

Forschungsprojekt Teil B

**Modellierung und Analyse multimodaler
Daten aus sensorbasierten
Lernumgebungen**

Jan Wankmüller

jan.wankmueller@student.htw-berlin.de

Inhaltsverzeichnis

1	Einleitung	1
1.1	Zusammenfassung von Forschungsprojekt A	2
1.2	Kurzbeschreibung von Forschungsprojekt B	3
2	Modellierung von Sensordaten zur Persistierung in dokumentenorientierten Datenbanken	4
2.1	Lern- und Sensordaten in LEMO2	5
2.1.1	Lerndaten im relationales LEMO2-Datenmodell	5
2.1.2	Schnittstellen des LEMO2-Daten-Backends	6
2.1.3	Sensordaten des Emotional Picture Experiments	6
2.2	Migration der Daten	7
2.2.1	dokumentenorientiertes Datenmodell	7
2.2.2	Migration der Lern- und Sensordaten	7
2.3	Entwicklung einer Anwendung zur Datenpersistierung	9
2.3.1	Datenpersistierung als eigenständiger Microservice	9
2.3.2	Kapselung der Anwendung durch Virtualisierung in einem Container	9
3	Analyse von Anomalien in Aufnahmen der elektrodermalen Aktivität	11
3.1	elektrodermale Aktivität in Sensordaten des EPE	12
3.1.1	elektrodermale Aktivität	12
3.1.2	Sensordaten des Emotional Picture Experiment	13
3.1.3	explorative Datenanalyse der EDA-Sensordaten des EPE	14
3.2	manuelle Analyse anomaler Merkmale der elektrodermalen Aktivität	16
3.2.1	Anomalieerkennung	16
3.2.2	Erkennung anomaler Merkmale bei Aufnahmen der EDA	18
3.2.3	Vorgehensweise bei der manuellen Erkennung anomaler Merkmale	18
3.2.4	Visualisierung der EDA-Sensordaten	19

INHALTSVERZEICHNIS

3.2.5	Ergebnisse	20
3.3	statistische Analyse zur automatisierten Erkennung anomaler Merkmale	23
3.3.1	kurzzeitiges Abfallen der elektrodermalen Aktivität . .	23
3.3.2	gekappte Maximalwerte der elektrodermalen Aktivität	25
3.3.3	hohe und niedrige Varianz der elektrodermalen Aktivität	27
3.3.4	starkes Ansteigen oder Abfallen über die gesamte Auf- nahme	28
4	Bewertung und Ausblick	31
4.1	Datenmodellierung von Lern- und Sensordaten	32
4.2	Anomalieerkennung	33

Kapitel 1

Einleitung

1.1 Zusammenfassung von Forschungsprojekt A

Der vorherigen Teil A dieses Forschungsprojektes beschäftigte sich mit der Datenmodellierung multimodaler Daten aus sensorbasierten Lernumgebungen.

Dabei wurde

1. eine Literaturrecherche zum aktuellen Stand der Forschungsgebiete *Learning Analytics* und *Multimodal Learning Analytics* durchgeführt um Informationen über den Kontext, in dem Sensordaten modelliert und persistiert werden, zu gewinnen
2. der aktuelle Stand von Anwendungen und Frameworks zur Datenpersistierung in *Multimodal Learning Analytics* untersucht und verglichen
3. eine Zusammenstellung von Anforderungen an die Modellierung und Persistierung von Sensordaten erstellt
4. ein Entwurf für ein dokumentenorientiertes Datenmodell für Sensordaten im LEMO2-Framework erstellt

Auf Grundlage des erstellten Datenmodells sollen im ersten Teil des Forschungsprojektes B Lern- und Sensordaten persistiert werden.

1.2 Kurzbeschreibung von Forschungsprojekt B

Dem Titel des Forschungsprojektes, *Modellierung und Analyse multimodaler Daten aus sensorbasierten Lernumgebungen*, entsprechend ist dieser Teil B in zwei weitere Teile aufgeteilt, die in dieser Arbeit in jeweils eigenständigen Kapiteln behandelt werden.

Im ersten Teil (Kapitel 2) wird das in Teil A entworfene Datenmodell implementiert und zur Persistierung von Lern- und Sensordaten genutzt. Dazu werden Lerndaten aus LEMO2 in eine dokumentenorientierte Datebank migriert und Sensordaten integriert. Der Zugriff die Daten in diesem Datenmodell wird in einer eigenständigen, unabhängigen Anwendung gekapselt. In dieser Arbeit werden die wesentlichen Aspekte des Vorgehens beschrieben, die Entwicklung selbst ist Teil eines eigenständigen Moduls (ICW2).

Der zweite Teil dieses Forschungsprojektes (Kapitel 3) beschäftigt sich mit der Analyse und Erkennung anomaler Merkmale in Sensoraufnahmen der elektrodermalen Aktivität, die im Rahmen des *Emotional Picture Experiments* an der HTW Berlin entstanden sind.

Dazu wird zunächst ein Überblick über elektrodermale Aktivität und die zu untersuchenden Aufnahmen gegeben. Anschließend werden bei einer manuellen Anomalieerkennung auffällige Merkmale in den EDA-Aufnahmen identifiziert. Diese Merkmale werden statistisch modelliert um sie automatisiert in Aufnahmen der elektrodermalen Aktivität identifizieren zu können.

Ziel soll dabei sein, mögliche bei der Aufnahme entstandenen Fehler zu identifizieren und automatisiert zu erkennen, um zukünftige weitere Aufnahmen einfacher prüfen zu können und später bei der Erstellung von Trainings- und Testdaten bei maschinellen Lernverfahren berücksichtigen zu können

Im abschließenden Kapitel 4 wird jeweils eine kurze Zusammenfassung, Bewertung der Ergebnisse und Ausblick für beide Teile des Forschungsprojektes gegeben.

Kapitel 2

Modellierung von Sensordaten zur Persistierung in dokumentenorientierten Datenbanken

2.1 Lern- und Sensordaten in LEMO2

Im den folgenden Unterabschnitten werden die bereits vorhandenen Lern- und Sensordaten sowie die existierende Systemarchitektur des Datenbackends und dessen Schnittstelle beschrieben.

2.1.1 Lerndaten im relationales LEMO2-Datenmodell

Vorhandene Lerndaten in LEMO2 werden in einem relationalen Datenmodell persistiert. In Abbildung 2.1 sind alle Entitäten der relationalen Datenbank in einem ERM-Modell dargestellt. Rot markiert sind die Entitäten, die die wesentlichen Informationen einer Lernaktivität (in einem actor-verb-object Triple) repräsentieren, weitere Entitäten enthalten Relationen, Metainformationen (wie die Entität *role*, die die Rolle einer Person in einem Context angibt) und Attributtabelle, die ein Hinzufügen weiterer Attribute erlauben.

