very common that every person can use it without adding extra hardware, which is not possible for the scenarios outside the home. Naeem et al. work on activity recognition of dementia patients inside the home [13]; they describe an approach for modelling and detecting activities of daily life based on a hierarchy of plans that contain a range of precedence relationships, representations of concurrency and other temporal relationships. Identification of activities of daily life is achieved by episode recovery models supported by using relationships expressed in the plans. In [14] the same researchers define a two tier approach that is used to detect the activities of daily life. The higher tier uses hidden markov modelling (HMM) to model the common goals and sub-goals associated with daily life activities while the lower tier is used to recognize the tasks from different sensor events. Paul L et al. recognize the activities of people working in a workshop by using motion sensors and sound sensors [16]. Motion data was captured with accelerometers and then analyzed using time series recognition techniques, i.e. HMM. Sound data was collected by using a microphone and LDA (Linear discriminate Analysis) used to partition the data which is further processed by a larger majority decision window to generate a single result for sound recognition. Both motion and sensor data are then combined to recognize the activity. This includes three steps: extracting the relevant data segments from chest and wrist microphones using intensity difference, independent classification of the actions based on acceleration or sound and removal of false positives. Unusual human behaviour is detected by K. Hara et. al. [22], they used motion sensors in an intelligent house to detect the activities and unusual patterns based on Markov Chain. Vector quantization is used to reduce the sensor states and then change between these states is observed by transition probability. They detect the unusual behaviour by computing the distance between the state transition probabilities or by the likelihood of user action. The distance between the state transition probabilities was calculated by using either Kullback-Leiber distance or Euclid distance.

The abundant usage of mobile phones [4] have made them an essential element of modern day communications; far from the primary usage of sending and receiving phone calls and text messaging, today's mobiles devices are capable of a range of activities from checking your email, surfing the web, playing games, etc. Different wireless connectivity options are available in many phones, one of the functions provided by most mobile devices is Bluetooth [19]. Bluetooth provides short range communications and was originally intended for use in personal area networks (PANs). Although the range of Bluetooth is relatively short, the extremely high market penetration among mobile users and the fact that it provides no cost communications between physically clustered mobile users means Bluetooth provides an excellent medium for collecting information about user behaviour.

There have been several recent projects that utilize Bluetooth

functionality to collect data that describes the interactions of everyday life - "reality data". As mentioned, Eagle and Pentland [12] use a passive running Bluetooth application to collect information about the daily interactions of 100 subjects within a university campus. They do this by recording the Bluetooth ID (BTID) of every device encountered by each user. Hermersdorf et al [9] also use Bluetooth to collect proximity data. Unlike Eagle and Pentland, they collect their data for 14 users in an office environment and demonstrated how behavioural patterns can be found in Bluetooth proximity data. They used Independent Component Analysis (ICA) to estimate the problem. There second model demonstrated that rich Bluetooth environments can be used for positioning without any base station. Pietilainen et. al. [1] studied users with Bluetooth enabled devices in different meeting areas and designed a social interaction communication software package, MobiClique; that run on mobile phone and successfully builds a local social network. They developed three applications: epidemic news groups, adhoc social connections and asynchronous messaging. T. Nicolai et. al. [15] also used Bluetooth to detect the social behaviours. Unlike Eagle and Pentland, they collect their data from amongst a set of relative strangers; conference participants. Users were provided with software that run on java enabled mobile phones and detect other Bluetooth devices and allow users to interact with member of existing groups or make new groups.

The reality data collected from the above projects have inspired a variety of different types of usage. Modelling of daily life individual routines and group behaviours based on Bluetooth proximity data and cell phone usage are studied in [5] and [6]. While [2, 15, 9, 8, 7] explains the social behaviours, routine classification and tracking and positioning of users.

3. Methodology

This section discusses the methodology of our analysis that we adopted to get the results.

