



Universität Paderborn — Fakultät EIM-I
Fachgebiet Analytic Information Systems and Business
Intelligence
Jun.-Prof. Dr. Artus Krohn-Grimberghe



Bachelorarbeit

Erkennung körperlicher Aktivitäten mittels Smartphone- und Smartwatch-Sensoren und Machine Learning

Christian Brüggemann

03.04.2017

Betreut von:

Jun.-Prof. Dr. Artus Krohn-Grimberghe

Bachelorarbeit

am Fachgebiet Analytic Information Systems and Business Intelligence
Jun.-Prof. Dr. Artus Krohn-Grimberghe

Institut für Informatik
Fakultät für Elektrotechnik, Informatik und Mathematik
Universität Paderborn

Vorgelegt von:
Christian Brüggemann
Matrikelnummer: 7004878
Salbeiweg 39
33100 Paderborn

am
03.04.2017

Betreut durch:
Jun.-Prof. Dr. Artus Krohn-Grimberghe
Zweitgutachter:
Prof. Dr. Eyke Hüllermeier

TODO: FG-Logo ändern

Zusammenfassung

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Abstract

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Motivation	1
1.2. Ziele der Arbeit	2
2. Grundlagen	3
2.1. Erläuterung der verwendeten Sensoren	3
2.1.1. Beschleunigungssensor	3
2.1.2. Gyroskop	3
2.1.3. Hautwiderstandssensor	3
2.1.4. Hauttemperatursensor	3
2.1.5. Laufgeschwindigkeitssensor	4
2.2. Wichtige Begriffe	4
3. Review of the State of the Art	5
4. Experiment	7
4.1. Beschreibung der Aktivitäten	7
4.1.1. Allgemeine Aktivitäten (nicht handorientiert)	8
4.1.2. Allgemeine Aktivitäten (handorientiert)	8
4.1.3. Essaktivitäten (handorientiert)	9
5. Methode	11
5.1. Implementierung der Aufzeichnungssoftware	11
5.1.1. Definition der Messdaten	11
5.1.2. Struktur der Aufzeichnungssoftware	11
5.2. Transformation der Daten	11
5.2.1. Beschreibung der Aggregatfunktionen	12
5.2.2. Anwendung der Aggregatfunktionen	14
5.2.3. Beispiel	16
5.3. Anwendung von ML-Klassifikationsalgorithmen	17
5.3.1. Random Forest (RF)	18
5.3.2. J48-Entscheidungsbaum (J48)	18
5.3.3. Instance-Based-Learner (IBk)	18
5.3.4. Naive Bayes (NB)	18
5.3.5. Multilayer Perceptron (MLP)	18
6. Evaluation	21
6.1. Effektivität der Datenkombination	22

6.1.1. Persönliche Modelle	22
6.1.2. Unpersönliche Modelle	23
6.2. Genauigkeit bei ungenauen Zeitstempeln	23
6.3. Genauigkeit bei einer niedrigen Sampling-Rate	24
6.4. Genauigkeit bei Überlappung der Intervalle	26
6.5. Genauigkeit bei Verschmelzung der Ess- und Trinkaktivitäten	26
7. Conclusions	29
A. What goes in the appendices	33
B. Used Acronyms	35
C. Literaturverzeichnis	37

Abbildungsverzeichnis

5.1. Struktur der Aufzeichnungssoftware	12
6.1. Genauigkeit der persönlichen Modelle in Prozent	23
6.2. Genauigkeit der unpersönlichen Modelle in Prozent	23
6.3. Genauigkeit der Modelle mit verrauschten Band-Zeitstempeln in Prozent . . .	24
6.4. Sampling-Raten der Sensoren	25
6.5. Genauigkeit der Modelle mit reduzierter Sampling-Rate in Prozent	25
6.6. Genauigkeit der Modelle mit Intervallüberlappung in Prozent	26
6.7. Genauigkeit der Modelle bei Verschmelzung der Ess- und Trinkaktivitäten in Prozent	26

1. Einleitung

1.1. Motivation

Smartphones haben im letzten Jahrzehnt an großer Bedeutung gewonnen. **TODO: Cite** Daneben existieren mittlerweile ergänzend dazu sogenannte *Smartwatches* und *Fitness-Tracker*. Smartwatches sind Armbanduhren, die in der Regel drahtlos mit einem Smartphone verbunden sind und Informationen wie beispielsweise Benachrichtigungen am Handgelenk zugänglich machen. Fitness-Tracker besitzen ähnliche Funktionen, zielen allerdings primär darauf ab, die Fitness und Gesundheit des Nutzers zu fördern, indem Statistiken wie beispielsweise die Schrittzahl des Nutzers pro Tag gesammelt und grafisch aufbereitet werden. In beiden Geräteformen werden üblicherweise Sensoren verbaut, mit denen sich die Bewegungen des Trägers nachvollziehen lassen.

Mit einigen Fitness-Trackern des Unternehmens *Fitbit* existieren bereits kommerzielle Produkte, die über die reine Sammlung und grafische Aufbereitung von Statistiken hinausgehen: Die Funktion *SmartTrack* erkennt kontinuierliche Aktivitäten mit hoher Bewegung teilweise automatisch, ohne dass der Anwender vorher manuell einstellen muss, welcher Aktivität er in den nächsten Minuten nachgehen wird. **TODO: Cite https://help.fitbit.com/articles/en_US/Help_article/1933**. Dies hat den Vorteil, dass der Nutzer sich nicht daran erinnern muss, im Fitness-Tracker die richtige Aktivität einzustellen, um kategorisierte Statistiken zu erhalten.

SmartTrack unterstützt die folgenden Aktivitäten: Gehen, Laufen, Fahrradfahren, Crosstrainer-Training und Schwimmen, sowie zwei allgemeine Kategorien "Sport" (Fußball, Basketball, Tennis, etc.) und "aerobes Training" (Zumba, Tanzen).

Es existieren weitere mögliche Anwendungsgebiete der automatisierten Aktivitätenerkennung. Für Smartphone-Betriebssysteme könnte das Wissen, dass der Anwender gerade Sport treibt, interessant sein, um eingehende Anrufe eines nicht als wichtig markierten Kontaktes zu unterdrücken. Des Weiteren könnte das Forschungsgebiet der "Transportation Mode Recognition" von solchen Methoden profitieren: Soll erkannt werden, mit welchem Verkehrsmittel sich der Nutzer gerade fortbewegt, könnte neben dem Parameter der Geschwindigkeit ebenfalls von Interesse ein, ob mithilfe der Methode die Aktivität "Fahrradfahren" erkannt wird oder nicht. So ließe sich die Fortbewegung mittels eines Mofas von der Fortbewegung mittels eines Fahrrads unterscheiden, was insbesondere für Dienste wie "Google Now" nützlich sein könnte. Diese dienen dem Nutzer als persönlicher Assistent und warnen diesen beispielsweise vor Stau auf einer häufig befahrenen Strecke. Eine solche Warnung könnte entfallen, wenn festgestellt wurde, dass der Nutzer die Strecke nicht mit einem Mofa, sondern mit einem Fahrrad bewältigt.

In der Literatur ist [16] hervorzuheben. Die Autoren vergleichen in ihrem Paper die Genauigkeit der Aktivitätenerkennung eines Smartphones mit der einer Smartwatch und kommen zu dem Schluss, dass die Güte der jeweiligen Erkennung insbesondere von der Aktivität selbst abhängig ist. Es liegt auf der Hand, dass nur mithilfe eines Smartphones beispielsweise eine

Unterscheidung zwischen "Zähneputzen" und "Stehen" schwer möglich ist, während analog dazu nur mithilfe einer Smartwatch die Unterscheidung zwischen "Gehen" und "Fußball schießen" ebenfalls herausfordernd ist. Naheliegend ist daher, eine Kombination beider Datenquellen einzusetzen, um die durchschnittliche Erkennungsrate zu verbessern, ohne eine Beschränkung der erkennbaren Aktivitäten einzuführen.

1.2. Ziele der Arbeit

Evaluiert werden soll ein zu entwickelndes Verfahren, das auf Basis von *maschinellern Lernen* (siehe Definition 1) und eben jenen gesammelten Daten feststellt, welcher Aktivität der Träger der Geräte in bestimmten Zeitintervallen nachgegangen ist. Hierzu wird zunächst eine Software benötigt, welche die synchrone Aufzeichnung von Sensordaten eines Fitness-Trackers oder einer Smartwatch und eines Smartphones ermöglicht.