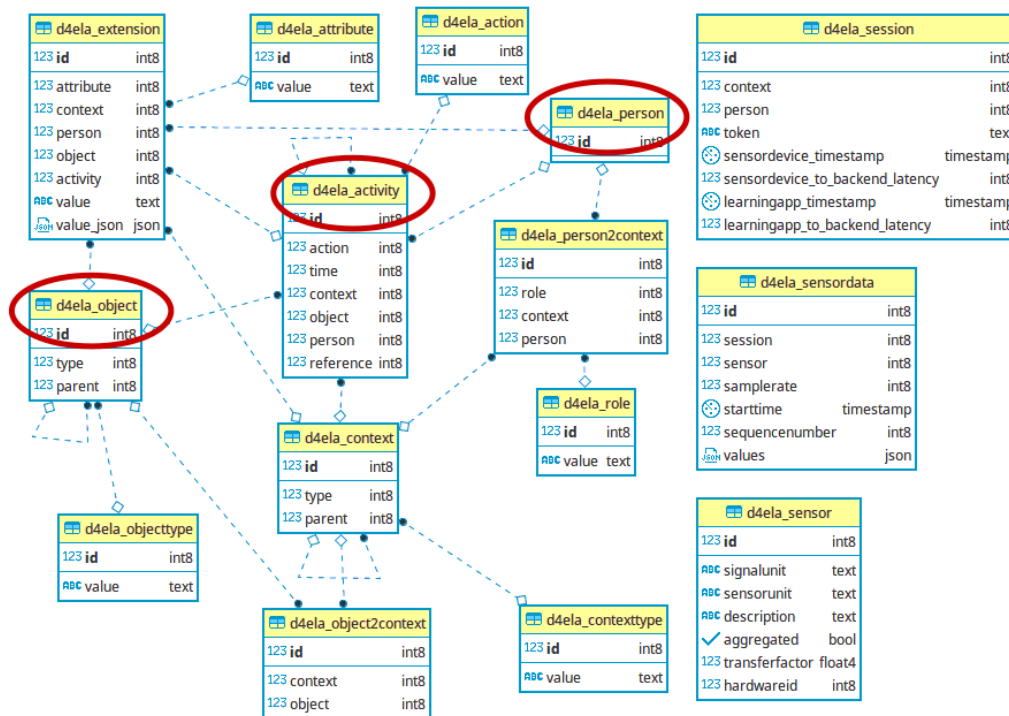


Abbildung 2.1: ERM-Modell der Lerndaten im LEMO2 Postgres-DBMS

Sensordaten sind zwar im Schema durch die Entitäten *sensor* und *sensordata* enthalten, wurden jedoch bisher nicht genutzt. Die Werte der Sensoren

KAPITEL 2. MODELLIERUNG VON SENSORDATEN ZUR PERSISTIERUNG IN DOKUMENTENORIENTERTEN DATENBANKEN

sind hier unstrukturiert als JSON-Feld in einer relationalen Datenbank modelliert. Lerndaten die in diesem Datenmodell gespeichert sind, wurden als Datenbankabbild zur späteren Migration in ein dokumentenbasiertes Datenmodell verwendet.

2.1.2 Schnittstellen des LEMO2-Daten-Backends

Die Datenpersistierung in LEMO2 ist in einem Datenbackend gekapselt, das mit dem Framework *Spring*¹ entwickelt wurde und über den ORM *Hibernate*² auf eine relationale Datenbank mit dem DBMS *PostgreSQL*³ zugreift.

Die REST-Schnittstellen des Backends zum Zugriff auf die Lerndaten sind mit dem Framework *Swagger*⁴ dokumentiert; für jede Resource sind die zugehörigen Routen und Zugriffsmethoden beschrieben.

Die Konzept der Schnittstelle, der Zugriff auf die wesentlichen Entitäten soll auch nach einer Migration zu einem neuen Datenmodell erhalten bleiben um einen Wechsel zwischen der Nutzung der Datenmodelle zu erleichtern. Da die Schnittstelle jedoch stark mit der Struktur der relationalen Datenbank verbunden ist können nicht alle Details übernommen werden.

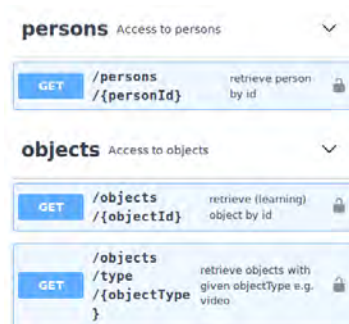


Abbildung 2.2: Ausschnitt aus der grafischen Darstellung der Schnittstelle von LEMO2 mit Swagger

2.1.3 Sensordaten des Emotional Picture Experiments

Um das entwickelte Datenmodell neben den LEMO2-Lerndaten auch mit Sensordaten zu nutzen wird auf Sensordaten des Emotional Picture Experiments (dessen Hintergrund in Abschnitt 3.1.2 in anderem Kontext erläutert wird) zugegriffen.

Bei diesen Daten handelt es sich zum einen um Rohdaten von Sensoren der elektrodermalen Aktivität und EKG, zum anderen um davon abgeleitete und aggregierte Daten. Dadurch können sowohl Sensordaten mit einer festen Intervallen (wie einer Abtastrate) als auch sporadisch auftretende (wie der Abstand zwischen Herzschlägen) verwendet werden.

¹<https://spring.io/>, Aufgerufen am 12.10.2019

²<https://hibernate.org/>, Aufgerufen am 12.10.2019

³<https://www.postgresql.org/>, Aufgerufen am 12.10.2019

⁴<https://swagger.io/>, Aufgerufen am 12.10.2019

2.2 Migration der Daten

In den folgenden Unterabschnitten wird das Vorgehen bei der Migration der Lern- und Sensordaten in das entwickelte dokumentenorientierte Datenmodell beschrieben.

2.2.1 dokumentenorientiertes Datenmodell

Das in Teil A des Forschungsprojektes übernommene Datenmodell wurde in wesentlichen Teilen übernommen. Zusätzlich wurden einige Ergänzungen und Erweiterungen vorgenommen:

- um große Datenmenge in einer Aufnahme eines Sensor abspeichern zu können, wurden Sequenznummern zur Aufteilung einer Aufnahme in mehrere Dokumente eingeführt. Damit ist die maximale Größe einer Aufnahme nicht von der maximalen Größe der verwendeten dokumentenbasierten Datenbank abhängig, sondern kann mit der Datenbank skalieren.
- die wesentlichen Entitäten der Lern- und Sensordaten des relationalen LEMO2-Datenmodells wurden übernommen, allerdings wurden Relationen und deren Metainformationen direkt in das die Entität repräsentierende Dokument eingefügt. Dadurch können auch Metainformationen und Relationen leicht ergänzt und erweitert werden

```
  _id: ObjectId("5d022b17aaf0e79ca2c16379")
  values: Array
    0: Object
      timestamp: "3106778"
      ibi: "682"
    1: Object
    2: Object
    3: Object
    4: Object
    5: Object
    6: Object
    7: Object
  end: 1561023191
  sequence_number: 4
  sensortype: Object
    description: "IBI modelled as sporadic events"
    session: ObjectId("5d022b17aaf0e79ca2c16374")
    start: 1560423191
    type: "sporadic"
```

Abbildung 2.3: Ausschnitt eines Sensordaten repräsentierenden Dokuments im dokumentenorientierten Datenmodell

In Abbildung 2.3 ist ein Ausschnitt eines Sensordaten-Dokuments mit sporadisch auftretenden Daten dargestellt. Hier ist sowohl die verwendete Struktur zur Modellierung von Zeitstempeln für jeden Wert sporadisch auftretender Sensordaten als auch die Sequenznummer des Dokuments, die zur Verteilung einer Aufnahme auf mehrere Dokumente notwendig ist, zu sehen.

2.2.2 Migration der Lern- und Sensordaten

Zur Migration der Lern- und Sensordaten wurde das LEMO2-Datenbankabbild in eine relationale Datenbank geladen. Anschließend wurde mit einem Python-Script

KAPITEL 2. MODELLIERUNG VON SENSORDATEN ZUR PERSISTIERUNG IN DOKUMENTENORIENTERTEN DATENBANKEN

die Entitäten Activity, Object, Context, Person sowie die zugehörigen Metadaten geladen und in Dokumente einer MongoDB-Datenbank migriert. Dabei wurden temporär die in der relationalen Datenbank verwendeten IDs zur Abbildung von Relationen verwendet um die Verweise von Relationen anschließend in die Object-ID-Identifikatoren der verwendeten MongoDB-Datenbank umzuwandeln.

Sensordaten wurden aus csv-Dateien importiert und Werte verschiedener sporadischer sowie kontinuierlicher Daten in Dokumenten gespeichert und mit zufällig gewählten Session- und Person-Dokumenten verknüpft.

2.3 Entwicklung einer Anwendung zur Datenpersistierung

In den folgenden Unterabschnitten wird die Entwicklung einer eigenständigen Anwendung zur Datenpersistierung beschrieben.

2.3.1 Datenpersistierung als eigenständiger Microservice

Der Zugriff auf Lern- und Sensordaten wurde in einer eigenständigen und unabhängigen Anwendung (Datenbackend) gekapselt um

- durch eine Modularisierung der Datenpersistierung einen schrittweisen Austausch des Datenbankends in LEMO2 zu ermöglichen
- einen unabhängigen Einsatz vom LEMO2-Framework zu ermöglichen

Zur Umsetzung des Backends wurde das Micro-Webframework *Flask*⁵ gewählt. Es ist in Python geschrieben und besitzt im Gegensatz zu umfangreicheren MVC-Frameworks wie Spring oder Django einen minimaleren Funktionsumfang, der sich speziell für Anwendungen ohne grafische Oberfläche eignet. Zusätzlich wurde die Erweiterung *FlaskRestPlus*⁶ eingesetzt, das die Entwicklung von REST-Schnittstellen erleichtert und die Generierung von auch in LEMO2 eingesetzten Schnittstellendokumentation mit *Swagger* ermöglicht.