Due to the nature of Bluetooth interactions; in which any device may send an inquiry beacon to find what other devices are available to connect to, and in response to that, any device can be configured to reply, devices can be detected in any sequence when they are in proximity with each other. Detection methodologies must allow for levels of variance in the detected sequences. For example, a person goes to his office daily and detects other colleagues and his own desktop Bluetooth device. While these interactions have a specific pattern of proximity and time scale, it is not required that they follow a strict order of occurrence. The interaction indicates that the person comes in office, meets his colleagues and stay there for specific time duration. However, daily patterns of spatial and temporal detections may differ as a result of his physical movement, absenteeism of employees and the presence of visitors. Therefore, using different levels of variance allows the detection of behaviour even when some temporary devices, like visitors are also detected. In this analysis, the sequences of detection of all Bluetooth proximate devices is considered and use n-gram techniques to find the longest string of repeated pattern of detected Bluetooth devices to know if there are some behavioural patterns that repeat on daily or weekly basis. n-gram is also used to detect repeated patterns in musical compositions [21], approximately duplicate database record [23] and in large scale clustering of DNA texts [24]. The reason to use n-gram here in this analysis is the simplicity and efficiency in detecting all patterns in the data. The n-gram [20] is a fragment of length 'n' of a word of length 'L'. For example *behaviour* is a word of length L = 9 and it has the following bigrams (n = 2) and trigrams (n = 3):

Bigrams = {be, eh, ha, av, vi, io, ou, ur}
Trigrams = {beh, eha, hav, avi, vio, iou, our}

All further n-grams can be calculated for $0 < n <= L$. If 'N' is the total number of n-grams then we can calculate 'N' from:

$$N = L - n + 1$$

Where we are assuming that a window of size 'n' is shifting one character at a time. So it means if L = 9 and n = 2, 3, 4, 5, 6, 7, 8 then N = 8, 7, 6, 5, 4, 3, 2. Classification and similarity measure between textual documents is a common example of n-gram usage.

4. Experimental Results

This section discusses the results and behaviour of the target user based on his duration of stay with other Bluetooth devices and the repeated patterns in detection of devices.

Fig.1 shows the total number of repeated patterns with different n-gram values that are detected for the whole month of activities. For example there are six different patterns of seven devices that repeat at least twice through out the whole month. The longest and most frequent repeated pattern was chosen for further analysis. Devices in this pattern are (101, 102, 104, 105, 41, 6, 16), and it repeated 6 times through out the whole month.

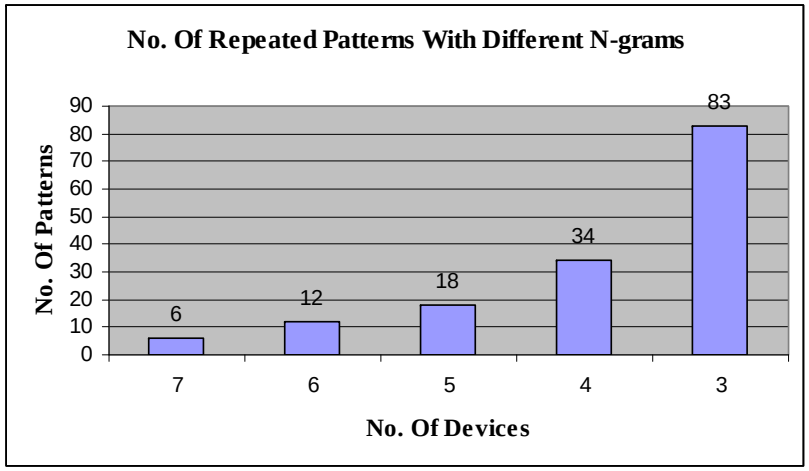

Fig.1 Number of repeated patterns with varying n-gram

As discussed, the combination of different Bluetooth devices in proximity can be detected in any pattern and if variation is allowed in the number of devices, as shown in Fig.2, the whole month is composed of many repeated activities that can be detected. Fig.2 shows the combinatorial analysis of repeated detections of the most frequently detected 7-gram, {101, 102, 104, 105, 41, 6, 16}, and adding different levels of variance. Zero level of variance refers to detection of the exact pattern. One level of variance means that 6 devices are from above mentioned sequence and they can combine in any order while one device is permitted to be different. As predicted, by adding each level of variance, the number of patterns that repeat throughout the month increases. This increase in the frequency of repeated patterns is because the window for the fixed repeated pattern has been compressed and the remaining devices are allowed to combine in any order. The maximum level of variance is 5, because the length of pattern under analysis is 7 and if we add more variance it doesn't make any pattern because only one device will remain fixed and all other devices will be allowed to be different.