Es ergibt sich insbesondere die Frage, inwiefern sowohl personalisierte, das heißt nutzerspezifische, als auch unpersonalisierte Modelle durch die Hinzunahme einer weiteren Datenquelle genauer werden.

Um eine Evaluation zu ermöglichen, wird ein Beispieldatensatz benötigt, der durch ein Experiment mit 10 Probanden aufgebaut wird. Orientiert ist diese Zahl an der Anzahl der Probanden in [16], an dessen Experiment 17 Probanden teilgenommen haben. Im Experiment sollen diese voneinander unabhängig mehreren definierten Aktivitäten nachgehen, während parallel dazu Sensordaten mithilfe der entwickelten Software aufgezeichnet werden. Um die Vergleichbarkeit mit den Ergebnissen aus [16] zu gewährleisten, werden die Probanden in diesem Experiment denselben Aktivitäten nachgehen.

Im Folgenden beschränken wir uns auf den uns zugänglichen Fitness-Tracker *Microsoft Band 2*, da dieser im Vergleich zur ebenfalls vorhandenen Smartwatch *Pebble Time* mehr Sensoren besitzt und letztere während der Durchführung des Experimentes nach einem erzwungenen Software-Update falsche Zeitstempel für Sensordaten lieferte.

2. Grundlagen

2.1. Erläuterung der verwendeten Sensoren

TODO: Grafiken zur Erläuterung Während in der Einleitung nur generisch von "Sensordaten" die Rede war, werden diese nun konkretisiert. Zur Aufnahme werden das Smartphone OnePlus 3 sowie der Fitness-Tracker Microsoft Band 2 verwendet. Beide Geräte besitzen einen Beschleunigungssensor sowie ein Gyroskop. Des Weiteren besitzt das Band unter anderem einen Sensor, der den elektrischen Widerstand der Haut des Trägers misst, sowie einen Hauttemperatursensor.

2.1.1. Beschleunigungssensor

Ein Beschleunigungssensor, auch *Accelerometer* genannt, misst die echte Beschleunigung eines Objektes im Raum je physikalischer Achse (x , y und z). Dies beinhaltet auch die Beschleunigung, die von der Erdgravitation ausgeht: Liegt das Objekt beispielsweise auf dem Boden, erfährt es auf der im rechten Winkel zum Boden stehenden Achse eine Beschleunigung von $g \approx 9.81 \text{ m/s}^2$ [6][15]. Über diesen Sensor kann demnach die Orientierung des Objektes im Raum bestimmt werden.

Die Daten dieses Sensors werden von beiden Geräten aufgezeichnet.

2.1.2. Gyroskop

Ein Gyroskop misst die Rotationsgeschwindigkeit je physikalischer Achse[6]. Demnach bedeutet der Messwert ($x = 0, y = 0, z = 0$), dass das Objekt relativ zur Erde nicht bewegt wird, während ($x = 1, y = 0, z = 0$) bedeutet, dass das Objekt um die x -Achse gedreht wird. Die Daten dieses Sensors werden von beiden Geräten aufgezeichnet.

2.1.3. Hautwiderstandssensor

Je nach aktueller körperlicher Betätigung ändert sich die elektrische Leitfähigkeit der Hautoberfläche durch Schweiß. Der Hautwiderstandssensor misst den elektrischen Widerstand der Haut, der sich umgekehrt proportional zur Leitfähigkeit verhält. Dieser Sensor ist nur im Band integriert, das daher für diesen die einzige Datenquelle ist.

2.1.4. Hauttemperatursensor

Der Hauttemperatursensor misst die Temperatur der Haut des Nutzers, die an der Unterseite des Band anliegt. Leider reagierte dieser Sensor in einem Test nur langsam auf Temperaturveränderungen: Nachdem der Tracker abgelegt wurde, übermittelte dieser den Temperaturabfall

erst mehrere Sekunden später. Aus diesem Grund wurde auf die Aufzeichnung dieser Daten verzichtet.

2.1.5. Laufgeschwindigkeitssensor

Der Laufgeschwindigkeitssensor ist ein virtueller Sensor, der auf Basis des elektronischen Schrittzählers im Microsoft Band 2 generiert wird. Er gibt die Laufgeschwindigkeit des Trägers in cm/s an.

2.2. Wichtige Begriffe

In den folgenden Kapiteln wird von einem grundlegenden Verständnis von Machine Learning-Klassifikation ausgegangen.

Definition 1 (ML-Klassifikation). *Gegeben sei ein Datensatz $D \ni (x_1, \dots, x_n, y)$ mit $|D| = m$ Instanzen. Dann sind $x = (x_1, \dots, x_n)$ die Input-Features und y das Target. Im Falle eines Klassifikationsproblems ist y nominal, d.h. ein deskriptiver Wert zur Identifikation. Gesucht ist nun eine Hypothesenfunktion $h(x) = \hat{y}$, die zu gegebenen Features den wahrscheinlichsten Wert des Targets bestimmt [13].*

Neben Definition 1 sind für das Verständnis dieser Arbeit die folgenden Begriffe wichtig:

1. (Überwachtes) Lernen, bzw. Training. Definiert in Kapitel 2.1 [7].
2. Overfitting und Train-Test-Split. Definiert und erläutert in Kapitel 7.2 [7].
3. Kreuzvalidierung. Definiert und erläutert in Kapitel 7.10 [7].

3. Review of the State of the Art

TODO: Review, grouped by idea not author

3. Review of the State of the Art

4. Experiment

In diesem Kapitel werden Aufbau und Durchführung des Experiments erläutert, durch das der Datensatz zusammengestellt wurde.

Insgesamt haben 10 Teilnehmer am Experiment teilgenommen und die in Abbildung 4.1 achtzehn gelisteten Aktivitäten durchgeführt. Weiss et al. nahmen jeweils zwei Minuten pro Teilnehmer und Aktivität auf [16], mussten allerdings zum Anfang und Ende jeder Aufnahme je zehn Sekunden abschneiden, um Ausreißer zu entfernen. Aus diesem Grund betrug die Dauer jeder Aufnahme in diesem Experiment drei Minuten, wodurch der gesamte Aufnahmeprozess pro Person rund zwei Stunden dauerte. Vor dem Experiment wurden Einverständniserklärungen der Teilnehmer eingeholt.

Die Aufnahme erfolgte mit dem Fitness-Tracker Microsoft Band 2 und dem Smartphone OnePlus 3, auf dem das Betriebssystem Android 6 installiert war. Der Fitness-Tracker wurde am Handgelenk des Teilnehmers befestigt, während das Smartphone in seiner Hosentasche platziert wurde, wie in [16] jeweils auf der dominanten Seite des Teilnehmers. Dies war notwendig, da insbesondere Aktivitäten wie das Dribbeln eines Basketballs mit der dominanten Hand durchgeführt werden. Während dies für Armbanduhren nicht notwendigerweise eine realistische Konfiguration ist, da diese üblicherweise auf der nicht-dominanten Seite getragen werden, besteht dieses Problem bei Fitness-Trackern nicht unbedingt. So lässt sich beispielsweise in Fitness-Trackern der Firma Fitbit einstellen, auf welcher Seite dieser getragen wird, was die Vermutung nahelegt, dass genügend Nutzer mit dem Tragen auf dieser Seite kein Problem haben [5].

Auf dem Smartphone wurde eine eigens für das Experiment entwickelte Anwendung ausgeführt, in der zunächst der Name des Teilnehmers eingegeben wurde. Sowohl auf dem Smartphone selbst als auch per drahtloser Fernsteuerung wurde die durchzuführende Aktivität eingestellt. Gestartet und gestoppt wurde die Aufnahme anschließend per Fernsteuerung, um Ausreißer am Anfang und Ende der Aufnahme zu vermeiden, obgleich trotzdem jeweils zehn Sekunden abgeschnitten wurden.