2.3.2 Kapselung der Anwendung durch Virtualisierung in einem Container

Zur Ausführung der Anwendung werden neben einer Python3-Installation auch verschiedene Python3-Module benötigt.

Um

- eine Ausführung unabhängig vom verwendeten Betriebssystem
- ohne Installation der notwendigen Laufzeitumgebung und Bibliotheken (bzw. Python-Module)
- eine gekapselte Ausführung ohne Zugriff auf die Umgebung des Betriebssystems

⁵<https://flask.palletsprojects.com>, Aufgerufen am 11.10.2019

⁶<https://flask-restplus.readthedocs.io>, Aufgerufen am 11.10.2019

KAPITEL 2. MODELLIERUNG VON SENSORDATEN ZUR PERSISTIERUNG IN DOKUMENTENORIENTERTEN DATENBANKEN

zu ermöglichen wurde die Anwendung in einem Docker-Container ⁷ virtualisiert. Dazu wurde ein eigener Docker-Container auf Grundlage eines Docker-Abbilds mit dem Betriebssystem Ubuntu erzeugt, auf dem Python und die notwendigen Bibliotheken installiert wurden. Auf diese Weise ist die Ausführung der Anwendung ohne zusätzlichen Installationsaufwand, abgesehen von der notwendigen Docker-Umgebung, möglich.

⁷<https://www.docker.com/>, Aufgerufen am 11.10.2019

Kapitel 3

Analyse von Anomalien in Aufnahmen der elektrodermalen Aktivität

3.1 elektrodermale Aktivität in Sensordaten des EPE

In den folgenden Unterabschnitten werden als Hintergrund des weiteren Vorgehens Eigenschaften der elektrodermalen Aktivität sowie Herkunft und Kontext der verwendeten Sensordaten beschrieben. Die später analysierten Sensordaten werden in einer explorativen Datenanalyse beschrieben.

3.1.1 elektrodermale Aktivität

Unter dem Überbegriff der *elektrodermalen Aktivität (EDA)* werden Änderungen der elektrischen Eigenschaften der Haut wie der Hautleitfähigkeit zusammengefasst [2, S.3]. Als Einheiten für gemessene Werte werden Microsiemens (μS) und Micromho (μmho) verwendet, beide Einheiten sind äquivalent ($1\mu\text{S} = 1\mu\text{mho}$) [2, S.4].

Mit der EDA werden sowohl tonische, als auch phasische Phänomene der Haut erfasst.

Der tonische Anteil ändert sich nur langsam, das Niveau wird mit dem Messwert *Skin Conductance Level (SCL)* erfasst. Dieses Leitwertniveau ergibt sich aus der Füllung der Schweißdrüsengänge und anderer Eigenschaften der Haut. Das SCL ändert sich zwar langsam aber ständig bei einem Individuum, zwischen verschiedenen Individuen kann es deutliche Unterschiede geben [2, S.4f.] [9, S. 16ff.]. So bewegt sich der tonische Anteil bei den meisten Menschen auf einem Niveau von 0-10 μS , allerdings können auch Werte von bis zu 100 μS vorkommen [9].

Der phasische Anteil der EDA tritt typischerweise als Reaktion auf externe Reize auf. Dabei verschiebt sich das Niveau der EDA im Sekundenbereich um 0,01 bis 5 μS . Bei der Bewertung von Fluktuationen als Reaktion auf einen Reiz können sich Überlagerungseffekte problematisch auswirken und eine eindeutige Interpretation erschweren [9, S.17f].

Zur Messung der EDA existieren zum einen endosomatische Methoden, bei denen keine elektrische Spannung an die Haut angelegt wird, die jedoch nur selten verwendet werden, da Messmethodik und Auswertung nach [9, S.19] [1] als problematisch gelten. Häufig zur Anwendung kommen exosomatische Methoden, bei denen eine geringe Gleich- oder Wechselspannung an



Abbildung 3.1: Beispiel für auf der Haut aufgebrachte Sensoren zur Messung der EDA, Quelle: www.zyto.com

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

die Haut angelegt wird [1, S.48ff.]. Abbildung 3.1 zeigt zwei solche auf der Hand aufgebrachten exosomatische Sensoren. Der gemessene Wert kann von klimatischen Bedingungen wie der Umgebungstemperatur und der Luftfeuchtigkeit beeinflusst werden. Auch physiologische Einflüsse wie die Temperatur, Durchblutung und Feuchtigkeit der Haut und ein Einfluss demografischer Eigenschaften konnten nachgewiesen werden, wie in einer umfangreichen Übersicht in [1] zusammengestellt wurde. Deshalb wird die zusätzliche Aufzeichnung meteorologischer Variablen zur Nachkontrolle empfohlen [9, S.15] [1].

3.1.2 Sensordaten des Emotional Picture Experiment

Im Rahmen des Projekts *Learning Analytics for sensor-based adaptive learning (LISA)* wurde ein *Emotional Picture Experiment (EPE)* an der HTW Berlin durchgeführt [10].

Dabei wurden Teilnehmern Bilder, die bei Betrachtung Emotionen hervorrufen sollen, und aus dem *International Affective Picture System (IAPS)* [6] stammen, gezeigt. Diesen Bildern (vier Beispielsbilder sind in Abbildung 3.2 dargestellt) sind jeweils akademischen Emotionen wie Aufregung, Wut, Ekel und Vergnügen zugeordnet um unabhängig von einer Selbstauskunft der Betrachter Informationen über die von ihnen empfundenen akademischen Emotionen zur Verfügung zu stellen.

Mit Sensoren wurden EDA- und ECG-Daten der Betrachter erhoben um später aus diesen Aufnahmen auf die von den Teilnehmern empfundenen akademischen Emotionen zu schließen zu können. Dazu wurden ihnen Elektroden auf Brust (zur Erkennung von ECG-Signale) und Finger (zur Erkennung der EDA-Signale) aufgeklebt, über einen Bausatz des Herstellers *Bitalino*¹ wurden die Sensordaten mit einer Auflösung von 10 Bit und einer Abtastfrequenz von 1000Hz aufgenommen und mit der Software *OpenSignals*² des Herstellers persistiert und in Echtzeit visuell überprüft [10].

Dabei wurden 60 Aufnahmen mit einer Länge



Abbildung 3.2: IAPS-Beispielsbilder [7]



Abbildung 3.3: Bitalino-Kit mit EDA- und ECG-Sensor [12]

¹<https://bitalino.com/en/>, Aufgerufen am 12.10.2019

²<https://bitalino.com/en/software>, Aufgerufen am 12.10.2019

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

von etwa 30 Minuten erstellt, die in ihnen enthaltenen Daten über die EDA werden in den folgenden Abschnitten dieser Arbeit auf auffällige Merkmale analysiert.

3.1.3 explorative Datenanalyse der EDA-Sensordaten des EPE

Mit einer explorativen Datenanalyse wurden eine erste Übersicht über Eigenschaften der OpenSignals-Aufnahmen der EDA gewonnen. Untersucht wurden 49 Aufnahmen, die jeweils mit einer Aufnahme Nummer zwischen 1 und 60 gekennzeichnet sind. Diese Nummern werden im Folgenden zur Identifikation der Aufnahmen verwendet.

Die Abtastrate der untersuchten Aufnahmen wurde bereits von 1000Hz auf 100Hz reduziert, die Länge der einzelnen Aufnahmen variiert. In Tabelle 3.1 werden statistische Eigenschaften der Aufnahmelängen und der Mittelwerte der EDA der Aufnahmen aufgeführt. Die Aufnahmelängen liegen zwischen etwa 45 und 65 Minuten, wobei sich die Länge der meisten Aufnahmen um einen Mittelwert von etwa 51 Minuten bewegt. Die mittlere EDA aller Aufnahmen liegt bei $13,5 \mu\text{S}$, unterscheidet sich aber zwischen den einzelnen Aufnahmen deutlich stärker als die Länge und reicht von $3,5 \mu\text{S}$ bis $23,2 \mu\text{S}$. Diese großen Unterschiede können aber durch die in Abschnitt 3.1.1 aufgeführten unterschiedlich hohen tonischen Anteile der EDA zwischen Individuen erklärt werden.