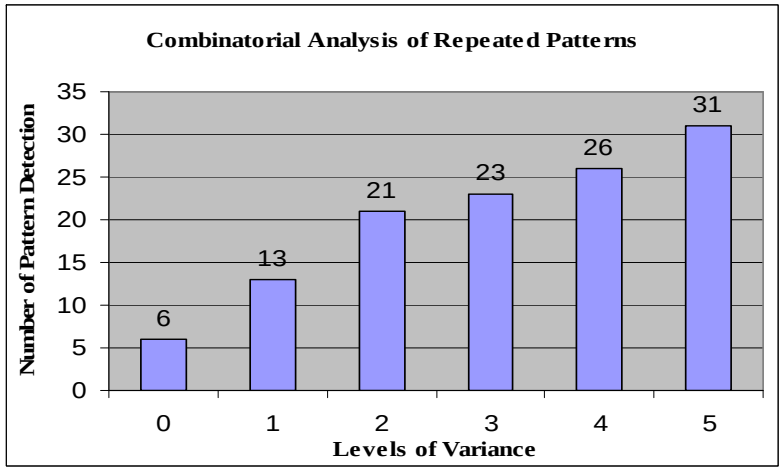

Fig. 2 Combinatorial Analysis of Repeated Patterns

Fig.3 shows the frequency of repeated patterns in specific hours for the whole month. With zero variance, the maximum length repeated patterns occur between 8am-10am, 10am-12pm and 4pm-6pm. As the level of variance is increased, the frequency of repeated patterns increases and time scale shows that these repeated patterns occur mostly between 10am-6pm, likely during the office hours. This means that the user is spending most of the time in office with his colleagues. It also shows that between 12pm-4pm

the detection of Bluetooth devices increases and activities reach at their peak because of the presence of most of the colleagues in the premises. For example if user is a student who normally work during usual office hours in research lab then this is the time when most of other students also work in lab and it can also because of lunch or coffee breaks when user spent his time with many other colleagues.

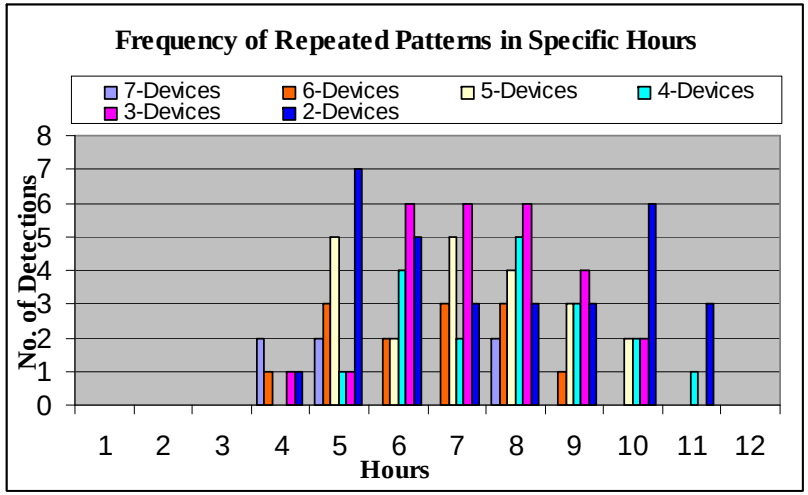

Fig. 3 Frequency of Repeated Patterns in Specific Hours

Now as the number of repeated patterns and the hours of their occurrence are detected, the next thing is to find out what percentage of these repetitions occurs per day for the whole month. Fig.4 shows the percentage of repetitions of different sequences through out the whole month with respect to time. For example longest sequence detected was identified and the percentage of time occupied by the repetitions of this sequence was calculated for those days in which those repetitions occurred. That longest sequence was of seven devices. In order to account for temporal variations as described above the sequence of six devices was taken and calculated the percentage in the same way as for seven devices. Those patterns of six devices that are subsets of patterns of seven devices and already calculated are neglected. The same procedure was applied to the patterns of five, four and three devices and then this analysis figure out those days in which repeated activities occur and also shows that percentage of repeated activities increases as we allow variance in number of devices.

This analysis was done on the most frequent repeated pattern. The next most frequent repeated pattern was also analysed and results showed that it increased the percentage of repetitions per day by 30.6 % for the whole month and also figure out some different days and repeated activities. So it is most likely that the remaining patterns that repeat almost with the same frequency will also increase the percentage of repetitions per day. Fig.5 shows that the frequently repeated patterns of 7-devices are not completely different from each other but there are some devices that are common in some sequences, that means there are more than one activity in which user is participating and some devices are part of different activities at the same time. Group-A has the device that is the part of all seven grams and so is the most common device. Group-B shows the devices that are part of four sequences and devices in Group-C are repeated three times while only two devices in Group-D are repeated twice and all devices in Group-E are individual devices that are not part of any other 7-gram.