4.1. Beschreibung der Aktivitäten

Im Folgenden werden die einzelnen Aktivitäten, die von den Teilnehmern ausgeführt wurden, genauer beschrieben. Um eine Vergleichbarkeit mit [16] zu ermöglichen, handelt es sich mangels detaillierter Beschreibungen zumindest um ähnliche Aktivitäten, die dieselben Bezeichnungen haben.

4.1.1. Allgemeine Aktivitäten (nicht handorientiert)

Gehen

Der Teilnehmer bewegt sich im Schrittempo auf einem Gehweg.

Joggen

Der Teilnehmer joggt auf einem Gehweg. Das Tempo variiert je nach sportlicher Verfassung des Teilnehmers.

Treppensteigen

Der Teilnehmer bewegt sich abwechselnd eine Treppe hoch unter herunter. Steigung und Länge sind variabel.

Sitzen

Der Teilnehmer sitzt auf einem Stuhl und versucht, sich dabei nicht unruhig zu verhalten.

Stehen

Der Teilnehmer steht und versucht, sich dabei nicht unruhig zu verhalten.

Fußball schießen

Der Teilnehmer schießt einen Fußball wiederholt mit mittlerer Kraft zu einem Mitspieler und versucht, Sprinten zu vermeiden.

4.1.2. Allgemeine Aktivitäten (handorientiert)

Basketball dribbeln

Der Teilnehmer schleudert einen Basketball wiederholt Richtung Boden und versucht, dabei möglichst an einer Stelle stehen zu bleiben.

Mit einem Tennisball Fangen spielen

Zwei Personen werfen sich abwechselnd einen Tennisball zu und versuchen dabei, möglichst an einer Stelle stehen zu bleiben.

Auf einer Tastatur tippen

Der Teilnehmer tippt sitzend seinen Gewohnheiten nach einen Text an einer beliebigen Computertastatur ab. Mindestens grobe Fehler sollten korrigiert werden. Um Verständnisproblemen aus dem Weg zu gehen, handelt es sich bei dem Text um ein Diktat für Siebtklässler.

Auf Papier schreiben

Der Teilnehmer schreibt sitzend seinen Gewohnheiten nach denselben Text wie mit der Tastatur mit einem Kugelschreiber auf ein Blatt Papier im Format DIN A4 ab.

Klatschen

Der Teilnehmer begleitet sitzend das Lied "Viva la Vida" von Coldplay klatschend.

Zähneputzen

Der Teilnehmer benutzt stehend eine Handzahnbürste (nicht elektrisch), um sich damit die Zähne zu putzen.

Kleidung falten

Der Teilnehmer faltet stehend der Reihe nach T-Shirts auf einem Tisch.

4.1.3. Essaktivitäten (handorientiert)

Spaghetti essen

Der Teilnehmer isst Spaghetti mit einer Soße.

Suppe essen

Der Teilnehmer isst eine Suppe seiner Wahl.

Brot essen

Der Teilnehmer isst ein belegtes Brot.

Chips essen

Der Teilnehmer isst Chips aus einer handelsüblichen Tüte.

Aus einer Tasse oder einem Glas trinken

Der Teilnehmer trinkt wiederholt aus einer Tasse oder einem Glas.

4. *Experiment*

5. Methode

5.1. Implementierung der Aufzeichnungssoftware

5.1.1. Definition der Messdaten

Definition 2. Ein *Reading* besteht aus einem Unix-Zeitstempel in Millisekunden, einer Gerätequelle, einem Sensortyp und den Werten des Sensors gruppiert nach Komponente. Ein Beispiel ist $(1000, \text{Band}, \text{Gyroscope}, \{x \rightarrow 0.0, y \rightarrow 1.0, z \rightarrow 1.0\})$.

5.1.2. Struktur der Aufzeichnungssoftware

Die Aufzeichnungssoftware wurde als Anwendung für das Android-Betriebssystem implementiert, da für diese Plattform auch ein *Software Development Kit (SDK)* für das Microsoft Band 2 existiert.

Abbildung 5.1 bietet einen Überblick über die Struktur der Aufzeichnungssoftware. Im Zentrum steht die Klasse *CollectionService*, die für die robuste Sammlung der Daten verantwortlich ist und daher als persistenter Android-Service implementiert ist, der im Gegensatz zu anderen Komponenten einer Android-App nicht ohne weiteres durch das System zur Schonung der Ressourcen beendet werden kann. Gesteuert wird die Aufnahme über eine grafische Benutzeroberfläche auf dem Smartphone selbst (*AndroidUI*), sowie optional auch über eine Weboberfläche, die über einen einfachen HTTP-Server zur Verfügung steht. In den Benutzeroberflächen kann der Name des Probanden, sowie die ausgeführte Aktivität angegeben werden. Des Weiteren wird hierüber die Aufnahme gestartet und gestoppt.

Der *CollectionService* delegiert die Aufnahme an den *BandDataRecorder* und den *LocalDataRecorder* weiter, die jeweils für das Microsoft Band 2 respektive die Smartphone-lokalen Sensoren verantwortlich sind. Über den *BandConnector* wird die Bluetooth-Verbindung zum Band hergestellt und aufrecht erhalten. *BandDataRecorder* und *LocalDataRecorder* zeichnen Daten in Form von *Readings* auf. Die Methode *getValuesPerComponent()* liefert beispielsweise Daten der Form $\{x \rightarrow 1.0, y \rightarrow 2.0, z \rightarrow 3.0\}$, wenn die Komponenten des jeweiligen Sensors der jeweiligen *Source* $\{x, y, z\}$ sind.

Wird die Aufnahme gestoppt, werden die Readings mitsamt den Metadaten Name des Probanden und Aktivität serialisiert und per SFTP auf einen entfernten Speicher hochgeladen. Anschließend werden die Daten von dort abgerufen und wie im folgenden Abschnitt beschrieben weiterverarbeitet.

5.2. Transformation der Daten

Jede einzelne Aufnahme eines Nutzers und einer Aktivität besteht aus einer Reihe von Daten mit Zeitstempeln. In dieser Form kann noch kein konventioneller ML-Klassifikationsalgorithmus

