Naïve Adaptive Probabilistic Sensor Fusion for Enhancing Context Recognition

A Thesis

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

in partial fulfillment
of the requirements for the degree

Master of Science

by

Shabnam Wahed

December 2019

APPROVAL SHEET

This Thesis
is submitted in partial fulfillment of the requirements
for the degree of
Master of Science

Author Signature: ______________________

This Thesis has been read and approved by the examining committee:

Advisor: Laura E. Barnes

Committee Member: Joanne Bechta Dugan

Committee Member: Mehdi O. Boukhechba

Committee Member: ______________________

Committee Member: ______________________

Committee Member: ______________________

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

December 2019

Abstract

Wearable and non-wearable smart devices are capable of capturing huge amount of information from their users, this is possible because of a variety of sensors installed on them. Sensor fusion technology can enable us to obtain a holistic picture of the users' context and accurately monitor their state, although it is very challenging to classify human context in real world. Handling multiple classes, uncertainty related to machine learning models, class-imbalance and large feature space are still some issues which need to be resolved, however recent researches propose different probabilistic models. Towards addressing these issues, in this study we explore a new probabilistic model fusion approach called Naive Adaptive Probabilistic Sensor (NAPS) fusion. This sensor fusion technique is capable of addressing uncertainty for multi- class classification in machine learning problems with large imbalanced Human Activity Recognition (HAR) dataset. Our empirical evaluation acclaims that NAPS fusion outperforms conventional sensor fusion technologies and enhances context recognition in natural environment. Our approach avoided dimensionality reduction techniques by developing structured feature-sets, which helps to resolve the class-imbalance issue. For multi-class classification task on Extrasensory dataset, NAPS fusion outperforms different UCSD sensor fusion models by 5%-56% in f1 score and 3%-9% in balanced accuracy.

Dedications

This is dedicated to people who have shaped me into who I am today! My parents who have taught me how to believe in myself always. My husband, M Arif Imtiazur Rahman, who has been there for me always and has guided me to accomplish my goals and my dreams. Without you, I can never be what I am today. I love you, will always love you, no matter what!

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Laura E. Barnes, for her mentoring and support. Thank you for giving me the time and providing me the opportunities to conduct my research. Thank you so much!

To Dr. Nicholas J. Napoli, I owe a debt of gratitude to you for your sincere guidance for my thesis and personal development. Thanks a lot!

To Mehrdad Fazli, thank you being wonderful collaborator.

Contents

Always, always, always believe in yourself, because if you don't, then who will, Sweetie?

Marilyn Monroe

1

Introduction

Activity recognition is a very hot research topic in the area of ubiquitous computing. Precise capture of human context can make a positive impact on quality of life. Technological advances in sensor and data fusion has enabled the research work more promising. Behavioral context recognition is receiving more and more attention recently due to the widely used digital devices especially wearable and smart devices such as cellphones, watches, tablets, etc. Accurate sensing of behavioral context recognition can develop many solutions for health monitoring, aging care, assisted living, sleep monitoring etc. Such applications with improved performance can enhance people's quality of life. For example, it can remind people of taking medications at the right time [3], keep track of activities of Diabetics patients so that caregivers can provide proper feedback of patient’s behavior [4], monitor people with Dementia to prevent any unwanted activities [5], and many more. Together with the smart sensing devices such as smartphones, smartwatches, and wearables collecting real-world data and proper context recognition, the possibility of smart living is near the doors.

1.1 Challenges of Human Context Recognition

The recognition of human context with big data still remains challenging. One such challenge is high dimensionality. Dimensionality means the quantity of features/attributes of a dataset. High dimensionality means that the quantity is so huge that we face difficulty in performing classification. In human activity recognition task, we extract a large number of features from the sensors of our smartwatches or smartphones. Without learning and establishing proper relationship between large feature space and different class variables, we are unable to classify activities properly [15]. Besides, it also adds up to the complexity of the models.

Another common challenge that machine learning models experience is uncertainty. Uncertainty occurs due to the lack of knowledge. Incomplete or missing sensor data causes uncertainty in the models. If we can handle uncertainty properly, we don't need to retrain the models with every new input.

One of the most common problems in every modeling especially classification, is to deal with skewed data. A simple example would be a binary classification which assigns each example to either positive class or negative class. The problem of data skewness arises when we have a considerable amount of imbalance among the classes. This issue can get worse if positive examples are inherently of more importance.

Human Activity Recognition (HAR) datasets are usually imbalanced if they capture natural human behavior. Multi-class classification in imbalanced dataset is very challenging because the classes no longer have equal relationship among them [16]. This study investigates several challenges to better HAR:

1. Dealing with high dimensionality with minimal cost of information loss
2. Handling uncertainty with each ML model
3. Minimizing class imbalances
4. Performing multi-class classification on imbalanced dataset

There also arise several other issues while fusing sensor data such as missing features and ambiguous ground truth. However, we chose ExtraSensory dataset [2] for our experiments to validate our NAPS fusion model because this dataset doesn't have ambiguous ground truth as only manually labeled ground truth data is used here. Besides, zero imputation is done for missing features.

1.2 Insights to Handle Challenges

1. **Dealing with High Dimensionality:** In order to handle high dimensionality, we can leverage dimensionality reduction techniques. In HAR domain, there are several commonly used dimensionality reduction methods such as Principle Component Analysis (PCA), RankFeatures, Correlation-based Feature Selection (CFS) method etc. [17]. Dimensionality reduction techniques reduce computational complexity. However, it comes with a cost. They can degrade the performance of the recognition process by ignoring important details of the data if the dataset is imbalanced. This can lead to substantial information loss which hinders the empirical learning [18]. This motivates us to find a way by which we can circumvent the utilization of different dimensionality reduction techniques. In order to accomplish this, we can create small models by leveraging subsets of features and later we can fuse them altogether as shown in Figure 1. This can take care of large feature space with minimal cost of information loss which can happen in dimensionality reduction techniques.

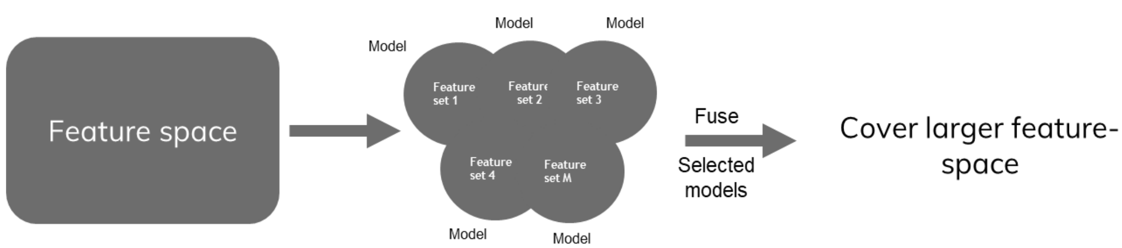

Figure 1: Handling high dimensional data

2. **Dealing with Uncertainty:** To make an accurate context inference, we mostly use probabilistic assignments when making decision under uncertain situations. Bayesian

network is trained on real data and gives probabilistic values for different nodes. Afterwards, this knowledge is used for inferring new contexts [19]. Fuzzy logic can play a great role to handle uncertainty. The study [20] generates fuzzy decision tree based on the fuzzy representation of context. It provides membership value while classifying activities. Thus, value can be used for representing the uncertainty of context to users. Another popular method of context aware reasoning is Dempster-Shafer theory of evidence [21], [22]. This technique performs well when combining information from multiple sources. It overcomes the ignorance due to the limited data from sensor fusion.