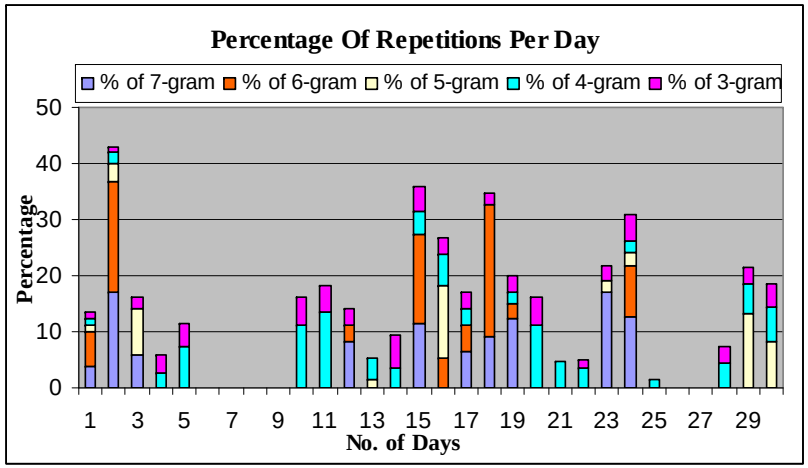

Fig.4 Percentage of repetitions per day through out the whole month

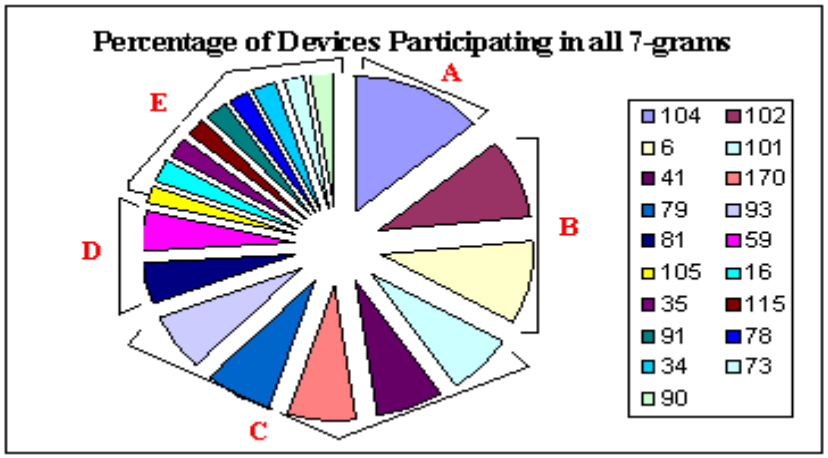

Fig.5 Percentage of Devices participating in all 7-grams

These initial results indicate that the user is performing some repeated set of activities during the specific number of hours and days. This likely shows that user spends most of his time at home in morning and nights and all detections of Bluetooth devices occur during the usual office hours and some evenings in week days. This is not the case with the people living with more entropic lives and having more variable behaviours because people with more entropic lives have diverse nature of routines. They have no specific daily routine patterns that they follow, while People with low entropy have certain specific daily/weakly routines that they follow as Nathan et. al. showed in his research work [12] that low-entropy lives tend be more predictable as they characterized by more repeated patterns in their daily lives. These low-entropic behaviours are helpful in detecting the human behaviour with more accuracy and it helps to identify the daily life routines.

5. Summary and Further Work

In this paper real time Bluetooth proximity data of one person was used and his behaviour for the whole month of November 2004 was studied. We successfully detected the devices with which user spent his most of the time throughout the month and then by using n-gram technique, we effectively discovered the repeated patterns in data set. These patterns show individual behaviour while in proximity with other Bluetooth devices. These successful extractions of repeated patterns help us to study and detect the wandering behaviour. By looking at the repeated patterns we can identify unusual behaviours and routines performed by the patients. In future work, we will try to discover different techniques to cluster and classify these repeated routines which will help us to identify the individual routines and behaviours more specifically.

6. References

[1] Anna-Kaisa Pietil¨ainen, Earl Oliver, Jason LeBrun, George Varghese, Jon Crowcroft, and Christophe Diot. Experiments in mobile social networking. Technical Report CR-PRL-2008-02-0003, Thomson, February 2008.