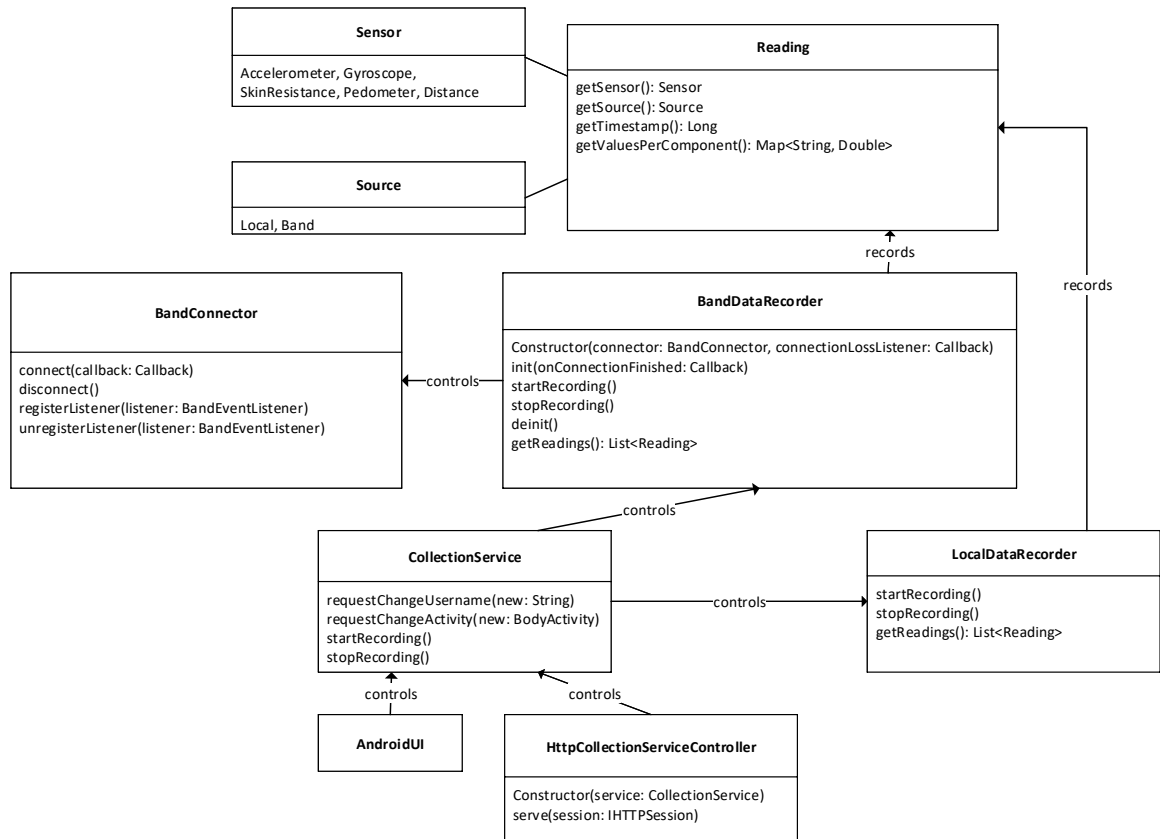


Abbildung 5.1.: Struktur der Aufzeichnungssoftware

mit den Daten arbeiten, da diese Daten in Form einer Liste von *Instanzen* erwarten. Eine Instanz besteht aus einer Menge von *Attributen/Features*, die in unserem Fall bis auf das nominale *Target*-Attribut allesamt kardinal sind. Das Target-Attribut ist dabei die durchgeführte Aktivität. Entsprechend ist eine Transformation der Daten notwendig, um diese tatsächlich nutzbar zu machen.

Um Instanzen zu erzeugen, teilt die Transformationssoftware die Ursprungsdaten zunächst in Intervalle mit zehnständiger Dauer auf. Ein solches Intervall soll schließlich eine Instanz, d.h. ein Beispiel für eine Aktivität bilden. Das erste und das letzte Intervall werden gelöscht, da in dieser Zeit die Aufnahme gestartet und gestoppt wurde, was ansonsten durch den anderen Bewegungsablauf die Ergebnisse verfälschen könnte. Aktuell umfasst das Intervall noch mehrere Readings, die aggregiert werden müssen.

5.2.1. Beschreibung der Aggregatfunktionen

Sei $A[1..N]$ eine Liste von Zahlen. Dann seien die folgenden Aggregatfunktionen in Anlehnung an Kwapisz et al.[11] wie folgt definiert:

Average

$$\text{Avg}(A) := \frac{1}{N} * \sum_{a \in A} a$$

Standard Deviation

$$\text{StdDev}(A) := \sqrt{\frac{1}{N} * \sum_{a \in A} (a - \text{Avg}(A))^2}$$

Average Absolute Difference

$$\text{AvgAbsDiff}(A) := \frac{1}{N} * \sum_{a \in A} |\text{Avg}(A) - a|$$

 k -Binned Distribution

Im Gegensatz zu den obigen Funktionen liefert diese Aggregatfunktion einen k -Vektor anstatt einen Skalar. Seien $i \in \{0, \dots, k-1\}$, $S := \frac{\max A - \min A}{k}$.

$$\text{Bin}(A, i) := \#\{a \in A \mid (\min A) + i * S \leq a < (\min A) + (i+1) * S\}$$

$\text{Bin}(A, i)$ ist demzufolge die Anzahl der Elemente in Korb i , wenn man A der Größe nach sortiert auf k Körbe verteilt. Der Bin ist hierbei als Multimenge zu verstehen und darf somit Elemente auch mehrfach enthalten.

Average Root of Squares

Auch diese Aggregatfunktion unterscheidet sich von den bisherigen, da sie mehrere Listen als Parameter erhält:

$$\text{Aros}(A_1, \dots, A_M) = \frac{1}{M} * \sum_{i=1}^M \sqrt{\sum_{a \in A_i} a^2}$$

Average Time between Peaks

Diese Aggregatfunktion gibt die durchschnittliche Zeit zwischen Höhepunkten in A zurück, wofür zusätzlich eine Funktion $\text{timeForIndex} : \mathbb{N} \rightarrow \mathbb{N}$ benötigt wird. Die Wahl der Höhepunkte erfolgt durch eine Heuristik.

Algorithmus 2 beschreibt die Berechnung des Wertes. Zunächst wird mittels Algorithmus 1 bestimmt, an welchen Indizes in der Liste Werte stehen, die deutlich größer als ihre Nachfolger sind. Dadurch wird definiert, was eine Spitze in dieser spezifischen Liste ausmacht. Anschließend werden Spitzen gesucht, die maximal um den Faktor *threshold* kleiner sind als die größte Spitze, die der Algorithmus gefunden hat. Diese Grenze wird solange gesenkt, bis genügend Spitzen gefunden wurden oder die Grenze so niedrig liegt, dass es sich nicht mehr um Spitzen handelt. Als letztes berechnet der Algorithmus die durchschnittliche Zeit zwischen den Spitzen mithilfe der Funktion *timeForIndex*.

Die im Pseudocode verwendeten Konstanten erwiesen sich im praktischen Test an den von uns gesammelten Daten als hinreichend. Für andere Datensätze kann eine Anpassung erforderlich sein.

Algorithm 1 IndicesOfPeaks(A, t), $t \in [0, 1]$

▷ Returns a list of indices i where $A[i] * t \geq A[i + 1]$

```

indices  $\leftarrow \emptyset$ 
for  $i \in \{1, \dots, |A| - 1\}$  do
  if  $A[i] * t \geq A[i + 1]$  then
    indices  $\leftarrow$  indices  $\cup \{i\}$ 
  end if
end for
return indices

```

5.2.2. Anwendung der Aggregatfunktionen

Die Transformationssoftware reduziert nun mithilfe der Aggregatfunktionen die Readings eines jeden Intervalls I auf eine konstante Anzahl skalarer Werte, welche die Features der Instanz darstellen. Im Folgenden sei I ein Array der Struktur $I[G][S][K] \in \mathbb{R}^\alpha$, das die Messungen gruppiert nach Gerät (G), Sensor (S) und Komponente (K) beinhaltet. Des Weiteren sei F_I die Menge der Features des Intervalls I . Die Folgenden Unterabschnitte zeigen den Aufbau von F_I mittels der Aggregatfunktionen:

Pro Kombination von Gerät G und Sensor S

Seien P die einzelnen Komponenten des Sensors, beispielsweise $P = \{x, y, z\}$.

$$\begin{aligned}
 \text{aros} &\leftarrow \text{Aros}(I[G][S][P[1]], \dots, I[G][S][P[|P|]]) \\
 F_I &\leftarrow F_I \cup \{((G, S), \text{aros})\}
 \end{aligned}$$

Im Kontext betrachtet liefert *Aros* am Beispiel des Beschleunigungssensors gewissermaßen die durchschnittliche Beschleunigung, die alle Achsen zusammen erfahren haben.

Pro Kombination von Gerät G , Sensor S und Komponente K

Sei *timeForIdx* hier eine Funktion, die den Zeitstempel des jeweiligen Datums im Intervall zurückgibt. Sei außerdem $k := 10$, sodass 10 Körbe verwendet werden, was sich experimentell als geeignet herausstellte.

$$\begin{aligned}
 F_I &\leftarrow F_I \cup \{((G, S, K), \text{Avg}(I[G][S][K]))\} \\
 F_I &\leftarrow F_I \cup \{((G, S, K), \text{StdDev}(I[G][S][K]))\} \\
 F_I &\leftarrow F_I \cup \{((G, S, K), \text{AvgAbsDiff}(I[G][S][K]))\} \\
 F_I &\leftarrow F_I \cup \{((G, S, K), \text{AverageTimeBetweenPeaks}(I[G][S][K], \text{timeForIdx}, 3))\} \\
 F_I &\leftarrow F_I \cup \{((G, S, K), \text{Bin}(I[G][S][K], i))\} \forall i \in \{1, \dots, k\}
 \end{aligned}$$

Algorithm 2 AverageTimeBetweenPeaks($A, \text{timeForIndex}, \text{minPeaks}$)

▷ *First, heuristically define what a peak is*
indicesOfPeaks \leftarrow IndicesOfPeaks($A, t = 0.8$)
if indicesOfPeaks = \emptyset **then**
 return Nil
end if
highestPeakIdx \leftarrow Index of highest peak in indicesOfPeaks