	Länge (min)	$\varnothing EDA (\mu\text{S})$
Mittelwert	50,6	13,5
Standardabweichung	3,9	4,9
Median	49,8	13,4
Minimum	45,8	3,5
Maximum	64,5	23,2
25%-Quantil	48,0	9,8
75%-Quantil	51,5	17,5

Tabelle 3.1: statistische Eigenschaften von Länge und Mittelwert der EDA bei Betrachtung aller 49 Aufnahmen

Ein visueller Eindruck bestätigt die unterschiedlichen Mittelwerte und zeigt zusätzlich, das Verlauf und kurzzeitige Änderungen zwischen den Aufnahmen teilweise stark variieren. In Abbildung 3.4 ist der zeitliche Verlauf der EDA von drei Aufnahmen dargestellt, neben den sich deutlich unterschei-

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

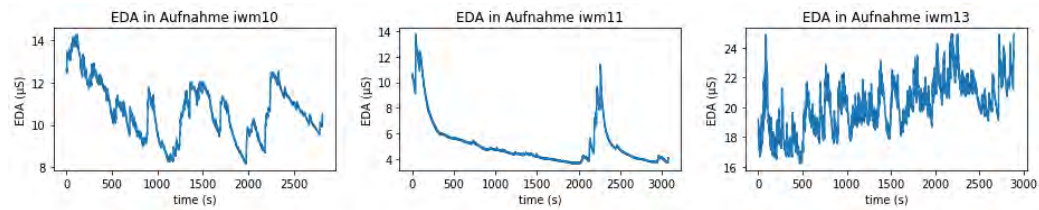


Abbildung 3.4: Verlauf der EDA in drei Aufnahmen mit deutlichen Unterschieden bei Niveau und Varianz in Abschnitten

denden Wertebereichen der EDA (4-14 μS bei Aufnahme 11, 16-25 μS bei Aufnahme 13) sind auch große Unterschiede beim Ausmaß der Änderungen der EDA in kurzen Zeitabschnitten von wenigen Minuten deutlich erkennbar. So variiert die EDA in Aufnahme 13 über den Verlauf der ganzen Aufnahme stark während Aufnahme 11 über einen langen Abschnitt kaum kurzzeitige Änderungen aufweist.

3.2 manuelle Analyse anomaler Merkmale der elektrodermalen Aktivität

In den folgenden Unterabschnitten wird das verwendete Verfahren zur (manuellen) Erkennung anomaler Merkmale in Aufnahmen der EDA beschrieben und dessen Umsetzung und Ergebnisse dokumentiert.

3.2.1 Anomalieerkennung

Anomalien

Anomalien sind nach [4, p. 3] Abweichung von einer Norm, die sich nicht durch eigenständige Eigenschaften, sondern durch die Gegenüberstellung zu erwartender Eigenschaften definieren. Die Erkennung von Anomalien basiert auf Modellen und Vorhersagen, die auf zuvor untersuchten Daten, aus denen eine Norm gebildet wurde, beruhen [8, p. 5].

Nach [4, p. 14] beschreiben die Schritte

1. Definition der Norm
2. Festlegung der Eigenschaften zur Bestimmung einer Abweichung von der Norm
3. Festlegung des Ausmaßes der Abweichung, die als anomal gilt

die notwendigen Ziele bei der Entwicklung eines Verfahrens zur Anomalieerkennung.

Nach [3] unterscheidet sich jedoch die Auffassung, was eine Anomalie darstellt, stark nach konkretem Anwendungsgebiet; so können Abweichungen sowohl als der Norm entsprechend als auch als anomal eingeordnet werden. Zusätzlich stellt die mangelnde Verfügbarkeit von gekennzeichneten (labeled) Daten ein großes Problem dar.

Verfahren zur Anomalieerkennung

Grundsätzlich können Verfahren zur Anomalieerkennung in

- beaufsichtigte (engl. *supervised*, mit voriger Kennzeichnungen aller Daten)
- teils beaufsichtigte (engl. *semi-supervised*, mit voriger Kennzeichnungen für einige Daten)

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

- unbeaufsichtigte (engl. *unsupervised*, ohne vorige Kennzeichnung von Daten)

unterteilt werden [4]. Bei unbeaufsichtigten Verfahren existiert im Vorfeld noch kein Modell der Daten, von denen eine Abweichung erkannt werden kann, während bei beaufsichtigten Verfahren historische Daten zur Modellierung von normalen und anomalen Eigenschaften verwendet werden.

Nach [3] wird, wie in Abbildung 3.5 dargestellt, eine angewandte Technik zur Erkennung von Anomalien zum einen durch Eigenschaften des zu lösenden Problems und zum anderen durch die Erkenntnisse verschiedener Wissenschaftsgebiete wie dem maschinellen Lernen und der Statistik beeinflusst. So werden beispielsweise neben fixen und dynamischen Schwellenwerten (wie beim Verfahren t-Digest zur Aktualisierung von Schwellenwerten bei großen Datenmengen [4, p. 20]) auch komplexe Verfahren des maschinellen Lernen wie Auto-Encoder zur Signalrekonstruktion verwendet [4, p. 30]). Verschiedene Verfahren zur Untersuchung von anomalen Ereignissen oder Abschnitten in einer Zeitreihe werden in [8, p. 155ff.] beschrieben.

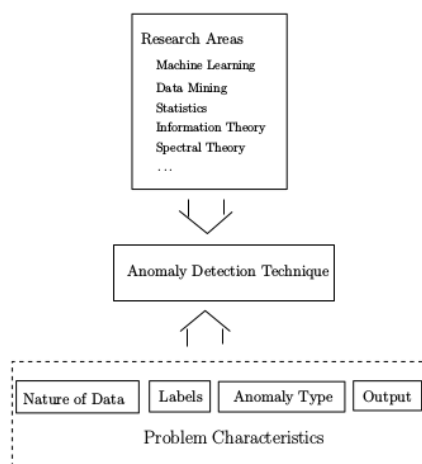


Abbildung 3.5: Übersicht zu Verfahren zur Anomalieerkennung, Ausschnitt aus [3]

Ziel einer Erkennung von Anomalien kann zum einen eine binäre Klassifikation, die einen Datensatz oder Bereich als anomal oder der Norm entsprechend einordnet, zum anderen ein Ranking sein, das es ermöglicht, Daten nach Grad der Anomalie zu sortieren.

Anomalieerkennung bei Sensordaten

Die zur Untersuchung von Anomalien in Sensordaten verwendete Lösung hängt von der konkreten Zielstellung und den verwendeten Daten ab.

In [11] wurden in einem drahtlosen Sensornetzwerk, das physiologischen Zustand und Vitalparameter von Patienten eines Krankenhauses überwacht eine Anomalieerkennung eingesetzt, um Patienten in kritischem Zustand zu klassifizieren. Dabei wurden als Verfahren Entscheidungsbäume und linearer Regression verwendet und ihre Eignung untersucht.

Zur Anomalieerkennung in großen und heterogenen Netzwerken von Umgebungssensoren in Echtzeit wurde in [5] eine auf univariater Autoregression

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

basierende Methode entwickelt und evaluiert um Ausfälle und Fehler von Sensoren zu erkennen.

Der Anwendungsfall dieser Studien ist die Echtzeitüberwachung von Sensornetzwerken, die große Datenmengen zur Verfügung stellen. Im Gegensatz dazu handelt es sich bei den in dieser Arbeit vorliegenden Aufnahmen um eine kleine Menge historischer Daten, eine Übertragung der in ihnen angewandten Verfahren wurde deshalb nicht weiter verfolgt.