3. **Handling Class imbalances:** Generally, we can mitigate the problem of class imbalances by generating artificial instances of minority classes. We oversample the instances of minority classes and thus equalize the minority and majority classes. However, in HAR generating synthetic training instances can render the classification procedure questionable.
4. **Multi-class Classification:** As the name suggests the multi-class classification task classifies examples/instances into one of N different classes. Here N is greater than 2. If N=2, then it is binary classification. We can extend this binary classification into multi-class classification scenario by taking some simplistic approaches such as One vs All (OVA), All vs All (AVA) [23]. OVA builds individual classifier per class. For N different classes, it leads to N different classifiers. On the other hand, AVA trains a separate classifier for each distinguished pair of labels. This generates N(N-1)/2 classifiers. A deep insight is required while performing multi-class classification on imbalanced dataset.

1.3 Contributions

The main contributions of this work are listed below:

- Adopting a probabilistic sensor fusion method called Naive Adaptive Probabilistic Sensor (NAPS) [1] to solve multi-class classification problem in human activity recognition, the approach outperforms different UCSD sensor fusion models discussed in [2]. We show results for four class activity recognition with Extrasensory dataset in our experiments section.
- Selecting appropriate model by uncertainty assignment and better handling of class imbalance by augmenting response variables as well as leveraging a popular oversampling technique.
- Mitigating the problem of class imbalances by generating artificial instances of minority classes. We oversample the instances of minority classes and thus equalize the minority and majority classes.

"I am not afraid of storms for I am learning how to sail my ship."

Louisa May Alcott

2

Related Works

The world is becoming smaller and people are trying to make daily living more comfortable with the invention of many smart devices. Smartphone has become an inseparable device in our everyday life. Our smartphones are equipped with powerful sensors and are capable of multitasking. With the help of these powerful sensors automatic activity recognition can capture user’s state and surroundings quite precisely. Automatic activity recognition such as, what the user is doing, who he is with, where he is etc. aims at developing many incredible solutions for improving quality of daily living. For example, recent health monitoring devices are mostly based on manual as well subjective reporting, time to time. We could improve the applications by automation of behavior recognition related to diet, exercise, sleep, stress etc. similarly, aging care, patient monitoring can leverage automatic activity recognition to take care of health and medication of elderly and sick people.

Many works have been done to classify behavioral sensing by leveraging sophisticated sensors of these devices. Lee et al. [6] showed the feasibility of a hierarchical probabilistic model-based approach to sense and classify human activities from accelerometer data only. However, it is found pretty difficult to classify due to the lack of sufficient information about context.

In order to improve accuracy and performance of context recognition, sensor fusion technology evolved. Usually a single sensor provides very limited information of the context [2]. That's why we continuously collect data from different sensors and combine these data to get a better view of the context. Luinge et al. [7] proved that accelerometer together with gyroscope can enhance the estimation of 3D orientation which can play a vital role in ambulatory human movement analysis. Mortazavi et al. [8] leveraged the discovery and introduction of a framework combining both sensors - gyroscope and accelerometer. This can be utilized for creating platform for physical activity recognition. Weiss et al. [9] conducted research to find out the combined performance of smartphones and smartwatches for human activity recognition. Smartwatches are capable of outperforming smartphones' performance of accurately identifying specific hand activities (for example, eating). Thus, fusing data from different smart devices can help to improve activity recognition results.

In order to accomplish robust measurements, there are 3 types of sensor fusion based on data processing [10] -

a) Data level fusion,

b) Feature level fusion, and

c) Decision level fusion.

a) Data level fusion: Direct data fusion combines all redundant raw sensor data directly. The fusion technique works for multiple homogeneous sensors where the sensor data can't be directly merged. Generally, raw sensor data are fused without performing any feature extraction. The concatenated sensor data is represented as a single unit and later is used for feature extraction. This type of fusion has good performance in enhancing signal to noise ratio (e.g. optimal averaging of sensor arrays), extracting original signal from the sensors estimating various combinations of signals from source (e.g. blind source separation) [11]. Liu et al [12] developed a composite health index by leveraging data level fusion which results into an upgraded degradation-based prognostic model.

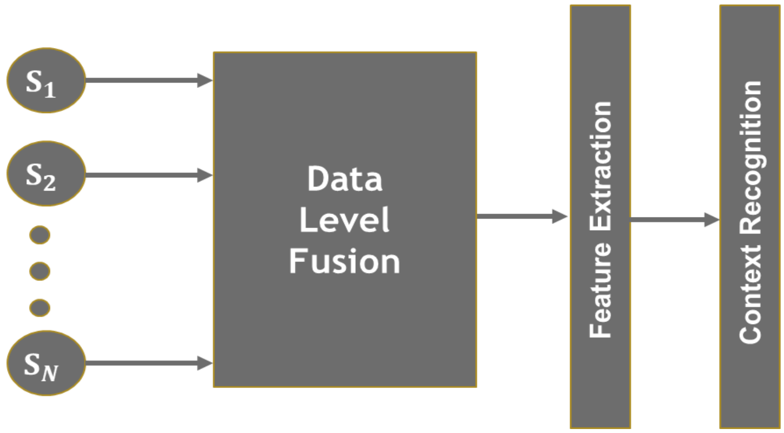

Figure 2: Data level fusion

b) Feature level fusion: Feature fusion adds up the illustrative features from various sensor modalities and select the suitable features, sometimes, transform them into lower dimension using dimensionality reduction methods such as Principle Component Analysis (PCA). The main application of feature level fusion is in classification purpose leveraging pattern recognition, neural network techniques etc. This type of fusion suffers from information loss as well as performance loss [19]. The feature level fusion is capable of deriving decision from both homogeneous and heterogenous sensors. Ross and Govindarajan [13] leveraged feature level fusion in biometric systems.

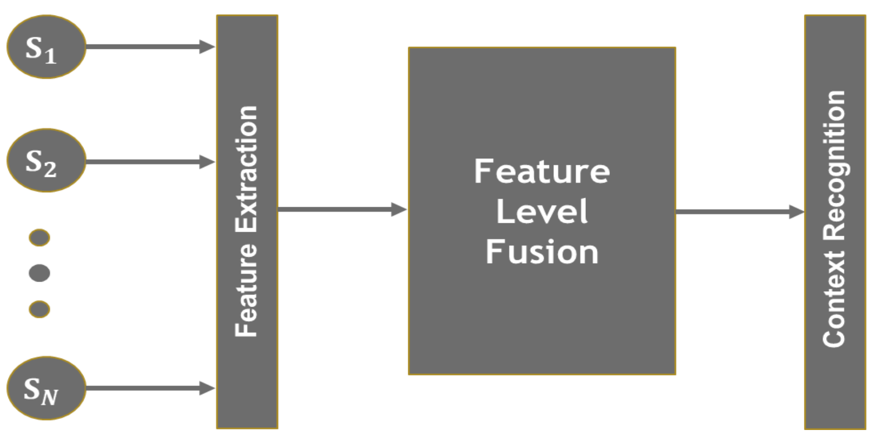

Figure 3: Feature level fusion

c) **Decision level fusion:** Decision level fusion method combines preliminary decisions from individual classifiers and then takes the final decision. This ensemble technique is capable of improving context recognition accuracy. The most commonly used decision fusion algorithms are- Dempster-Shafer's (DS) theory, weighted decision, Bayesian Inference method etc. The fusion technique also works for both homogeneous as well as heterogenous sensors. Gunatilaka et al [14] leveraged both feature level and decision level fusion for non-commensurate sensor data and found significant performance improvements compared to singe sensor data.