▷ *Lower the threshold until we have found enough peaks or the threshold is too low*
otherPeakIndices $\leftarrow \emptyset$
threshold $\leftarrow 0.8$
repeat
 otherPeakIndices \leftarrow Indices of A where $A[i] \geq A[\text{highestPeakIdx}] * \text{threshold}$
 threshold $\leftarrow \text{threshold} * 0.9$
until $|\text{otherPeakIndices}| \geq \text{minPeaks} \vee \text{threshold} < 0.3$

▷ *Check whether we have found enough peaks*
if $|\text{otherPeakIndices}| < \text{minPeaks}$ **then**
 return Nil
end if

▷ *Calculate the average time between the peaks*
timesBetweenPeaks $\leftarrow \emptyset$
for $i \in \{2, \dots, |\text{otherPeakIndices}|\}$ **do**
 time $\leftarrow \text{timeForIndex}(\text{otherPeakIndices}[i]) - \text{timeForIndex}(\text{otherPeakIndices}[i - 1])$
 timesBetweenPeaks $\leftarrow \text{timesBetweenPeaks} \cup \{\text{time}\}$
end for
return AVG(timesBetweenPeaks)

5.2.3. Beispiel

Wir betrachten für die Aktivität *Jogging* den Beispieldatensatz

$$D = \{(0, \text{Band}, \text{Gyroscope}, x \rightarrow 1.0, y \rightarrow 0.0), \\ (0, \text{Band}, \text{Accelerometer}, x \rightarrow 1.0, y \rightarrow 1.0), \\ (1000, \text{Band}, \text{Gyroscope}, x \rightarrow 2.0, y \rightarrow 2.0), \\ (1000, \text{Band}, \text{Accelerometer}, x \rightarrow 0.0, y \rightarrow 1.0), \\ (2000, \text{Band}, \text{Gyroscope}, x \rightarrow 3.0, y \rightarrow 0.0), \\ (2000, \text{Band}, \text{Accelerometer}, x \rightarrow 1.0, y \rightarrow 3.0)\}$$

Zur Vereinfachung wird von 2D-Sensoren ausgegangen, sodass keine z -Komponente vorhanden ist. Zur Instanzerzeugung verwenden wir als Aggregatfunktionen den Durchschnitt, die *Binned Distribution* mit zwei *Bins* und den *Average Root of Squares*. Die gewählte Intervallgröße beträgt zwei Sekunden, sodass sich die folgenden Intervalle ergeben:

$$I_1 = \{(0, \text{Band}, \text{Gyroscope}, x \rightarrow 1.0, y \rightarrow 0.0), \\ (0, \text{Band}, \text{Accelerometer}, x \rightarrow 1.0, y \rightarrow 1.0), \\ (1000, \text{Band}, \text{Gyroscope}, x \rightarrow 2.0, y \rightarrow 2.0), \\ (1000, \text{Band}, \text{Accelerometer}, x \rightarrow 0.0, y \rightarrow 1.0)\},$$

$$I_2 = \{(2000, \text{Band}, \text{Gyroscope}, x \rightarrow 3.0, y \rightarrow 0.0), \\ (2000, \text{Band}, \text{Accelerometer}, x \rightarrow 1.0, y \rightarrow 3.0)\}$$

Nun werden die Aggregatfunktionen auf I_1 angewendet:

$$\text{Avg}_{\text{Band, Gyroscope, x}} = (1 + 2)/2 = 1.5$$

$$\text{Avg}_{\text{Band, Gyroscope, y}} = (0 + 2)/2 = 1$$

$$\text{Avg}_{\text{Band, Accelerometer, x}} = (1 + 0)/2 = 0.5$$

$$\text{Avg}_{\text{Band, Accelerometer, y}} = (1 + 1)/2 = 1$$

$$\text{Bin}_{\text{Band, Gyroscope, x}}(0) = \#\{1\} = 1$$

$$\text{Bin}_{\text{Band, Gyroscope, x}}(1) = \#\{2\} = 1$$

$$\text{Bin}_{\text{Band, Gyroscope, y}}(0) = \#\{0\} = 1$$

$$\text{Bin}_{\text{Band, Gyroscope, y}}(1) = \#\{2\} = 1$$

$$\text{Bin}_{\text{Band, Accelerometer, x}}(0) = \#\{0\} = 1$$

$$\text{Bin}_{\text{Band, Accelerometer, x}}(1) = \#\{1\} = 1$$

$$\text{Bin}_{\text{Band, Accelerometer, y}}(0) = \#\{0\} = 0$$

$$\text{Bin}_{\text{Band, Accelerometer, y}}(1) = \#\{1\} = 0$$

$$\text{AroS}_{\text{Band, Gyroscope}} = (1/2) * (\underbrace{\sqrt{1^2 + 2^2}}_x + \underbrace{\sqrt{0^2 + 2^2}}_y) \approx 2.12$$

$$\text{AroS}_{\text{Band, Accelerometer}} = (1/2) * (\underbrace{\sqrt{1^2 + 0^2}}_x + \underbrace{\sqrt{1^2 + 1^2}}_y) \approx 1.2$$

Die Berechnung für I_2 erfolgt analog. Aus den obigen Features ergibt sich der folgende Datensatz, der von einem herkömmlichen Klassifikationsalgorithmus verarbeitet werden kann:

Intervall	$\text{Avg}_{\text{Band, Gyroscope, x}}$	$\text{Avg}_{\text{Band, Gyroscope, y}}$...	sampleClass
I_1	1.5	1	...	Jogging
I_2	Jogging

Die Intervallspalte dient nur der Illustration.

5.3. Anwendung von ML-Klassifikationsalgorithmen

Um die Vergleichbarkeit mit [16] sicherzustellen, wurde die Java-basierte ML-Software WEKA zum maschinellen Lernen verwendet. Diese bietet fertige Implementierungen diverser ML-Algorithmen an, die sowohl mittels einer grafischen Benutzeroberfläche, als auch in Code verwendet werden können. Die in dieser Arbeit verwendeten Algorithmen entsprechen denen aus [16] und werden im Folgenden kurz beschrieben. Falls nicht anders angegeben, wurden für jedes Modell die standardmäßigen Hyperparameter verwendet. Dabei handelt es sich um Parameter, die der jeweilige Algorithmus nicht selbst lernen kann. **TODO: Hyperparameter aktualisieren, falls nötig**

5.3.1. Random Forest (RF)

Ein *Random Forest* ist ein Wald von Entscheidungsbäumen (erklärt in Kapitel 14.4 [3]), die randomisiert entstanden sind. Soll eine neue Instanz klassifiziert werden, treffen alle Entscheidungsbäume eine Entscheidung und die am häufigsten vorhergesagte Klasse gewinnt. WEKA nutzt eine Implementierung nach [4]. Experimentell haben sich die folgenden Hyperparameterwerte als sinnvoll erwiesen: Die Anzahl der Bäume in Wald wurde auf 50, die Anzahl der verwendeten Features pro Baum auf 10, sowie die maximale Tiefe auf 25 festgelegt.

5.3.2. J48-Entscheidungsbaum (J48)

J48 ist eine Implementierung des C4.5-Algorithmus[14]. Der Entscheidungsbaum wird anhand von Trainingsdaten T iterativ durch das Hinzufügen von Entscheidungsknoten erzeugt, die den Informationsgehalt der Teilmengen von T minimieren, die aus dem Split an jenem Entscheidungsknoten hervorgehen. Der Informationsgehalt umso höher, je mehr verschiedene Klassen sich in einer Menge von Instanzen befinden. Ein niedriger Informationsgehalt besagt demnach, dass nach einer Entscheidung an einem Entscheidungsknoten klarer geworden ist, welcher Klasse die Instanzen in den Mengen nach dem Split zuzuordnen sind.

5.3.3. Instance-Based-Learner (IBk)

Der IBk-Algorithmus ist ein k NN-basierter Klassifikationsalgorithmus. Soll einer neuen Instanz $x \in \mathbb{R}^n$ eine Klasse zugeordnet werden, werden anhand eines Distanzmaßes für den Vektorraum \mathbb{R}^n die k nächsten Nachbarn von x gesucht. x wird danach die Klasse zugeordnet, die unter den k Nachbarn am häufigsten vorkam. Weitere Details sind zu finden in [1]. k ist dabei ein Hyperparameter, der auf 3 eingestellt wurde.