3.2.2 Erkennung anomaler Merkmale bei Aufnahmen der EDA

Ziel ist die Erkennung von Aufnahmen der EDA des EPE, die Anomalien enthalten, die möglicherweise durch Fehler verursacht wurden. Dadurch soll eine spätere Überprüfung mit umfangreichem Domänenwissen bei der manuellen Interpretation der EDA möglicher Fehler bei der Messung der EDA erleichtert werden. Im Gegensatz zu den im vorigen Abschnitt 3.2.1 vorgestellten Arbeiten liegen dieser Aufgabe weder große Datenmengen noch ein in Echtzeit Daten erzeugendes Sensornetzwerk zugrunde.

Die Aufgabe besitzt folgende wesentliche Eigenschaften:

- wie in 3.1.1 beschrieben gibt es große Unterschiede der EDA zwischen Individuen
- die Aufnahmen entstanden während des EPE-Experiments, bei dem starke Emotionen auslösende Bilder gezeigt wurden; Veränderungen der EDA sind Ziel des Experiments und stellen nicht automatisch eine Anomalie im Aufgabenkontext dar
- es werden nur 49 Aufnahmen untersucht, da jede Aufnahme komplett betrachtet wird, ist diese Menge an Datensätzen gering

aufgrund der die weitere Vorgehensweise bestimmt wurde. Da nur wenige Datensätze Grundlage einer Anomalieerkennung sind wurde von probabilistischen Verfahren zur Erkennung anomaler Datensatz abgesehen. Stattdessen sollen anomale Merkmale und davon betroffene Datensätze manuell von mehreren Teilnehmern des Forschungsprojekte identifiziert werden um als Basis für eine spätere automatisierte Erkennung verwendet werden zu können.

3.2.3 Vorgehensweise bei der manuellen Erkennung anomaler Merkmale

Um eine manuelle Identifikation von anomalen Merkmalen zu erhalten wurden drei weitere Teilnehmer des Forschungsprojekts in einem gemeinsamen

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

Treffen nach ihren Einschätzungen der Aufnahmen befragt. Alle Teilnehmer waren durch ihre Arbeit in Forschungsprojekten mit den besprochenen Daten vertraut. Dazu wurde folgender Ablauf geplant:

- Beschreibung des Ziels der Anomalieerkennung (Eigenschaften, die auf Fehler hindeuten können)
- visuelle Präsentation aller Anomalien
- Diskussion von Ausschnitten der Aufnahmen, die als anomal eingeschätzt werden
- gemeinsame Abstimmung der identifizierten anomalen Merkmale und der davon betroffenen Datensätze

Durch dieses Vorgehen soll sichergestellt werden, dass eine Auszeichnung von Datensätzen nicht nur auf einer Einzelmeinung beruht, sondern bestehende Erfahrungen mit den untersuchten Datensätzen mehrerer Personen berücksichtigt werden können. Die Art der Visualisierung der Aufnahmen soll sicherstellen, dass alle Aufnahmen betrachtet und berücksichtigt werden.

3.2.4 Visualisierung der EDA-Sensordaten

Um in einem begrenzten Zeitfenster alle Datensätze berücksichtigen zu können sollte sichergestellt werden, dass

- alle Datensätze zur Übersicht gezeigt werden können
- Details (Abschnitte in Zeit und Wertebereich) interaktiv und schnell dargestellt werden können

Dazu wurden zum einen eine Übersichtsgrafik mit einer Visualisierung der Verlaufs aller 49 Aufnahmen (Ausschnitt in Abbildung 3.6) erstellt, um ohne Ladezeit (die von einer Echtzeitvisualisierung benötigt würde) einen schnellen Eindruck über alle Aufnahmen zu ermöglichen.

Zum anderen wurde mit dem Framework *Dash* des Unternehmens *Plotly*³ eine interaktive Anwendung zur Visualisierung der Aufnahmen entwickelt. Wie in einem

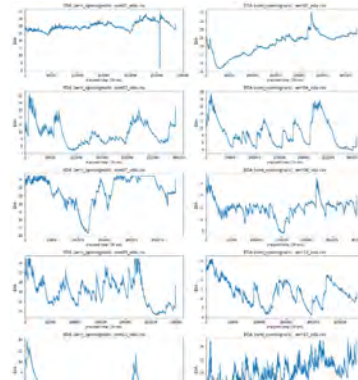


Abbildung 3.6: Ausschnitt der Visualisierung aller Aufnahmen der EDA

³<https://plot.ly/dash/>, Aufgerufen am 9.10.2019

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

Screenshot in Abbildung 3.7 dargestellt können Aufnahmen ausgewählt und einzeln oder gemeinsam über einer Zeitachse dargestellt werden. Mehrere gleichzeitig angezeigte Aufnahmen werden dabei durch unterschiedliche Farben, die in einer Legende aufgeführt sind, repräsentiert. Eine schnelle Auswahl von Ausschnitten, sowohl der Zeit als auch des Wertebereichs, ist durch eine Selektion von Bereichen möglich. Mit dieser Anwendung sollen Teilbereiche einer Aufnahme, die in der Übersichtsvisualisierung über ihre Nummer ausgewählt wurde genauer analysiert werden können.

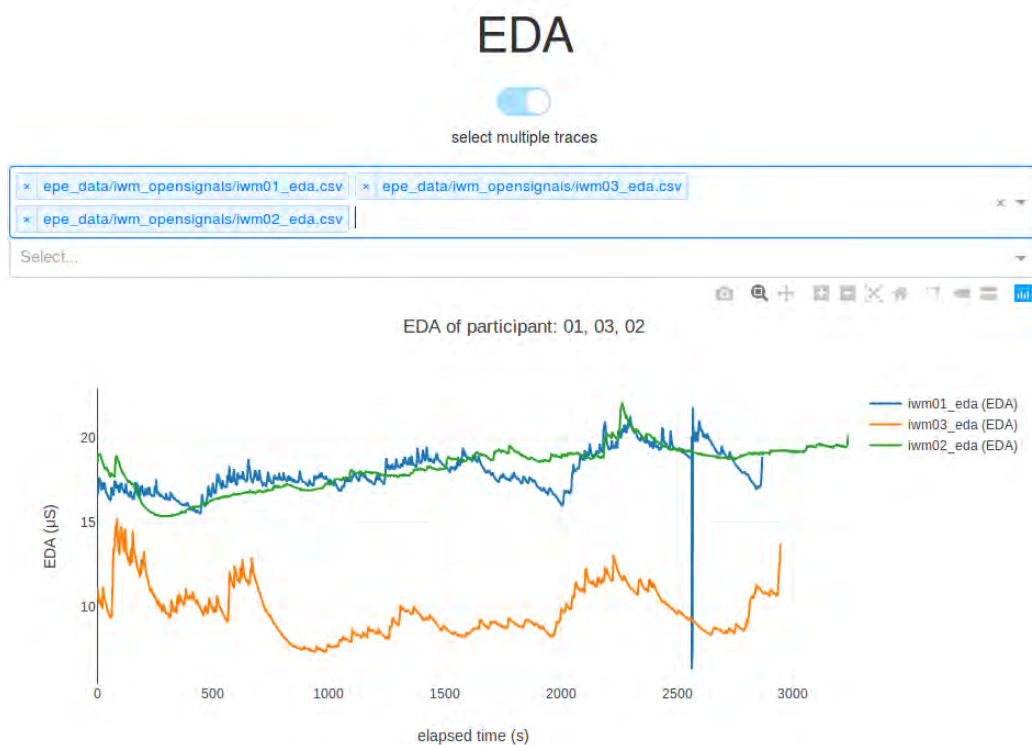


Abbildung 3.7: Entwickelte Anwendung zur interaktiven Visualisierung der Aufnahmen

3.2.5 Ergebnisse

Wie in Abschnitt 3.2.3 beschrieben wurde ein Treffen mit insgesamt vier Forschungsprojektteilnehmern, die in ihren Projekten mit den analysierten Datensätzen arbeiteten, genutzt um gemeinsam eine manuelle Identifikation anomaler Merkmale vorzunehmen. Den Teilnehmer wurden Zweck und

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

Als anomal bewertetes Merkmal	Teilnehmer
hohe Werte sind gekappt	7, 9, 15, 16, 56
kurzzeitiges starkes Abfallen der EDA	1
Existenz von Abschnitten mit auffallend niedriger Varianz (starkes Steigen/Fallen über die gesamte Aufnahme)	11, 14, 59 38 u.a.