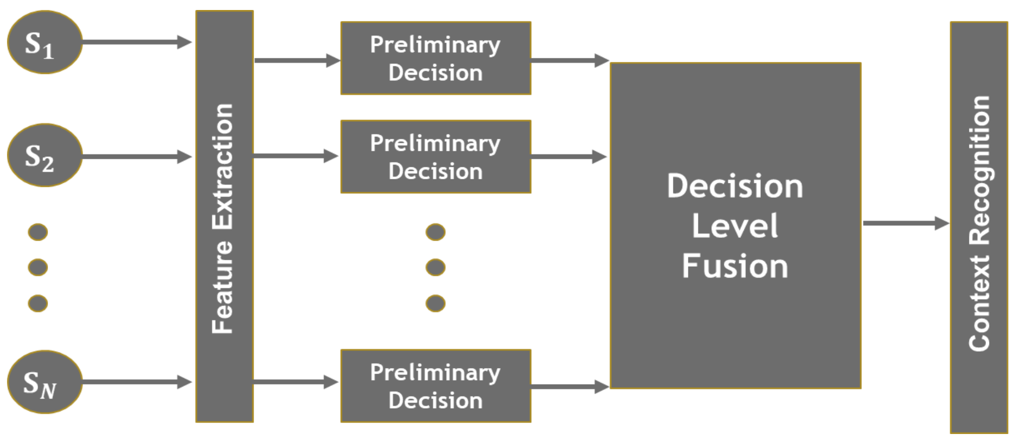

Figure 4: Decision level fusion

Vaizman et al. [2] discussed the difficulties related to detailed human context recognition and showed how sensor fusion can solve this issue. The study has explored both feature level and decision level sensor fusion methods –

I. Feature level fusion
   - Early Fusion (EF)

II. Decision level fusion

- Late fusion using Averaging (LFA), and
- Late fusion with Learned weights (LFL)

- **Early fusion (EF):** In this fusion method, features from multiple sensors are combined together to make a single long length feature vector x with d dimension. Then we learn a d-dimensional logistic regression classifier from a training set. Finally, a binary classification along with a probabilistic value is obtained by applying this d-dimensional logistic regression classifier to a testing set.
- **Late fusion using averaging (LFA):** This fusion method leverages bagging approach. Then averages output probabilities of single sensor classifiers and finally gives the final probability value. The fusing method places equal weight to each sensor and avoids additional training after the learning of single sensor classifiers. The output is positive if average probability value is greater than 0.5 and is negative if vice-versa.
- **Late fusion with learned weights (LFL):** This fusion method learns sensor weights from data instead of putting equal value to each sensor like LFA. The reason behind this is some sensors can perform better for different labels. The method learns an N-dimensional logistic regression model which has N probability outputs of the single-sensor classifiers as input and the output is the final probability value.

Vaizman et. al [25] in their later work have proposed a multi-class classification approach with the use of multiple layer perceptron (MLP) or Neural Networks to recognize the context. The dataset is randomly divided into train and test sets. The model is trained with initial set of labels and the performance for missing sensors is evaluated. Objective function is adjusted to handle the unbalanced and missing labels. Thus, incomplete and unbalanced labeling and missing sensor problems are addressed in the study. When collecting new data, transfer learning is used. Accuracies are defined per class and then these are averaged over all the classes to compute overall performance of the model. The results show that MLP maximizes the accuracy rate and thus, improves the performance. In our research, we compare our results against these techniques here to elaborate the improvement, details follow in experiment section.

"Nothing is impossible., the word itself says 'I'm possible'!"

Audrey Hepburn

3

Approach and Solution

Naive Adaptive Probabilistic Sensor (NAPS) Fusion proposed by Napoli et. al [1] is a probabilistic model fusion approach which is capable of addressing uncertainty for multi-class classification problems. This fusion method is based on the DST and treats each class as one single proposition (one possible outcome). DST requires these propositions to be mutually exclusive and this implies that we are not able to use this method for conflicting label contexts. The power set is also consisting of all the possible permutations of classes, namely response permutations. Then, the author aimed at building a model for each two exclusive and complementary subsets of the power set. These pairs are called augmented response variables and will be utilized to create binary classification models. This would leave us with a handful of models with different degrees of imbalance. All these models contribute to our decision-making process as sources of evidence that will be combined together later to make a decision.

Napoli, also equipped their fusion model with random selection of the features to reduce the complexity and dimensionality of the models. To this end, they defined a specific structure for the reduced feature sets which determines the number of features selected from each sensor. Now within this structure the features are randomly selected to create one reduced feature set. This random selection of features can be repeated many times to ensure that we have one or more good representatives of the entire feature space amongst reduced feature sets. For each pair of reduced feature set and augmented response variable, we may build a model.

To further improve the performance of the fusion model and reduce the variance, they applied bootstrap aggregating (bagging) on each model. Also, Synthetic Minority Over-sampling Technique (SMOTE) is used to help with the data skewness. What makes the NAPS so flexible is that it creates a huge pool of models to select from. Additionally, model selection is done for each single example separately. In other words, NAPS allows each example to select the best performing models and then combines them to make a proper decision for that certain example.

3.1 Fundamentals of Dempster-Shafer Theory (DST)

Dempster-Shafer theory (DST) also known as evidence theory, is regarded as the generalization of the probability theory by taking uncertainty into consideration. Uncertainty is the central concept in DST that quantifies the degree of certainty in our decisions. Belief and plausibility are other concepts that have been formed around the uncertainty in DST to measure the minimum and maximum certitude about the decision. DST provides a framework for combining evidences (possibly conflicting evidences) from different sources and making a decision. DST offers several advantages such as problems with specifying priors can be avoided, uncertainty and ignorance can be expressed. DST starts by assuming a Frame of Discernment, Ω which is complete and contain all possible subsets. The equation below represents the property.

$$m(\varphi) = 0 \tag{1}$$

A probability mass is assigned to each subset of Ω in a given domain. We call this assignment Basic Probability Assignment (BPA). Basic assignment functions are defined on the power set of 2^{Ω}. If the subset is assigned to a non-zero probability mass, then it is called a focal element of BPA. BPA fulfils the following property [24]:

$$\sum_{x \subseteq \varphi} m(X) = 1 \tag{2}$$

The equation normalizes all statements of a single data source. It confirms that all data is equally assessed i.e. each data sources is given same weight. Dempster-Shafer allows the combination of basic assignments with the help of Rule of Combination [22]. Equation below shows the rule.

$$m(D) = \frac{\sum_{x \cap y = D \neq \varphi} m(X).\, m(Y)}{1 - \sum_{x \cap y = \varphi} m(X).\, m(Y)} \tag{3}$$

Where $X, Y, D \subseteq \varphi$. The denominator and the numerator represent the conflicts between the sets and the cumulative confirmation from the sets to support hypothesis D respectively.

3.2 Solution Steps

We applied NAPS Fusion on Extrasensory Dataset with 4 mutually exclusive selected classes/activities. We focused on only 4 class classification problem here in order to reduce computational complexity. NAPS Fusion works by creating different propositions of response variables which are later on augmented to produce various ML models. The number of models increase with the number of response variables. So, we kept the number of response variables constrained to 4. We selected 4 mutually exclusive context labels such as Lying down, Walking, Sitting, and Running out of 6 mutually exclusive activities. The steps to our solution are outlined below.

a) **Step 1 - Developing feature-sets:** ExtraSensory dataset [2] has total 175 features with 6 core sensors as below.

Table 1: Number of Features for Each Sensor of Extrasensory Dataset

Accelerometer	Gyroscope	Audio	Location	Phone State	Watch Accelerometer	Total 175
26	26	26	17	34	46	

We took 10% of features from each sensor. When there was a fraction, we used the integer value. For example, if the 10% of the total number of Accelerometer features is 2.6, we took 2 features. Thus using 10% of features from each sensor, we got a total of 14 features.

Table 2: Number of Features for Each Sensor for creating Feature-sets

Accelerometer	Gyroscope	Audio	Location	Phone State	Watch Accelerometer	Total 14
2	2	2	1	3	4	

Taking into consideration the system's performance by calculating f1 score with different number of feature-sets as shown in Figure 1, we created 200 feature-sets in order to make a balance between computational complexity and highest performance.