5.3.4. Naive Bayes (NB)

Naive Bayes ist eine probabilistische Methode zur Klassifikation[9]. Mit $x = (x_1, \dots, x_n)$ wird die Klasse C gesucht, die $\mathbb{P}(C|x_1, \dots, x_n)$ maximiert. x besteht hier aus nominalen Attributen. Der Satz von Bayes besagt, dass $\mathbb{P}(C|x) = \frac{\mathbb{P}(C)\mathbb{P}(x|C)}{\mathbb{P}(x)}$. Da über C optimiert wird, kann der Nenner für diesen Ansatz ignoriert werden. Nimmt man *naiv* an, dass x_1, \dots, x_n voneinander unabhängig sind, kann somit statt $\mathbb{P}(C|x)$ auch folgender Term über C maximiert werden: $\mathbb{P}(C) \prod_{i=1}^n \mathbb{P}(x_i|C)$. Durch den Trainingsdatensatz sind $\mathbb{P}(C)$ als Prior der Klasse C , sowie alle $\mathbb{P}(x_i|C)$ bekannt, sodass der Term durch einfaches Ausprobieren aller Klassen maximiert werden kann.

Enthält x numerische Attribute, müssen diese zunächst durch nominale Attribute ersetzt werden. Der Wertebereich aller x_i über den Trainingsdatensatz hinweg kann in eine feste Anzahl von Intervallen aufgeteilt werden. Für jedes Intervall I wird nun eine Indikatorvariable als Feature angelegt, die angibt, ob für eine Instanz x gilt, dass $x_i \in I$.

5.3.5. Multilayer Perceptron (MLP)

Bei dieser Methode handelt es sich um ein mehrschichtiges Netzwerk von sogenannten *Perzeptren*, auch (*künstliches*) *neuronales Netzwerk* genannt. Ein Perzeptron erhält n Eingaben

x_1, \dots, x_n und gewichtet diese mit Parametern $\Theta_1, \dots, \Theta_n$, die gelernt werden. Das Perzeptron berechnet $y = g(\sum_{i=1}^n \Theta_i x_i)$ als Ausgabewert, wobei g eine Aktivierungsfunktion ist, die den Ausgabewerte beschränkt, beispielsweise die Sigmoidfunktion.

Die Ausgabe eines Perzeptron kann einem Perzeptron der nächsten Schicht als Eingang dienen. Die letzte Schicht fungiert als Ausgabe des Netzwerks, dessen Fehler anhand eines Trainingsdatensatzes bestimmt werden kann. Jedes Perzeptron berechnet, wie die eigenen Parameter Θ angepasst werden müssen, um den Fehler zu minimieren, wofür der Fehler der Perzeptren der letzten Schicht an die vorherige Schicht weiter propagiert wird (*Backpropagation*). Zur Optimierung der Parameter wird ein stochastischer Gradientenabstieg verwendet. Zur Klassifikation können die Ausgangswerte der Perzeptren der letzten Schicht als Indikatorvariablen für die vorhandenen Klassen interpretiert werden. Weitere Informationen sind zu finden in [8].

6. Evaluation

Dieses Kapitel widmet sich der Auswertung der Ergebnisse anhand verschiedener Metriken. Eingebettet in der Transformationssoftware befindet sich ein Präprozessor, der die aufgenommen Daten vor der Transformation manipulieren kann. Auf diese Weise werden mehrere Trainingsdatensätze erstellt, die in der Evaluation daraufhin analysiert werden, wie genau ein Klassifikator nach dem Training mit ihnen Vorhersagen treffen kann.

Die folgenden Datensätze werden generiert:

1. *Combined*: Alle Daten werden ohne Veränderungen mit in die Transformation einbezogen.
2. *NoisyBandTimestamps*: Die Zeitstempel der Readings des Bands werden mit gaussischem Rauschen versehen
3. *Band combined*: Alle Daten des Smartphones werden vor der Transformation verworfen, sodass nur noch die Daten des Band verbleiben..
4. *Band accel*: Nur die Daten des Beschleunigungssensors des Microsoft Band 2 werden mit in die Transformation einbezogen.
5. *Band gyro*: Nur die Daten des Gyroskops des Microsoft Band 2 werden mit in die Transformation einbezogen.
6. *Phone comb.*: Alle Daten des Microsoft Band 2 werden vor der Transformation verworfen, sodass nur noch die Daten des Smartphones verbleiben.
7. *Phone accel*: Nur die Daten des Beschleunigungssensors des Smartphones werden mit in die Transformation einbezogen.
8. *Phone gyro*: Nur die Daten des Gyroskops des Smartphones werden mit in die Transformation einbezogen.
9. *SamplingRate1Hz*: Es werden Readings verworfen, als hätten Band und Smartphone jeweils nur Readings bei einer Rate von 1 Hz geliefert.
10. *SamplingRate5Hz*: Es werden Readings verworfen, als hätten Band und Smartphone jeweils nur Readings bei einer Rate von 5 Hz geliefert.

TODO: Genauigkeiten aus Weiss zum Vergleich mit einbinden, side by side in einem diagramm

6.1. Effektivität der Datenkombination

In diesem Abschnitt widmen wir uns der Frage, wie effektiv die Kombination der Daten von Band und Smartphone hinsichtlich der Vorhersagegenauigkeit gegenüber einer einzigen Datenquelle ist. Weiss et al.[16] werten in ihrem Papier die Genauigkeit von Modellen aus, die entweder auf Daten des Beschleunigungssensors eines Smartphones, des Beschleunigungssensors einer Smartwatch oder des Gyroskops einer Smartwatch zurückgreifen. Eine Kombination der Daten schlagen die Autoren vor, führen diese allerdings nicht durch.

Abbildung 6.1 zeigt die Genauigkeit der persönlichen Modelle und Abbildung 6.2 die der unpersönlichen Modelle. Ein persönliches Modell basiert auf Daten des jeweiligen Benutzers, während ein unpersönliches Modell noch keine Daten des Nutzers gesehen hat, für den eine Vorhersage getroffen werden soll. Die *Genauigkeit* ist definiert als der Anteil der korrekt klassifizierten Instanzen, die dem Modell zum Test vorgelegt wurden.

Die erste Spalte der in diesem Abschnitt referenzierten Tabellen gibt den Algorithmus an, der zur Klassifikation verwendet wurde. Eine Erläuterung der Algorithmen ist in Abschnitt 5.3 zu finden. Fettgedruckt ist immer der jeweils höchste Wert einer Spalte.

6.1.1. Persönliche Modelle

Die Evaluierung der persönlichen Modelle erfolgt wie folgt: Für jeden Benutzer b wird eine 10-fache Kreuzvalidierung durchgeführt, wobei die Trainingsdaten nur Daten von b beinhalten. Anschließend wird der Durchschnitt der einzelnen Genauigkeiten über alle Benutzer gebildet, um einen globalen Genauigkeitswert zu erhalten.

Nutzt man nur je ein Gyroskop, ist die Qualität der persönlichen Modelle am schlechtesten. Im Gegensatz dazu haben bereits Modelle, die lediglich Daten von einem der Beschleunigungssensoren verwenden, eine gute Aussagekraft. Weiss et al.[16] stellten in ihrem Experiment fest, dass die Genauigkeit des *Phone accel*-Modells 17.8% über dem *Watch accel / Band accel*-Modell lag. In unserem Experiment ist der Unterschied trotz gleichem Experimentaufbau mit 1.6% weit geringer. Eine denkbare Erklärung wäre, dass die Probanden von Weiss et al. über verschiedene Aktivitäten hinweg mehr Variation bei ihrer Bewegung des Unterkörpers eingebracht haben könnten.