[2] Eagle N. *Machine Perception and Learning of Complex Social Systems.* MIT, 2005

[3] IT Facts Wireless Data: http://www.itfacts.biz/24-bln-bluetooth-enabled-devices-to-ship-worldwide-by-2013/11380

[4] IT Facts Mobile Phone Usage: http://www.itfacts.biz/41-bln-mobile-subscriptions-in-2008/12765

[5] K. Farrahi, Daniel Gatica-Perez, Daily Routine Classification from Mobile Phone Data. *IDIAP Research Institute, Martigny, Switzerland.*

[6] K. Farrahi, Daniel Gatica-Perez, What did you do today? Discovering daily routines from Large-Scale Mobile Data. *IDIAP Research Institute, Martigny, Switzerland.*

[7] Kotanen A. and Hannikainen M. and Leppakoski H. and Hamalainen T. Experiments on local positioning with Bluetooth. In *Proceedings of the International Conference on Information Technoogy: Computers and Communications (ITCC)*, 2003.

[8] Lamarca A., Chawathe Y., and Consolov S. Placelab: Device positioning using radio beacons in the wild. In *Third International Conference on Pervasive Computing, 2005.*

[9] M Hermersdorf, H Nyholm, J Perkiö, V Tuulos. Sensing in Rich Bluetooth Environments- Workshop on WorldSensorWeb, in Proc. SenSys, 2006 - sensorplanet.org

[10] McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G.: Automatic Analysis of Multimodal Group Actions in Meetings. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI) 27(3), 305–317 (2005)

[11] Munguia Tapia, E., Intille, S.S., Haskell, W., Larson, K., Wright, J., King, A., Friedman, R.: Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart monitor. In: Proc. Int. Symp. on Wearable Comp., Boston (2007)

[12] N. Eagle, A. Pentland, Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing 2006 – Springer.*

[13] Naeem, U. Bigham, J. (2008). "*Activity Recognition using a Hierarchical Framework*". Ambient Technologies for Diagnosing and Monitoring Chronic Patients Workshop, Tampere, Finland, January 29th, 2008

[14] Naeem, U. Bigham, J. (2007). "*A Comparison of Two Hidden Markov Approaches to Task Identification in the Home Environment*". Proceedings of the 2nd International Conference on Pervasive Computing and Applications, Page(s): 383-388

[15] Nicolai T., Behrens N., and Yoneki E. Wireless Rope: Experiment in social proximity sensing with Bluetooth. In *fourth annual IEEE International Conference on Pervasive Computing, 2006.*

[16] Paul L, Jamie A W, Holger J, M stager, Gerhard T, Amin A, Thad S, "*recognizing workshop activity using body worn microphones and accelerometers"*.

[17] Stiefelhagen, R., Bernardin, K., Ekenel, H.K., McDonough, J., Nickel, K., Voit, M., Woelfel, M.: Audio-Visual Perception of a Lecturer in a Smart Seminar Room. In: Signal Processing - Special Issue on Multimodal Interfaces, vol. 86 (12). Elsevier, Amsterdam (2006)

[18] Wren, C., Ivanov, Y., Kaur, I., Leigh, D., Westhues, J.: SocialMotion: Measuring the Hidden Social Life of a Building. In: Hightower, J., Schiele, B., Strang, T. (eds.) LoCA 2007. LNCS, vol. 4718, pp. 85–102. Springer, Heidelberg (2007)

[19] Bluetooth Specifications, retrieved on 01-08-2009, retrieved from http://www.bluetooth.com/Bluetooth/Technology/Building/Specifications/

[20] N-gram: retrieved on 2 July 2, 2009, retrieved from http://www.cs.man.ac.uk/~mary/CS3421lectures/node3.html

[21] Nikunj Patel, Padma Mundur, "*An n-gram Based Approach to Finding the Repeating Patterns in Musical Data*", Computer Science and Electrical Engineering Department University of Maryland Baltimore County, Baltimore

[22] Hara. K. Omori. T. Ueno. R. "*Detection of unusual human behaviour in intelligent house*"; Proceedings of the 2002 12th IEEE workshop on Neural Networks for Signal Processing, 2002.

[23] Z. Tian, H. Lu, W. Ji, A. Zhou, Zhong Tian, "*An n-gram-based approach for detecting approximately duplicate database records*"; 22 August 2001, Springer-Verlag 2001

[24] Z. Volkovich, V. Kirzhner, A. Bolshoy, E. Nevo, A. Korol; "*The method of n-grams in large-scale clustering of DNA texts*"; The journal of pattern recognition society, accepted 2 May 2005.