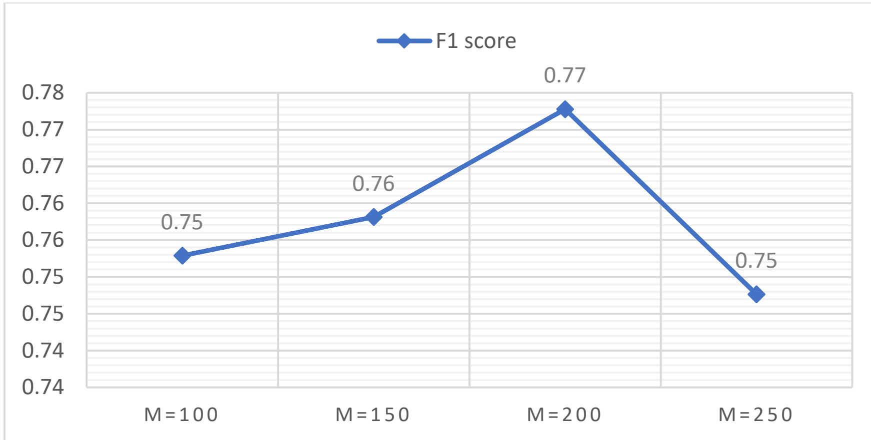

Figure 5: f1 score for different number of feature sets

b) **Step 2 - Augmented Response Variables and their Model Permutations:** In the light of DST, we used mutually exclusive and exhaustive propositions as our Frame of Discernment (FoD), Ω.

$$\Omega = \{\beta1, \beta2, \beta3, \beta4\} \quad (4)$$

Then we took the power set of it to create all possible propositions to model uncertainty related to FoD. So, the selected 4 activities would be our FoD. We got 16 possible propositions from our FoD.

$$2^{\Omega} = \{\emptyset, \{\beta_1\}, \{\beta_2\}, \{\beta_3\}, \{\beta_4\}, \quad (5)$$
$$\{\beta_1, \beta_2\}, \{\beta_3, \beta_4\}, \{\beta_2, \beta_3\}, \{\beta_1, \beta_4\}, \{\beta_1, \beta_3\}, \{\beta_2, \beta_4\},$$
$$\{\beta_1, \beta_{2,}, \beta_3\}, \{\beta_2, \beta_{3,}, \beta_4\}, \{\beta_1, \beta_{3,}, \beta_4\}, \{\beta_1, \beta_{2,}, \beta_4\}, \Omega\}$$

In order to reduce the four-class ML problem to a two-class ML problem, we create ($P = 2^{\Omega-1} - 1 = 7$) total 7 proposition assignments as shown in the Table below.

Table 3: Proposition Assignments of the Response Variables (P=7)

Response Permutations	Proposition Assignments	
1	$\{\beta_1\}$	$\{\beta_2, \beta_{3,}, \beta_4\}$
2	$\{\beta_2\}$	$\{\beta_1, \beta_{3,}, \beta_4\}$
3	$\{\beta_3\}$	$\{\beta_1, \beta_{2,}, \beta_4\}$
4	$\{\beta_4\}$	$\{\beta_1, \beta_{2,}, \beta_3\}$
5	$\{\beta_1, \beta_2\}$	$\{\beta_3, \beta_4\}$
6	$\{\beta_1, \beta_3\}$	$\{\beta_2, \beta_4\}$
7	$\{\beta_1, \beta_4\}$	$\{\beta_2, \beta_3\}$

Besides, by augmenting response variables we are reducing the amount of uncertainty. This approach creates huge number of models which are later combined to get a larger feature space.

c) **Step 3 - SMOTE and Bagging:** The response variables have different number of examples i.e. there are class imbalances. We are adding class imbalance with augmented response variables. That's why we implemented Synthetic Minority Over-sampling Technique (SMOTE). It creates synthetic instances of minority classes in order to balance the majority and minority classes.

With 7 proposition assignments and 200 feature-sets, SMOTE produces 7x200 models. All these models can't be always generalized with the datasets. In order to reduce the variance, we leveraged bagging or bootstrap aggregation approach. Bags mean creating a number of subsets of data. We created T=200 bags here. We used 60% of our training data with replacement, i.e., there were repetitions of instances. Then we used those bags to train our different models. Thus, we have an ensemble of distinct models.

d) **Step 4 - Model Uncertainty Calculation:** Now we have a large number of models for classification task. Obviously not all the models are reliable. In order to find out the most reliable ones to perform in the final stage, we can calculate the uncertainty value related to each model. We leveraged normalized bagging voting approach to find out the model uncertainty. The equation below calculates the uncertainty value by taking into account the number of votes for each proposition assignment.

$$\beta_1 = 1 - \left(\frac{\sqrt{(\frac{a_1}{B} - \frac{1}{H})^2 + \cdots + (\frac{a_H}{B} - \frac{1}{H})^2}}{\sqrt{\frac{H-1}{H}}} \right) \tag{6}$$

Here H represents total number of hypotheses/classes, a_H stands for the number of votes for a_H th hypothesis/classes and B denotes total number of bags in the model.

Besides, biased models are unable to perform properly. They tend to miss the relevant relations between input features and target labels. Generally, class imbalances cause biased models. If we train a model without fixing this issue, it degrades the model's performance. In order to avoid this issue, we can leverage original dataset to figure out imbalances associated to each class as shown in the equation below.

$$\beta_2 = \frac{\sqrt{(\frac{E_1}{X} - \frac{1}{H})^2 + \cdots + (\frac{E_H}{B} - \frac{1}{H})^2}}{\sqrt{\frac{H-1}{H}}} \tag{7}$$

Here X is the total examples from the original dataset. The amount of class imbalance of β_e depends on the ratio of the number of examples from the class E_H and X. For the total uncertainty, β we can simply can the mean value of the β_1 and β_2.

$$\beta = \frac{\beta_1}{2} + \frac{\beta_2}{2} \tag{8}$$

5. **Step 5 - Model Selection and Fusion:** With the uncertainty value associated with each model, we selected the top performing models and fused them to perform the classification task. We selected six feature-sets with highest reliability value (i.e. lowest uncertainty value) for all 7 proposition assignments. Then by leveraging One vs All approach, we reduced them to 4 proposition assignments. Finally, we selected 6x4 models to fuse together using Dempster's Combination Rule (DCR) [22] to classify the context labels.

"Only do what your heart tells you."

Audrey Hepburn

4

Experimental Results

Activity recognition is a very hot research topic in the area of ubiquitous computing. Precise capture of human context can make a positive impact on quality of life. Technological advances in sensor and data fusion has enabled the research work more promising.