Vergleicht man die beste Datenquelle (Band accel) mit den Kombinationsmöglichkeiten, so lässt sich feststellen, dass die Hinzunahme der anderen Sensoren des Band eine Genauigkeit von 97.1% erzielt, wohingegen die Kombination der beiden Smartphone-Sensoren lediglich eine Verbesserung von 0.7% gegenüber *Phone accel* und gegenüber *Band accel* sogar eine Verschlechterung erwirkt. Hebt man jedoch die künstlichen Einschränkungen auf, so wird eine Genauigkeit von 99.4% erreicht, sodass Benutzer nach dreiminütiger Aufnahme einer Aktivität davon ausgehen könnten, dass diese in Zukunft in aller Regel automatisch erkannt wird. Insgesamt konnte gegenüber *Phone Accel* damit eine Verbesserung um 7.8% erreicht werden, obwohl die Verbesserung einer bereits hohen Genauigkeit herausfordernd ist. **TODO: Kann man das belegen?** Gegenüber dem besten Ergebnis eines einzelnen Sensors (*Band accel*) mit 75.2% liefert die Kombination aller Daten eine Verbesserung um 7.8%.

Algo.	Phone accel	Phone gyro	Band accel	Band gyro	Comb.	Band comb.	Phone comb.
RF	88.7	68.2	91.6	80.5	99.4	97.1	90.7
J48	90.0	64.3	86.5	72.7	92.8	90.1	89.7
IB3	62.2	44.8	77.0	59.4	82.3	76.9	60.1
NB	87.6	60.4	90.9	78.8	96.2	92.2	85.5
MLP	78.9	51.5	88.9	67.8	94.8	90.7	75.1
∅	81.5	57.8	87.0	71.8	93.1	89.4	80.2

Abbildung 6.1.: Genauigkeit der persönlichen Modelle in Prozent

Algo.	Phone accel	Phone gyro	Band accel	Band gyro	Comb.	Band comb.	Phone comb.
RF	39.2	35.1	75.2	61.7	78.5	76.5	41.3
J48	32.9	29.1	61.6	49.2	59.7	61.3	32.6
IB3	24.3	22.6	60.9	45.0	52.9	58.0	26.0
NB	31.8	30.0	67.3	56.9	60.9	62.3	35.0
MLP	28.8	29.3	70.3	53.5	54.9	74.3	31.5
∅	81.5	29.2	67.1	53.3	61.4	66.5	33.3

Abbildung 6.2.: Genauigkeit der unpersönlichen Modelle in Prozent

6.1.2. Unpersönliche Modelle

Für die Evaluierung der unpersönlichen Modelle wird eine Kreuzvalidierung über die Nutzer durchgeführt, das heißt der Datensatz D wird für alle Benutzer b aufgeteilt in $D = \text{Train} \uplus \text{Test}$ mit $\text{Train} = \{\text{Intervall ist nicht von } b\}$, $\text{Test} = D \setminus \text{Train}$.

Das schlechteste Ergebnis wird mit 35.1% bei den unpersönlichen Modellen erzielt, wenn nur das Gyroskop des Smartphones verwendet wird. Nur wenig besser sind die Modelle, die nur den Beschleunigungssensor des Smartphones (39.2%) oder beide Sensoren des Smartphones (41.3%) nutzen. Für den praktischen Gebrauch sind diese Modelle aufgrund ihrer Ungenauigkeit untauglich. Schon rund 20% besser sind die Modelle, die nur das Gyroskop des Bands nutzen, wobei der Wechsel zum Beschleunigungssensor eine weitere Verbesserung um fast 14% ermöglicht. Offenbar erklären die Daten des Beschleunigungssensors einen Großteil der Daten, die auch durch das Gyroskop und die anderen Sensoren des Band erklärt werden, sodass eine Kombination aller Band-Sensoren lediglich eine Verbesserung um rund 1% ermöglicht. Selbiges gilt auch für die Hinzunahme der Daten des Smartphones: Es wird eine Verbesserung um 2% erzielt, was darauf schließen lässt, dass die Armbewegungen unter den Probanden ähnlich sind, die Bewegungen auf Hüfthöhe hingegen weniger. Gegenüber dem besten Ergebnis eines einzelnen Sensors (*Band accel*) mit 75.2% liefert die Kombination aller Daten eine Verbesserung um 3.3%.

6.2. Genauigkeit bei ungenauen Zeitstempeln

Die Problemstellung, Daten mehrerer Quellen miteinander zu kombinieren, wirft die Frage auf, inwiefern die Synchronisierung der Daten einen Einfluss auf die Genauigkeit der resultierenden Modelle besitzt. Smartphone und Band besitzen je eine eigene Uhr, sodass

Algo.	Comb. (Pers.)	Verrauscht (Pers.)	Comb. (Unpers.)	Verrauscht (Unpers.)
RF	99.4	99.2	78.5	78.3
J48	92.8	93.5	59.7	57.3
IB3	82.3	81.5	52.9	52.9
NB	96.2	95.7	60.9	60.5
MLP	94.8	94.2	54.9	55.4
\emptyset	93.1	92.8	61.4	60.9

Abbildung 6.3.: Genauigkeit der Modelle mit verrauschten Band-Zeitstempeln in Prozent

die Datenquellen zum selben Zeitpunkt t *Readings* mit Zeitstempeln $t + \epsilon_{\text{Smartphone}}$, bzw. $t + \epsilon_{\text{Band}}$ mit $|\epsilon_{\text{Smartphone}} - \epsilon_{\text{Band}}| > 0$ liefern könnten. Wird die Differenz zu groß, schadet die Kombination der Quellen möglicherweise mehr als sie nützt.

Ein erster Ansatz, um dieses Problem zu bekämpfen, war daher zunächst, die beiden Uhren via Bluetooth und ein Verfahren wie das *Network Time Protocol (NTP)* [12] zu synchronisieren. Leider ermöglicht das Microsoft Band SDK nicht die Ausführung beliebigen Codes auf dem Band, sondern nur das Abonnieren von Events, die Sensordaten ausliefern. Das SDK versteht diese Events lokal auf dem Smartphone mit Zeitstempeln, sodass weder direkt noch indirekt Zugriff auf die Uhr des Bands besteht und eine Synchronisierung somit unmöglich ist.

Um zu prüfen, inwiefern unsere Modelle durch eine mögliche, nicht detektierbare Differenz der Uhren beeinträchtigt werden, generiert die Transformationssoftware zusätzlich zum normalen vollständigen Datensatz auch einen, in dem die Zeitstempel des Bands additiv mit gausschem Rauschen versehen wurden. Die Parameter des gausschen Rauschens wurden auf einen Mittelwert von 0 und eine Standardabweichung von 500ms festgelegt. Die Ergebnisse der Auswertung werden in Abbildung 6.3 aufgeführt. Betrachtet man die jeweils besten Modelle, so lässt sich sowohl bei den persönlichen, als auch bei den unpersönlichen Modellen eine Verschlechterung um lediglich 0.2% feststellen. Auch die anderen Modelle sind im Durchschnitt mit Verschlechterungen um 0.3%, bzw. 0.5% hinreichend resistent gegen das Rauschen.

Um zu prüfen, ob eine Standardabweichung von 500ms sinnvoll ist, wurde über drei Tage hinweg um jeweils 10, 15 und 18 Uhr geprüft, mit welchem zeitlichen Abstand das Umspringen von der vollen Stunde auf die nächste Minute erfolgte. Eine Verzögerung war in keinem der Fälle mit dem Auge beobachtbar, weshalb davon auszugehen ist, dass die Uhren aufgrund der bereits bestehenden Synchronisierung durch die vom Hersteller gelieferte Band-App für Android in der Regel weniger als 500ms voneinander abweichen. Dies führt zu der Annahme, dass die Genauigkeit der Zeitstempel im Rahmen dieser Arbeit keinen negativen Einfluss auf die Ergebnisse hat und somit nicht weiter betrachtet werden muss.