Tabelle 3.2: Einschätzung von EDA-Aufnahmen des EPE-Experiments als Anomalien

Kontext der gesuchten Merkmale erläutert und Visualisierungen alle EDA-Aufnahmen gezeigt. Auffällige Aufnahmen wurden in der interaktiven Visualisierungsanwendung (Abbildung 3.7) detaillierter betrachtet und deren Merkmale diskutiert.

Die als Ergebnis entstandenen Merkmale und die betroffenen Datensätze sind in Tabelle 3.2 aufgeführt. Während die ersten beiden dort aufgeführten Merkmale (Kappung und Abfallen der EDA) von allen Teilnehmern als von möglicherweise von Messfehlern verursacht eingeordnet wurden, herrschte bei den Merkmalen *abschnittsweise niedrige Varianz* und *Steigen/Fallen über die gesamte Aufnahme* keine einheitliche Meinung über deren Ursache vor. Eine genaue Einordnung erfordert umfangreiches Domänenwissen in der Interpretation der elektrodermalen Aktivität und möglicherweise eine Analyse zusätzlicher Aufnahmen. In einer späteren Diskussion wurden statt einer abschnittswisen niedrigen Varianz eine hohe Varianz in Abschnitten als auffallendes Merkmal interpretiert, deshalb wurden bei der statistischen Modellierung in den folgenden Abschnitten beide Merkmale berücksichtigt.

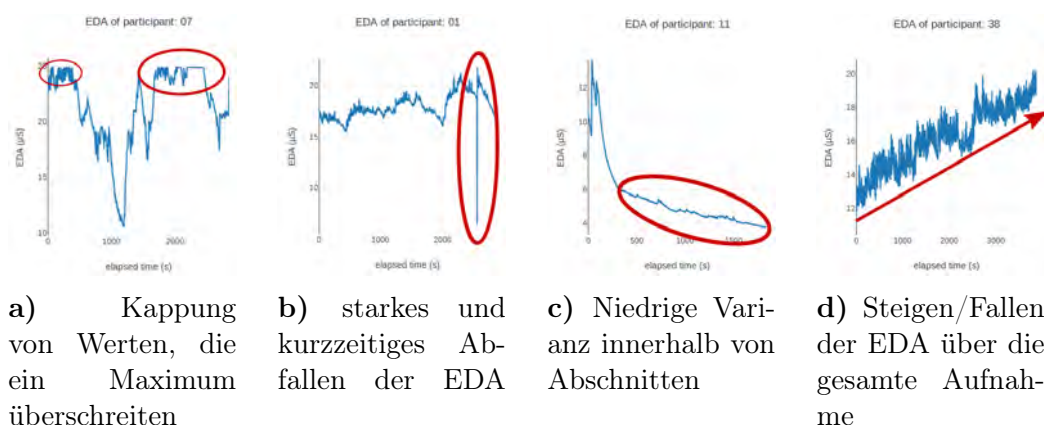


Abbildung 3.8: Beispielaufnahmen mit manuell identifizierten Anomalien, die jeweiligen Merkmale sind rot markiert

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

In Abbildung 3.8 sind beispielhaft Aufnahmen, die jeweils von einem der beschriebenen anomalen Merkmale betroffen sind, abgebildet. In diesen Aufnahmen treten diese Merkmale visuell besonders deutlich hervor, in anderen sind sie weniger auffällig vertreten.

3.3 statistische Analyse zur automatisierten Erkennung anomaler Merkmale

In den folgenden Unterabschnitten werden anomale Merkmale in EDA-Aufnahmen statistisch modelliert und Verfahren zu ihrer automatisierten Erkennung beschrieben.

3.3.1 kurzzeitiges Abfallen der elektrodermalen Aktivität

Von einem kurzzeitigen Abfallen der EDA war nur eine Aufnahme (iwm01, siehe Abbildung 3.8b) betroffen. Hier fiel die EDA in einem Zeitraum von etwa 5s von über $20 \mu\text{S}$ auf ca. $6\mu\text{S}$ und stieg wieder auf das ursprüngliche Niveau an. Da keine weiteren Aufnahmen mit diesem Problem im untersuchten Datensatz existierten, wird dieses Verhalten als repräsentativ für eine statistische Erkennung betrachtet.

Bei einem kurzzeitigen starken Ansteigen oder Abfallen steigt die Standardabweichung in einem diesen Vorgang umgebenden Zeitfenster stark an. Abbildung 3.9 zeigt zwei Aufnahmen im Vergleich, hier zeigt sich, dass die Standardabweichung innerhalb eines 3s-Fensters bei der betroffenen Aufnahme deutlich höher ist als bei der Vergleichsaufnahme *iwm16*, die über die gesamte Länge eine überdurchschnittlich hohe Standardabweichung besitzt. Dieses Merkmal wurde als Grundlage für eine statistische Erkennung mit einem Schwellenwert verwendet. Dabei wurde für jeder Aufnahme mit folgendem Vorgehen ein Skalarwert berechnet:

1. *MinMax-Normalisierung* für jede Aufnahme, damit sich alle Werte zwischen 0 und 1 befinden und eine Vergleichbarkeit zwischen Aufnahmen mit unterschiedlichem Niveau der EDA gewährleistet ist:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

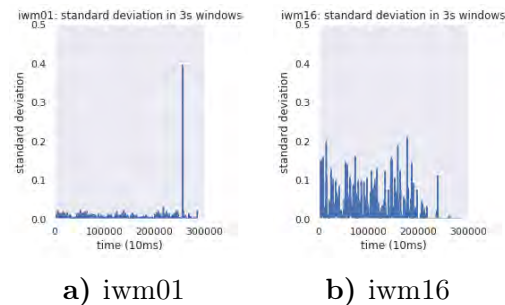


Abbildung 3.9: Standardabweichung in 3s Fenster bei Aufnahme mit (l) und ohne (r) plötzlichem Abfallen der EDA

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

2. Berechnen der Standardabweichung in einem Zeitfenster, dabei wurde für jeden Datenpunkt einer Aufnahme die Standardabweichung innerhalb eines ihn umgebenden n Sekunden langen Zeitfensters berechnet
3. Verwenden des Maximalwert jeder Aufnahme, der eine möglichen kurzzeitiges Abfallen/Ansteigen repräsentieren soll.

Um die Größe des Zeitfensters zu optimieren wurden Werte von 1s bis 15s getestet (siehe Abbildung 3.10), als Vergleichswert wurde dabei die Differenz des mit dem jeweiligen Zeitfensters gebildeten Maximalwerts zwischen der betroffenen Aufnahme (iwm01) und derjenigen mit dem zweithöchsten Maximalwert berechnet. Diese Differenz soll möglichst hoch sein, um die Unterscheidung von betroffenen und nicht betroffenen Aufnahmen zu maximieren. Sie war bei einem Zeitfenster von einer Sekunde am höchsten.

Mit dieser Größe des Zeitfenster wurde der Maximalwert der Standardabweichung für alle Aufnahmen berechnet, in Abbildung 3.11 ist eine sortierte Liste aller Aufnahmen abgebildet, die Aufnahme mit dem anomalen Merkmal besitzt den höchsten Wert und kann mit einem Schwellenwert, beispielsweise bei einer maximalen Standardabweichung von mehr als 0.2, eindeutig identifiziert werden.