4.1 Dataset Description

Table 4: Brief overview of Extrasensory dataset

<table>
<tr><th colspan="2">Users' Statistics</th><th colspan="3">Analyzed Sensors -measurements & Examples</th><th rowspan="2">Diverse Contexts</th></tr>
<tr><th>Statistics</th><th>Range</th><th>Sensor</th><th>Raw measurements</th><th>No of examples</th></tr>
<tr><td>Age in years</td><td>18-42</td><td>Accelerometer</td><td>3-axis (40 Hz)</td><td>308,306</td><td rowspan="6">1. Lying down
2. Sitting
3. Walking
4. Running
5. Bicycling
6. Standing
Mutually Exclusive Activities
7. With friends
8. Phone in pocket
9. Or inside
10. Surfing the internet
11. Eating
12. Phone in hand
13. Watching tv
14. Or outside
15. Phone in bag
16. Or exercise
17. Drive I am the driver
18. With co-workers
19. In class
20. In a car
21. In a meeting
22. Loc Home
23. Cooking
24. Lab work
25. Cleaning
26. Grooming
27. Toilet
28. Drive I am a passenger
29. Dressing
30. Fix restaurant
31. Bathing shower
32. Shopping
33. On a bus
34. At a party
35. Drinking alcohol
36. Washing dishes
37. At the gym
38. Phone on table
39. Strolling
40. Stairs going up
41. Stairs going down
42. Singing
43. Loc beach
44. Doing laundry
45. At a bar
46. Elevator
47. Sleeping
48. At school
49. Computer work
50. Talking
51. Loc main workspace</td></tr>
<tr><td>Weight in kg</td><td>50-93</td><td>Gyroscope</td><td>3-axis (40 Hz)</td><td>291,883</td></tr>
<tr><td>Height in cm</td><td>145-188</td><td>Watch accelerometer</td><td>3-axis (40 Hz)</td><td>210,716</td></tr>
<tr><td>Participation duration in days</td><td>2.9-28.1</td><td>Location</td><td>Long-lat-alt</td><td>273,737</td></tr>
<tr><td>Labeled examples</td><td>685-9,706</td><td>Audio</td><td>13 MFCC (46-ms frames)</td><td>302,177</td></tr>
<tr><td>Additional unlabeled examples</td><td>2-6,218</td><td>Phone State</td><td>Once per Example</td><td>308,320</td></tr>
</table>

ExtraSensory dataset [2] is a rich, publicly available dataset with 51 diverse context labels and 175 different features across multiple sensor modalities. There are 300k examples of 60 participants over 7 days. The dataset involves data from multiple sensors such as accelerometer, gyroscope, magnetometer, audio, location, and phone-state from the person's phone both iPhone and Android, as well as accelerometer data from an additional smartwatch. Table 4 reports a brief overview of the dataset.

4.2 Evaluation and Metrics

There are many approaches to improve the performance of context recognition system such as machine learning algorithms, sensor fusion, video analysis etc. It is not emblematic to combine several approaches such as machine learning algorithms and sensor fusing to boost up the recognition rate.

In the study we selected two fusion technologies and compared their performance by applying both of them to a publicly available HAR dataset, the Extrasensory dataset. The dataset is comprised of natural behavior of human subjects which is much more challenging to classify.

We selected 4 activities - Lying down, Sitting, Walking, and Running. We considered examples from all of the core sensors i.e. Accelerometer, Gyroscope, Watch Accelerometer, Audio, Phone state, and Location. We evaluated different classifier performance using 5-fold cross validation. Finally, we calculated the performance metrics to compare their performances. We selected and measured the following metrics:

1. **Precision:** We can also call Precision as Positive predictive rate in the realm of classification. In other words, we can gain the knowledge of the proportion of actually correct instances from the predicted positive instances.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

2. **Recall:** We calculated the recall value to find out what proportion of actually positive instances was recognized correctly. So with recall we can minimize false negatives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3. **Specificity:** We calculated the specificity value to find out what proportion of actually negative instances was recognized correctly. Specificity takes into account false positives. High specificity value results into low type-I error rate.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

4. **f1 Score:** f1 score is the measure of test's accuracy. It is simply the harmonic mean of precision and recall value. It takes on account both false positives and false negatives. So f1 score plays a significant role specially for the dataset which has uneven class distribution.

$$\text{f1 score} = \frac{\text{2 x Precision x Recall}}{\text{Precision} + \text{Recall}}$$

5. **Balanced Accuracy (BA):** Accuracy measures the correctly identified instances from total number of instances. However, it provides inaccurate information when measured in imbalanced dataset. So, we do not solely depend on the accuracy values. For both binary and multi-class classification, the performance of balanced accuracy is promising. It deals

with imbalanced datasets pretty well. It is the mean value of the recall and Specificity obtained from each class individually.

$$\text{Balanced Accuracy} = \frac{\text{Recall} + \text{Specificity}}{2}$$

4.3 Results and Analysis

We implemented the Late Fusion technologies- LFA and LFL and evaluated 5-fold cross validation performance. We measured 2 performance metrics - f1 score and balanced accuracy and got almost similar results as in [2]. We recreated the entire sensor fusion pipeline from the paper [2] and got almost similar results.

We calculated the performance metrics such as balanced accuracy, f1 score, precision, recall, specificity from NAPS fusion and both UCSD late fusion technologies- LFA, LFL as shown in the Table below. Best achieved results are marked as bold.

Table 5: Performance metrics comparison of different sensor fusion technologies

Performance metrics	Classes	Sensor fusion technologies		
		LFA	LFL	NAPS
Precision	Lying Down	0.70	0.77	**0.89**
	Sitting	0.65	0.64	**0.83**
	Walking	0.28	0.26	**0.58**
	Running	0.02	0.03	**0.61**
Recall	Lying Down	**0.91**	0.88	0.88
	Sitting	**0.86**	0.80	0.77
	Walking	**0.90**	0.81	0.77
	Running	**0.83**	0.73	0.58
Specificity	Lying Down	0.80	0.86	**0.91**
	Sitting	0.70	0.71	**0.82**
	Walking	0.74	0.74	**0.79**
	Running	0.51	0.74	**0.74**
f1 score	Lying Down	0.79	0.77	**0.88**
	Sitting	0.74	0.64	**0.80**
	Walking	0.43	0.26	**0.66**
	Running	0.03	0.03	**0.59**
Balanced Accuracy	Lying Down	0.85	0.77	**0.92**
	Sitting	0.78	0.64	**0.83**
	Walking	0.82	0.26	**0.85**
	Running	0.67	0.73	**0.76**

The Figures below represent the comparison of performance metrics calculated for different fusion techniques- LFA, LFL, and NAPS.

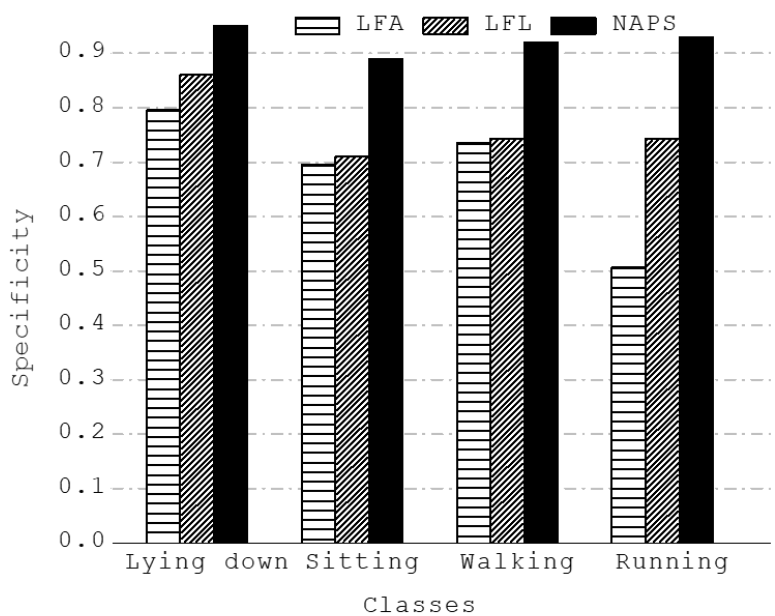

Figure 6: Performance Metrics Comparison- Specificity

Specificity indicates the ratio of negative instances which are classified correctly as negative. High specificity means that we have correctly ruled out all negative examples. The Figure 6 shows that we have high specificity value for each our context labels.