6.3. Genauigkeit bei einer niedrigen Sampling-Rate

Da beide verwendete Geräte primär akkubetrieben sind, muss bei einer marktreifen Umsetzung einer Aufnahmesoftware auch bedacht werden, dass das Sammeln der Daten hinsichtlich der Energieressourcen nicht kostenlos ist. Eine Rolle beim Energieverbrauch spielt auch die Sampling-Rate (Abtastrate): Ist sie hoch eingestellt, können die entsprechenden Sensoren nicht

Gerät	Sensor	Durchsch. Sampling-Rate
Smartphone	Accelerometer	200 Hz
Smartphone	Gyroskop	200 Hz
Fitness-Tracker	Accelerometer	8 Hz
Fitness-Tracker	Gyroskop	8 Hz
Fitness-Tracker	Laufgeschwindigkeit	1 Hz
Fitness-Tracker	Hautwiderstand	5 Hz

Abbildung 6.4.: Sampling-Raten der Sensoren

Algo.	5 Hz (Pers.)	1 Hz (Pers.)	5 Hz (Unpers.)	1 Hz (Unpers.)
RF	98.6	95.3	70.1	62.1
J48	93.8	87.8	53.4	48.7
IB3	71.0	14.6	43.5	7.8
NB	91.6	75.5	53.3	44.2
MLP	90.0	78.9	63.6	51.2
Ø	89.0	70.4	56.8	42.8

Abbildung 6.5.: Genauigkeit der Modelle mit reduzierter Sampling-Rate in Prozent

in einen Energiesparmodus wechseln und außerdem müssen von der anfordernden Anwendung mehr Daten verarbeitet werden, wodurch wiederum Anforderungen an den Prozessor gestellt werden [10]. Beim Band kommt hinzu, dass eine höhere Bluetooth-Datenrate in Kauf genommen werden muss, die insbesondere auf den bauartbedingt kleinen Akku des Fitness-Trackers einen negativen Einfluss hat. Aus diesem Grund untersuchen wir, inwiefern die Sampling-Rate einen Einfluss auf die Genauigkeit unserer Modelle besitzt, um möglicherweise Energie sparen zu können.

Die Sampling-Raten, mit denen die Aufnahmesoftware Daten aufgezeichnet hat, befinden sich in Abbildung 6.4. Um Sampling-Raten miteinander zu vergleichen, wurden dieselben Rohdaten verwendet wie im normalen, vollständigen Modell. Um eine Sampling-Rate von k Hz zu simulieren, wurden nach einem behaltene Reading (t_0 secs, sensor, source, ...) alle weiteren Readings (t secs, sensor, source, ...) mit $t \in [t_0, t_0 + 1/k]$ gelöscht.

Wir vergleichen nun die besten Modelle miteinander und rufen uns in Erinnerung, dass *Comb. (Pers.)* eine Genauigkeit von 99.4% und *Comb. (Unpers.)* eine Genauigkeit von 78.5% erzielte. Reduziert man die Sampling-Rate des persönlichen Modells auf 5 Hz, so reduziert sich die Genauigkeit um 0.8%. Eine dramatischere Verschlechterung wird durch die Reduzierung auf 1 Hz verursacht, durch die sich die Genauigkeit um 4.1% verschlechtert. Noch stärker beeinträchtigt werden durch die Reduzierung der Sampling-Rate die unpersönlichen Modelle. Beschränkt man diese auf 5 Hz, so ergibt sich gegenüber *Comb. (Unpers.)* eine Verschlechterung um 8.4%. Bei einer Beschränkung auf 1 Hz beträgt die Verschlechterung sogar 16.4%, sodass zu einer Reduzierung der Sampling-Rate nur bei den ohnehin bereits genauen persönlichen Modellen zu raten ist.

Algo.	Comb. (Pers.)	Überlappung (Pers.)	Comb. (Unpers.)	Überlappung (Unpers.)
RF	99.4	99.2	78.5	73.4
J48	92.8	96.3	59.7	57.5
IB3	82.3	81.8	52.9	50.7
NB	96.2	95.8	60.9	59.7
MLP	94.8	95.0	54.9	59.4
∅	93.1	93.6	61.4	60.9

Abbildung 6.6.: Genauigkeit der Modelle mit Intervallüberlappung in Prozent

Algo.	Comb. (Unpers.), verschmolzen	Comb. (Unpers.), nicht verschmolzen
RF	87.1	78.5
J48	73.7	59.7
IB3	63.9	52.9
NB	72.4	60.9
MLP	77.6	54.9
∅	74.9	61.4

Abbildung 6.7.: Genauigkeit der Modelle bei Verschmelzung der Ess- und Trinkaktivitäten in Prozent

6.4. Genauigkeit bei Überlappung der Intervalle

Bao et al. nutzten 2004 auf der Basis bereits bestehender Werke für die Transformation eine Intervallüberlappung von 50% [2]. Weiss et al. probierten dies laut ihrem Paper nicht aus [16]. Um zu prüfen, ob dies eine weitere Verbesserung der Genauigkeiten erzielen könnte, wurden für die Intervalle $(t_0, t_0 + 10)$, $((t_0 + 10) - 10 * 0.5, (t_0 + 10) - 10 * 0.5 + 10)$, ... Features generiert und wie zuvor evaluiert. **TODO: Grafik zur Illustration der Intervalle**

Anders als erhofft verschlechterten sich die persönlichen RF-Modelle wie in Abbildung 6.6 ersichtlich durch die Überlappung um 0.2%, während sich die unpersönlichen Modelle sogar um 5.1% verschlechterten.

6.5. Genauigkeit bei Verschmelzung der Ess- und Trinkaktivitäten

TODO: Confusion-Matrix all impersonal rf einbinden Betrachtet man die Konfusionsmatrix der vollständigen unpersönlichen RF-basierten Modelle, so lässt sich feststellen, dass die Klassen, die sich mit Essen oder Trinken beschäftigen, nur ungenau erkannt werden und somit die Durchschnittsgenauigkeit entsprechend senken. Zudem liegt die fehlerhaft vorhergesagte Klasse häufig **TODO: Zahl nennen** im Bereich des Essens und Trinkens. Zurückzuführen ist dies auf die Tatsache, dass sich diese Aktivitäten untereinander ähneln, da die zentralen Bewegungsmerkmale das Sitzen sowie das Führen der Hand zum Mund sind.

Abbildung 6.7 zeigt, wie sich die Genauigkeit der unpersönlichen Modelle verändert, wenn man diese Aktivitäten zu einer einzigen Aktivität verschmilzt. **TODO: Zeige auch neue Konfusionsmatrix.** Bei den RF-Modellen ist eine Verbesserung um 8.6% zu erkennen, wovon

ein Teil auch darauf zurückzuführen sein könnte, dass es insgesamt weniger Klassen gibt und die Wahrscheinlichkeit einer Fehlklassifikation somit sinkt.

Aus diesen Daten lässt sich schließen, dass bei der Verwendung von unpersönlichen Modellen darauf geachtet werden sollte, dass die zu erkennenden Aktivitäten nicht zu feingranular voneinander getrennt werden, da dies die Klassifikation stark erschweren kann.

7. Conclusions

TODO: Check template for suggestions on how to write this

7. *Conclusions*

Anhang

A. What goes in the appendices

A. What goes in the appendices

B. Used Acronyms

B. Used Acronyms

C. Literaturverzeichnis

- [1] D. Aha and D. Kibler. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.
- [2] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer, 2004.
- [3] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [5] Fitbit. How accurate are fitbit trackers? https://help.fitbit.com/articles/en_US/Help_article/1136. Abgerufen am 20.01.2017.
- [6] Google Inc. *Android Sensors Overview*, 12 2016.
- [7] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2009.
- [8] Simon Haykin. *Neural networks, a comprehensive foundation*. -, 1994.
- [9] George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo, 1995. Morgan Kaufmann.
- [10] A. Krause, M. Ihmig, E. Rankin, D. Leong, Smriti Gupta, D. Siewiorek, A. Smailagic, M. Deisher, and U. Sengupta. Trading off prediction accuracy and power consumption for context-aware wearable computing. In *Proc. Ninth IEEE Int. Symp. Wearable Computers (ISWC'05)*, pages 20–26, October 2005.
- [11] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, March 2011.
- [12] D Mills, J Martin, J Burbank, and W Kasch. Rfc5905. *Network Time Protocol Version*, 4:2010–06, 2010.
- [13] AY Ng. Cs229 lecture notes on machine learning. Technical report, Technical report, Stanford University, Department of Computer Science, Stanford, USA, 2011. 39, 116, 140, 2011.
- [14] Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

- [15] B. N. Taylor and A. Thompson. The international system of units (si). National Institute of Standards and Technology, Special Publication 330, Gaithersburg, MD, 2008.
- [16] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber. Smartwatch-based activity recognition: A machine learning approach. In *Proc. IEEE-EMBS Int. Conf. Biomedical and Health Informatics (BHI)*, pages 426–429, February 2016.

Erklärung der Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Ort, Datum

Unterschrift