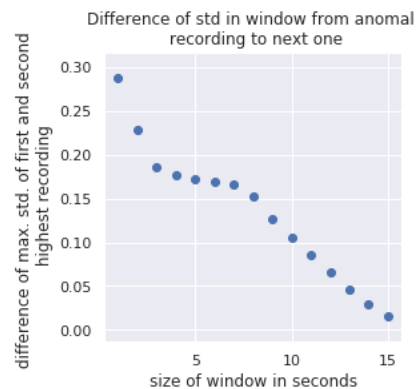


Abbildung 3.10: Differenz der Maximalwerte der Standardabweichung in unterschiedlich großen Zeitfenster zwischen der betroffenen und der nächsthöher bewerteten Aufnahme

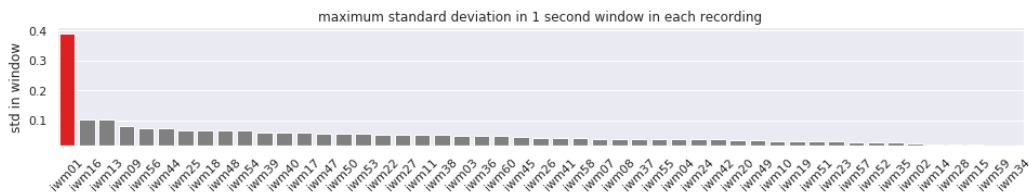


Abbildung 3.11: Aufnahmen, sortiert nach ihrer maximalen Standardabweichung in 1s-Zeitfenster, Aufnahme mit anomalem Merkmal rot markiert

Da im Datensatz nur eine betroffene Aufnahme zur Verfügung stand, müssen Schwellenwert und gewähltes Zeitfenster in Annahme der Repräsen-

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

tativität der Aufnahme gewählt wären; mit mehr betroffenen Aufnahmen könnten weitere Optimierungen durchgeführt werden.

3.3.2 gekappte Maximalwerte der elektrodermalen Aktivität

In Aufnahmen wie in Abbildung 3.8a dargestellt, sind alle Werte über einem Maximalwert gekappt. Um solche Aufnahmen statistisch zu erkennen sollen solche, bei denen sich in einem hohen Quantil nur Maximalwerte befinden, klassifiziert werden.

Mit folgendem Vorgehen wurden die Aufnahmen klassifiziert:

1. *MinMax-Normalisierung* für jede Aufnahme, damit sich alle Werte zwischen 0 und 1 befinden und eine Vergleichbarkeit zwischen Aufnahmen mit unterschiedlichem Niveau der EDA gewährleistet ist:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

2. für jede Aufnahme wird ein n-Quantil untersucht, Aufnahmen, in denen sich in diesem Quantil nur Datenpunkte mit Maximalwert (1.0) befinden, werden als betroffene interpretiert
3. das Quantil wird erhöht, bis weitere Aufnahmen im untersuchten Quantil bei einem Wert von 1.0 liegen. Diese neuen Aufnahmen wurden manuell auf Anomalien untersucht. Wenn sie keine aufweisen, wird abgebrochen

In Tabelle 3.3 werden die gefundenen Quantile, ab denen neue Anomalien gefunden wurden, aufgelistet. Das jeweilige Quantil gibt ebenso an, welcher Anteil einer Aufnahme nicht auf den Maximalwert gekappt ist.

Die bei der manuellen Analyse gefundenen Aufnahmen (7, 9, 15, 16, 56) wurden im 98,1%-Quantil entdeckt. Zusätzlich konnten weitere Aufnahmen (13, 34, 41, 53) ab einem Quantil von 99,5% entdeckt werden. Diese sind in Abbildung 3.13 in einer Übersicht dargestellt. Hier sind die gekappten Maximalwerte zunächst nicht visuell erkennbar, bei Betrachtung eines Abschnitts in vergrößerter Ansicht (beispielsweise bei Aufnahme 41 in Abbildung

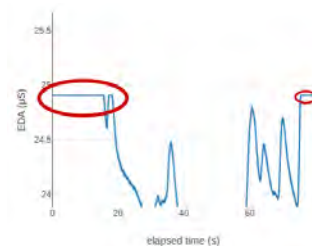


Abbildung 3.12: Ausschnitt von Aufnahme 41, mit abgeschnittenen Werten (rot markiert)

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

Quantil (gerundet)	identifizierte Aufnahmen
69,30%	16
86,10%	16, 7
94,50%	16, 7, 15
98,10%	16, 7, 15, 9, 56
99,20%	16, 7, 15, 9, 56, 53, 34
99,30%	16, 7, 15, 9, 56, 53, 34, 41
99,50%	16, 7, 15, 9, 56, 53, 34, 41, 13

Tabelle 3.3: Schwellenwerte für das Quantil, ab die in der zweiten Spalte gelisteten Aufnahmen identifiziert werden können

3.12) fallen jedoch kleine betroffene Bereiche auf.

Bei allen identifizierten Aufnahmen außer Nummer

34 befindet sich der abgeschnittene Maximalwert zwischen 24 und 25 μS , bei dieser jedoch bei etwa 5 μS , möglicherweise existiert bei dieser Aufnahme ein zeitweises Plateau auf dem Maximalwert, das von einer anderen Ursache bedingt ist.

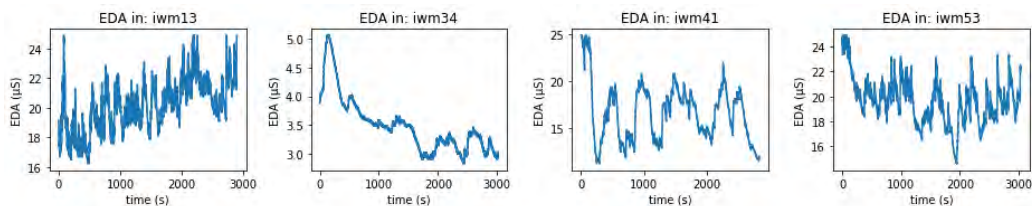


Abbildung 3.13: Aufnahmen mit abgeschnittenen Werten, die visuell nur in stark vergrößerter Ansicht auffällig sind

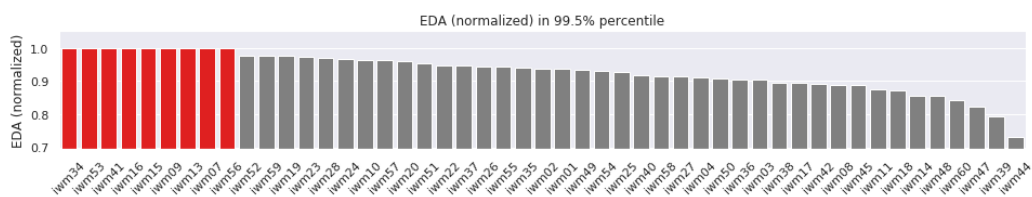


Abbildung 3.14: normalisierter EDA-Wert im 99,5%-Quantil aller Aufnahmen, sortiert, Aufnahmen mit Maximalwert im Quantil sind rot markiert

In Abbildung 3.14 ist der normalisierte EDA-Wert aller Aufnahmen im 99,5%-Quantil sortiert dargestellt. Aufnahmen, in denen sich in diesem Quan-

til nur Datenpunkte mit Maximalwert (1.0) befinden, sind rot markiert und werden als Anomalie erkannt.

3.3.3 hohe und niedrige Varianz der elektrodermalen Aktivität

Wie in Tabelle 3.2 dargestellt, wurde eine abschnittsweise vergleichsweise niedrige Varianz in einer Aufnahme von mehreren Teilnehmern des Forschungsprojektes als anomales Merkmal beurteilt, dessen Ursache ein Fehler in der Messung der EDA sein könnte. Anders als die beiden in den vorigen Abschnitten analysierten Merkmale, *gekappte Maximalwerte* (Abschnitt 3.3.2) und *kurzzeitiges Abfallen* (Abschnitt 3.3.1), ist dieses Merkmal jedoch weniger eindeutig abgrenzbar. Deshalb wird statt einer automatisierten Klassifizierung von Aufnahmen mit einer hohen oder niedrigen Varianz eine Sortierung der Aufnahmen umgesetzt. Dadurch wird eine gezielte manuelle Analyse durch Experten mit Domänenwissen in der Interpretation der EDA sowohl nach hoher als auch niedriger Varianz erleichtert.

Als Ziel einer automatisierten Sortieren wird ein Ranking der manuell identifizierten Aufnahmen (11, 14, 59) an höchster Stelle definiert. Dieses wurde mit folgender Vorgehensweise erreicht:

1. *MinMax-Normalisierung* für jede Aufnahme, damit sich alle Werte zwischen 0 und 1 befinden und eine Vergleichbarkeit zwischen Aufnahmen mit unterschiedlichem Niveau der EDA gewährleistet ist:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

2. Berechnen der Varianz in einem rollendem Zeitfenster (120s). Getestet wurden unterschiedliche Zeitfenster von 10s bis 300s, bei allen wurde das Ziel erreicht.
3. Berechnen des Medienwerts der Varianz in den Zeitfenstern
4. Aufsteigende Sortierung der Aufnahmen nach dem Median

Wie im Ergebnis in Abbildung 3.15 zu sehen ist, stehen bei einer aufsteigenden Sortierung der Aufnahmen nach diesem Verfahren die manuell identifizierten Aufnahmen an den ersten drei Positionen.