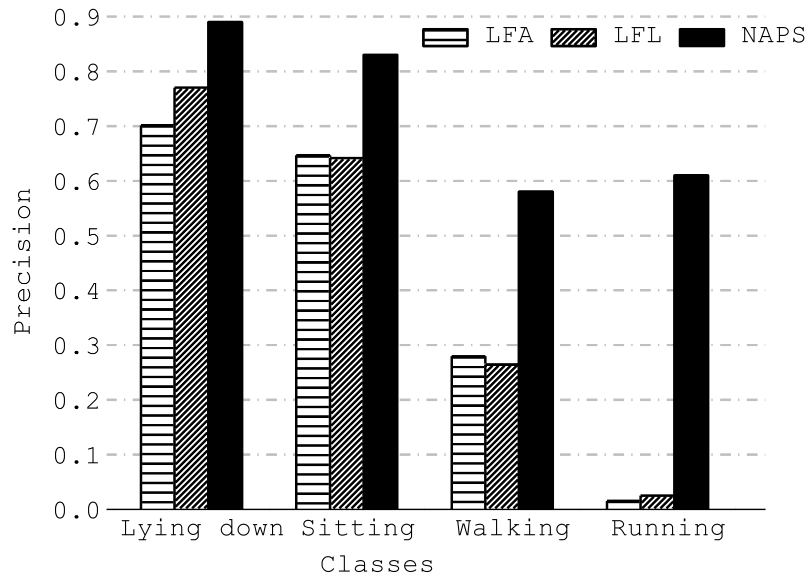

Figure 7: Performance Metrics Comparison- Precision

From Figure 7 it is noticeable that NAPS Fusion has higher precision value for each activity. That means it has higher exactness than UCSD Late fusion methods. Precision-recall counter each other, i.e., with the improvement of precision value, recall is penalized. Same has happened here in Figure 8.

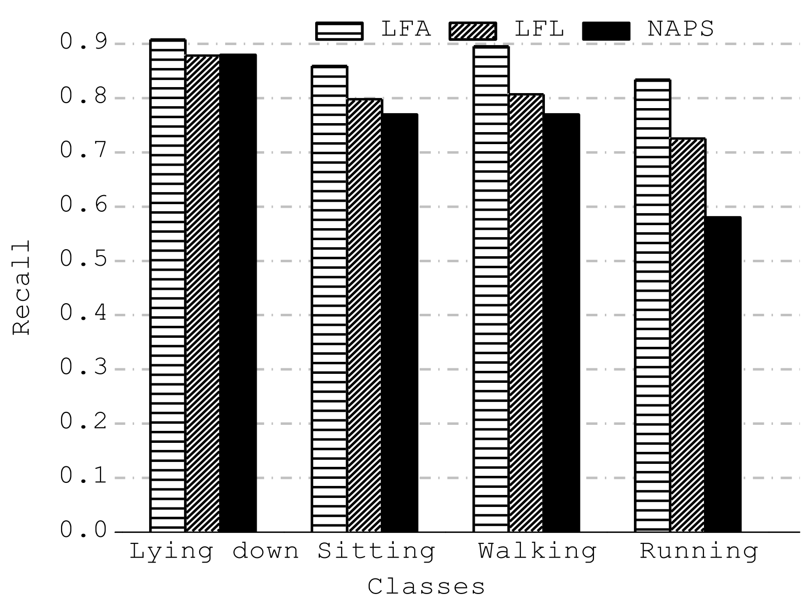

Figure 8: Performance Metrics Comparison- Recall

In NAPS fusion, we acquired lower recall value for each context label. The takeaway from these graphs is that we have better performance with NAPS except for recall. There is room for improvement by restructuring the feature space or tuning the parameters of our models.

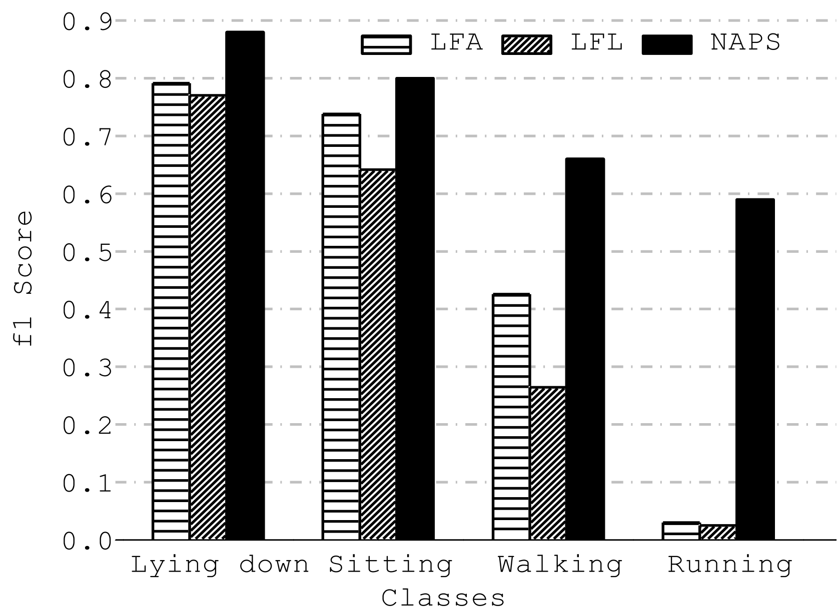

Figure 9: Performance Metrics Comparison- f1 score

F1 score is the harmonic mean of recall and precision. Higher f1 score indicates that we have balanced values of precision and recall for our dataset. The Figure 9 shows that with NAPS we can greatly improve the f1 score. So, we can ensure that we are having less miss-classification rate.

Balanced accuracy takes into account specificity and recall value. Figure 10 shows that we have higher balanced accuracy for each of the context labels with NAPS than other 2 fusion techniques. The reason behind that is NAPS has balanced the classes properly and has improved classification rate.

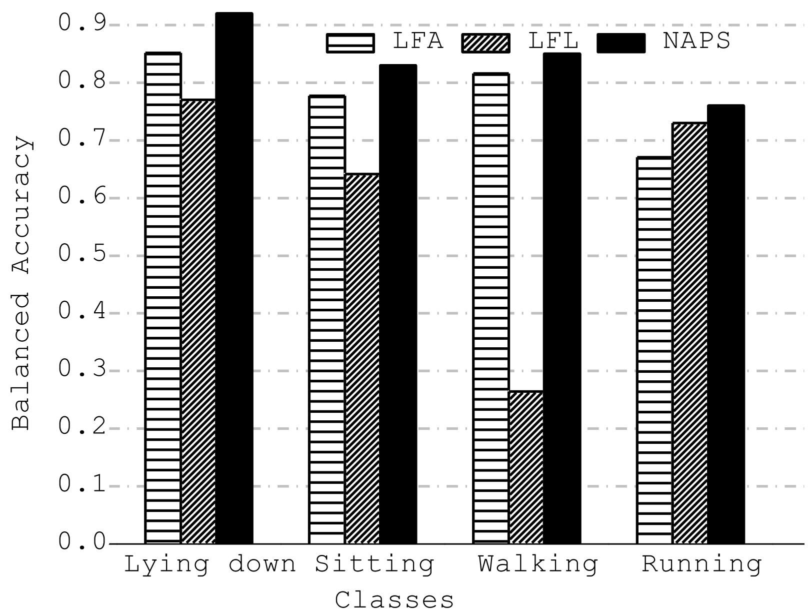

Figure 10: Performance Metrics Comparison- Balanced Accuracy

For example, class “Running” has a smaller number of examples than other 3 classes as shown in Figure 11.

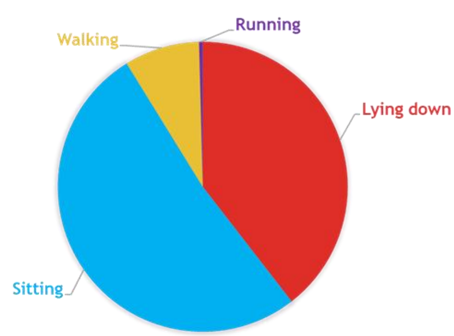

Figure 11: Total number of available examples for each activity

Both LFA and LFL classify “Running” with pretty good accuracy (about 0.70) but an extremely low f1 score (0.04). It means that both UCSD sensor fusion techniques have failed to handle class imbalances which results into such poor f1 score. F1 score accounts for the performance of all classes. However, with NAPS we can acclaim that we are resolving the issue and thus enhancing the performance of classifiers as shown in Figure 12.

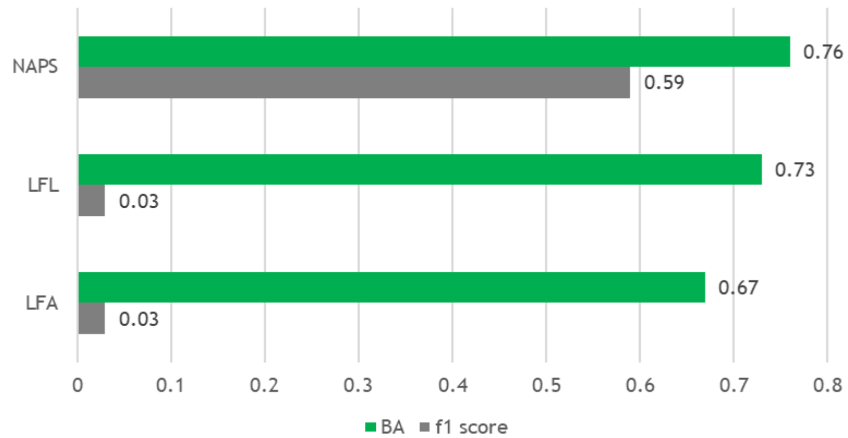

Figure 12: Balanced accuracy and f1 score comparison for Running

We compared NAPS with some baseline approaches such as Gaussian Naïve Bayes, Random Forest (RF), SVM, Gradient Boosting Machine and reported their performance in Table 6 in order to compare with NAPS.

Table 6: Comparison of NAPS with some baseline approaches

Balanced Accuracy Comparison					
Classes	**NAPS**	**Naïve Bayes**	**SVM**	**Random Forest**	**Gradient Boosting Machine**
Lying Down	**0.92**	0.74	0.83	0.51	0.67
Sitting	**0.83**	0.78	0.62	0.75	0.78
Walking	**0.85**	0.77	0.54	0.51	0.53
Running	**0.76**	0.67	0.50	0.5	0.50

A group of randomized decision trees are the foundation of random forest. A random training subset as well as a random features' subset is used for learning for every decision tree. It simply takes the mean value of the individual outputs of each decision tree in order to make a prediction on a test sample. In simple, we get the final probability score by averaging the probability score from each tree and the class with highest probability is the predicted activity for that test example. Krupitzer et al. [26] used a random forest classifier to fuse features from different sensors attached to different positions to classify Activity of Daily livings (ADL). Besides, Bayesian networks is one of the common algorithms used for sensor fusion. Gaussian Naïve Bayes is a kind of simple Bayesian networks [27]. It is noticeable that NAPS outperforms other sensor fusion algorithms as shown in Figure 13.

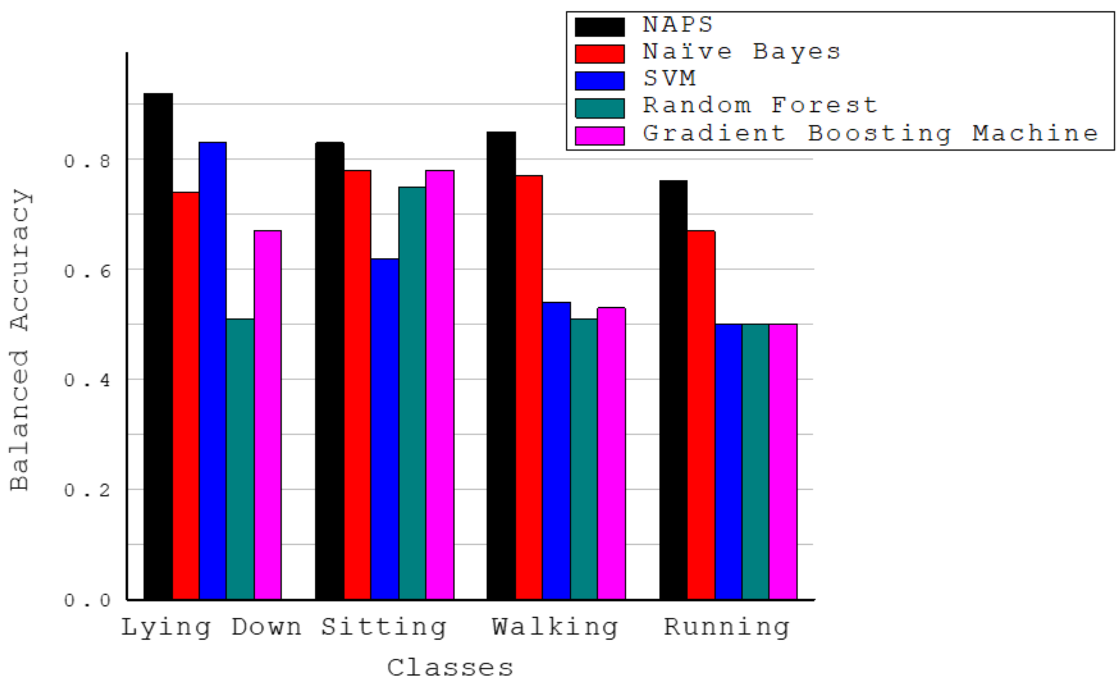

Figure 13: Comparison of NAPS with some baseline approaches

NAPS Fusion has better performance in behavioral context recognition in uncontrolled environment. The reason behind the improved recognition rate are:

1. **Dealing with High Dimensionality:** NAPS fusion creates a set of feature spaces which are randomly chosen out of the original feature space. When we fuse models together with the associated feature-sets, we actually extend the feature space and thus, deals with high dimensionality without using any dimensionality reduction techniques.
2. **Handling uncertainty:** Model uncertainty can degrade the performance of SMOTE. SMOTE leverages kNN approach. It clusters the input feature space using k-means. There arises unavoidable uncertainty in the model with the lack of density in input feature space. In order to reduce sparsity, we augment the response variables. NAPS is built upon DS theory and benefits from dynamically weighting of the sensors that ensures reliable and unbiased models with improved accuracy.
3. **Multi-class classification:** By assigning a mass to each element in the powerset of the FOD, NAPS makes it possible to account for the classes happening together.
4. **Class imbalances:** NAPS augments the response variables which allows to create more data instantaneously. NAPS also leverages SMOTE. This allows us to build models on an enhanced and balanced dataset.

“Think like a queen. A queen is not afraid to fail. Failure is another stepping stone to greatness.”

Oprah Winfrey

8

Conclusion

Human activity recognition is very important in improving both doctor-patient interaction and personal daily living. Unbalanced data, large feature space, multi-class classification, uncertainty in the decision boundary of models make the classification task very challenging. We tend to use some sampling approaches such as SMOTE to make a balance among the classes by generating synthetic instances. However, SMOTE can't perform well with high dimensional dataset. When we aim to tackle high dimensions with dimensionality reduction techniques, we prone to lose some valuable information. If we are unable to resolve these issues, we may end up with unreliable and biased models with very poor recognition rate for multi-class classification.

A novel sensor fusion approach like NAPS Fusion is capable of overcoming these challenges. It creates structured sub-set of features which are later merged to extend the dimension of feature-space. NAPS resolves the class imbalance by leveraging SMOTE as well as combining response variables. By assigning uncertainty value to each model, the fusion technique smartly selects models that increases prediction accuracy. With NAPS we are capable of minimizing mis-classification rate and ensure accurate recognition of a person's behavioral context.

NAPS fusion framework with its ability to reduce model uncertainty in unbalanced large HAR dataset for multi-class classification lifts up recognition performance. Currently, this fusion technology can classify only 4 mutually exclusive activities at a time. As there is always room for betterment, this leaves a scope for future work.

Bibliography

[1] Nicholas J. Napoli, "Characterizing uncertainty in sensor fusion to improve predictive models". [Doctoral Dissertation]

[2] Y. Vaizman, K. Ellis and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE Pervasive Computing,* vol. 16, no. 4, pp. 62-74, 1 10 2017.

[3] K. Stawarz, A. L. Cox and A. Blandford, "Don't forget your pill!," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, New York, New York, USA, 2014.

[4] Y. Jia, "Dietetic and exercise therapy against diabetes mellitus," in *ICINIS 2009 - Proceedings of the 2nd International Conference on Intelligent Networks and Intelligent Systems*, 2009.

[5] J. Yin, Q. Yang and J. J. Pan, "Sensor-based abnormal human-activity detection," in *IEEE Transactions on Knowledge and Data Engineering*, 2008.

[6] Y. S. Lee and S. B. Cho, "Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011.

[7] H. J. Luinge, P. H. Veltink and C. T. Baten, "Estimating orientation with gyroscopes and accelerometers.," *Technology and health care : official journal of the European Society for Engineering and Medicine,* vol. 7, no. 6, pp. 455-9, 1999.

[8] B. J. Mortazavi, M. Pourhomayoun, G. Alsheikh, N. Alshurafa, S. I. Lee and M. Sarrafzadeh, "Determining the single best axis for exercise repetition recognition and counting on smartwatches," in *Proceedings - 11th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2014*, 2014.

[9] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach," in *3rd IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2016*, 2016.

[10] C. Zhao and Y. Wang, "A new classification method on information fusion of wireless sensor networks," in *Proceedings - The 2008 International Conference on Embedded Software and Systems Symposia, ICESS Symposia*, 2008.

[11] P. Tsinganos and A. Skodras, "On the Comparison of Wearable Sensor Data Fusion to a Single Sensor Machine Learning Technique in Fall Detection.," *Sensors (Basel, Switzerland),* vol. 18, no. 2, 14 2 2018.

[12] K. Liu, N. Z. Gebraeel and J. Shi, "A Data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Transactions on Automation Science and Engineering,* vol. 10, no. 3, pp. 652-664, 7 2013.

[13] A. Ross and R. Govindarajan, "Feature Level Fusion in Biometric Systems".

[14] A. H. Gunatilaka and B. A. Baertlein, "Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 577-589, 6 2001.

[15] I. Elmoudden, B. Benyacoub and S. Elbernoussi, "Modeling human activity recognition by dimensionality reduction approach," *Proceedings of the 27th International Business Information Management Association Conference - Innovation Management and Education Excellence Vision 2020: From Regional Development Sustainability to Global Economic Growth, IBIMA 2016,* no. May 2016, pp. 1800-1805, 2016.

[16] B. Krawczyk, *Learning from imbalanced data: open challenges and future directions,* vol. 5, Springer Verlag, 2016, pp. 221-232.

[17] R. Damaševičius, M. Vasiljevas, J. Šalkevičius and M. Woźniak, "Human activity recognition in AAL environments using random projections," *Computational and Mathematical Methods in Medicine,* vol. 2016, 2016.

[18] S. Pathical, "Classification in High Dimensional Feature Spaces through Random Subspace Ensembles," no. December, p. 233, 2010.

[19] T. Gu, T. Gu, H. K. Pung, D. Q. Zhang, H. K. Pung and D. Q. Zhang, "A Bayesian Approach For Dealing With Uncertain Contexts," *AUSTRIAN COMPUTER SOCIETY,* vol. 176, p. 2004, 2004.

[20] H. E. Byun and K. Cheverst, "Supporting Proactive ' Intelligent ' Behaviour : the Problem of Uncertainty," *Management,* 2003.

[21] A. P. Dempster, "Upper and Lower Probabilities Induced by a Multivalued Mapping," *The Annals of Mathematical Statistics,* vol. 38, no. 2, pp. 325-339, 4 1967.

[22] G. Shafer, A mathematical theory of evidence, Princeton University Press, 1976, p. 297.

[23] Senzhang Wang, Zhoujun Li, Wenhan Chao and Qinghua Cao, "Applying adaptive over-sampling technique based on data density and cost-sensitive SVM to imbalanced learning," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012.

[24] K. SENTZ and S. FERSON, "Combination of Evidence in Dempster-Shafer Theory," Albuquerque, NM, and Livermore, CA (United States), 2002.

[25] Y. Vaizman, N. Weibel and G. Lanckriet, "Context Recognition In-the-Wild," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* vol. 1, no. 4, pp. 1-22, 8 1 2018.

[26] C. Krupitzer, T. Sztyler, J. Edinger, M. Breitbach, H. Stuckenschmidt and C. Becker, "Beyond position-awareness—Extending a self-adaptive fall detection system," *Pervasive and Mobile Computing,* vol. 58, 1 8 2019.

[27] S. Veetil and Q. Gao, "Real-time Network Intrusion Detection Using Hadoop-Based Bayesian Classifier," in *Emerging Trends in ICT Security*, Elsevier Inc., 2013, pp. 281-299.

[28] R. Zhu and Z. Zhou, "A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 12, no. 2, pp. 295-302, 6 2004.

[29] W. Zhao, A. Bhushan, A. D. Santamaria, M. G. Simon and C. E. Davis, "Machine learning: A crucial tool for sensor design," *Algorithms,* vol. 1, no. 2, pp. 130-152, 12 2008.

[30] H. Zhao and Z. Wang, "Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion," *IEEE Sensors Journal,* vol. 12, no. 5, pp. 943-953, 2012.

[31] M. L. Zhang and Z. H. Zhou, *A review on multi-label learning algorithms,* vol. 26, IEEE Computer Society, 2014, pp. 1819-1837.

[32] J. Read, B. Pfahringer, G. Holmes and E. Frank, "Classifier chains for multi-label classification," *Machine Learning,* vol. 85, no. 3, pp. 333-359, 30 12 2011.

[33] J. Read and J. Read, "A pruned problem transformation method for multi-label classification," *IN: PROC. 2008 NEW ZEALAND COMPUTER SCIENCE RESEARCH STUDENT CONFERENCE (NZCSRS,* pp. 143--150, 2008.

[34] L. V. D. Maaten, L. Van Der Maaten, E. Postma and J. Van Den Herik, "Dimensionality reduction: A comparative review".

[35] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2011.

[36] R. Heideklang and P. Shokouhi, "Decision-level fusion of spatially scattered multi-modal data for nondestructive inspection of surface defects," *Sensors (Switzerland),* vol. 16, no. 1, 15 1 2016.

[37] A. N. (. N. Gorban', Principal manifolds for data visualization and dimension reduction, Springer, 2007, p. 334.

[38] N. Ghamrawi and A. McCallum, "Collective multi-label classification," in *International Conference on Information and Knowledge Management, Proceedings*, 2005.

[39] D. L. Donoho and D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS CONFERENCE ON MATH CHALLENGES OF THE 21ST CENTURY,* 2000.

[40] M. Di, E. M. Joo and L. H. Beng, "A comprehensive study of kalman filter and extended kalman filter for target tracking in wireless sensor networks," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2008.

[41] A. Clare and R. D. King, "Knowledge Discovery in Multi-label Phenotype Data," 2001, pp. 42-53.

[42] R. Bellman, Adaptive control processes : a guided tour.