Die Aufnahmen mit der niedrigsten abschnittswisen Varianz sind in Abbildung 3.16 dargestellt, bei einer absteigenden Sortierung können die Aufnahmen mit der höchsten automatisiert erkannt werden (Abbildung 3.17). Hier ist schon im visuellen Vergleich beider Aufnahmen ein deutlicher Unterschied erkennbar.

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

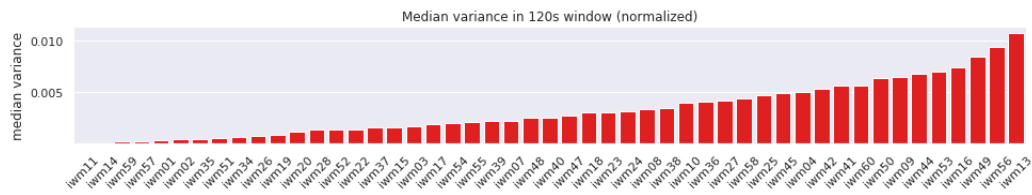


Abbildung 3.15: Aufnahmen sortiert nach dem Median der Varianz in einem Zeitfenster von 120s

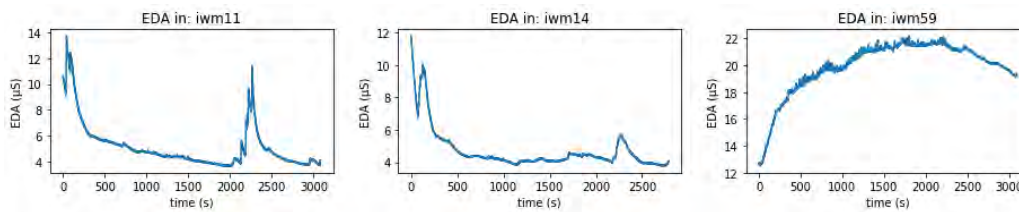


Abbildung 3.16: Die drei Aufnahmen mit niedrigster Varianz

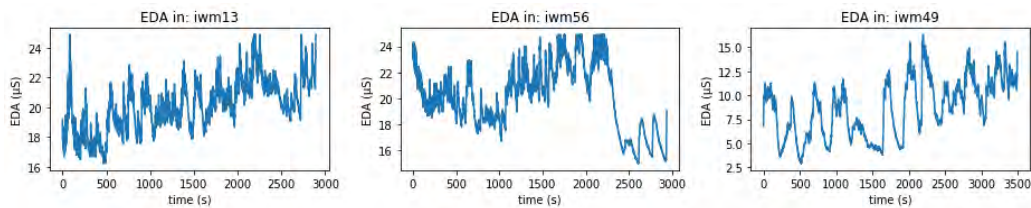


Abbildung 3.17: Die drei Aufnahmen mit höchster Varianz

3.3.4 starkes Ansteigen oder Abfallen über die gesamte Aufnahme

Bei der manuellen Identifikation von anomalen Merkmalen wurde auch das deutliche Ansteigen und Abfallen der EDA innerhalb einer Aufnahme genannt. Da nicht alle beteiligten Teilnehmer von einem Fehler der Messung als Ursache ausgingen, wurde nur eine betroffene Aufnahme (38) beispielhaft genannt.

Um eine Sortierung der Aufnahmen nach absoluter Steigung umzusetzen wurde folgendes Verfahren gewählt:

1. *MinMax-Normalisierung* für jede Aufnahme, damit sich alle Werte zwischen 0 und 1 befinden und eine Vergleichbarkeit zwischen Aufnahmen mit unterschiedlichem Niveau der EDA gewährleistet ist:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT

2. Berechnung der Steigung der Geraden durch Lineare Regression mit dem geringsten quadrierten Fehler E (least square error):

$$E = \sum_{i=0}^n |p(x_i) - y_i|^2$$

3. Verwendung des Betrags des Steigungskoeffizienten um sowohl das Ansteigen als auch das Abfallen zu berücksichtigen
4. Erneute MinMax-Normalisierung, allerdings zwischen den Aufnahmen, da die Steigungskoeffizienten teils sehr kleine Werte gegen Null sind
5. Absteigende Sortierung

Eine Übersicht aller Geraden, die durch lineare Regression jeweils eine Aufnahme repräsentieren, ist in Abbildung 3.18 dargestellt.

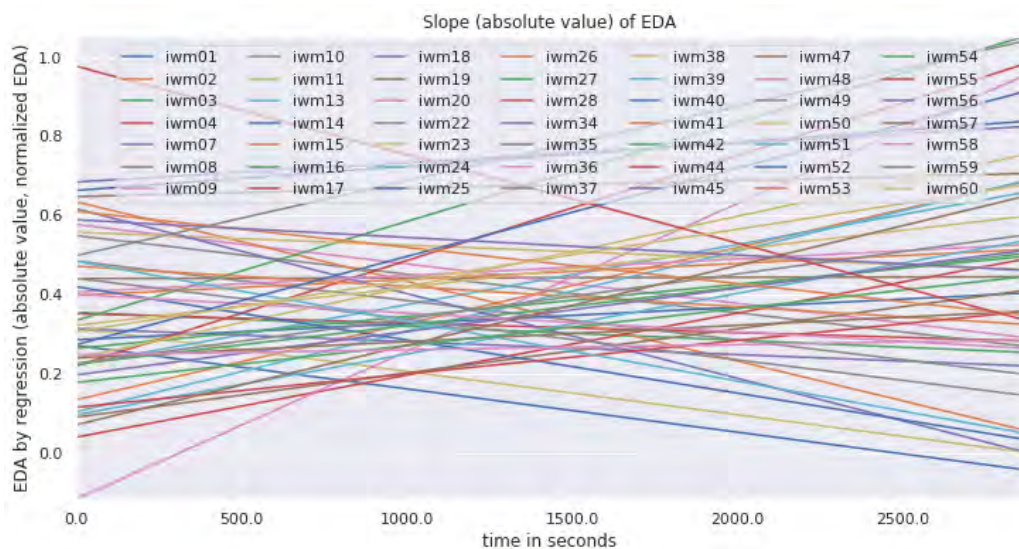


Abbildung 3.18: Übersicht der durch lineare Regression erhaltenen Geraden, die jeweils eine Aufnahme repräsentieren

Als Ergebnis wurden die absoluten Steigungen absteigend sortiert, diesem in Abbildung 3.19 dargestelltem Ranking sind an den ersten Stellen die Aufnahmen mit der höchsten absoluten Steigung zu entnehmen.

In Abbildung 3.20 sind vier Aufnahmen mit der höchsten absoluten Steigung dargestellt. Unter diesen Aufnahmen ist auch die Aufnahme 16, die schon bei der Analyse auf den Maximalwert gekappter Werte auffiel, da ein großer Teil (fast 30%) der Aufnahme auf Maximalwerten besteht.

KAPITEL 3. ANALYSE VON ANOMALIEN IN AUFNAHMEN DER ELEKTRODERMALEN AKTIVITÄT



Abbildung 3.19: Absolutwerte der Steigungen aller Aufnahmen, absteigend sortiert

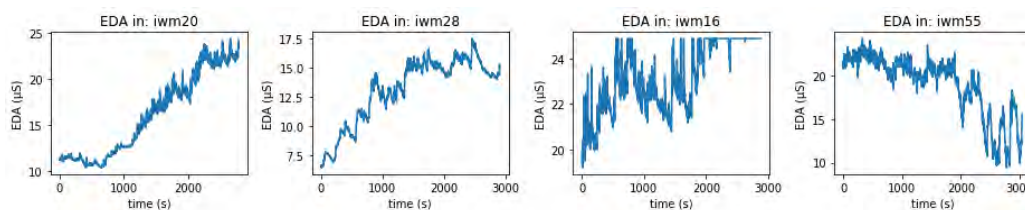


Abbildung 3.20: Vier Aufnahmen mit der höchsten absoluten Steigung

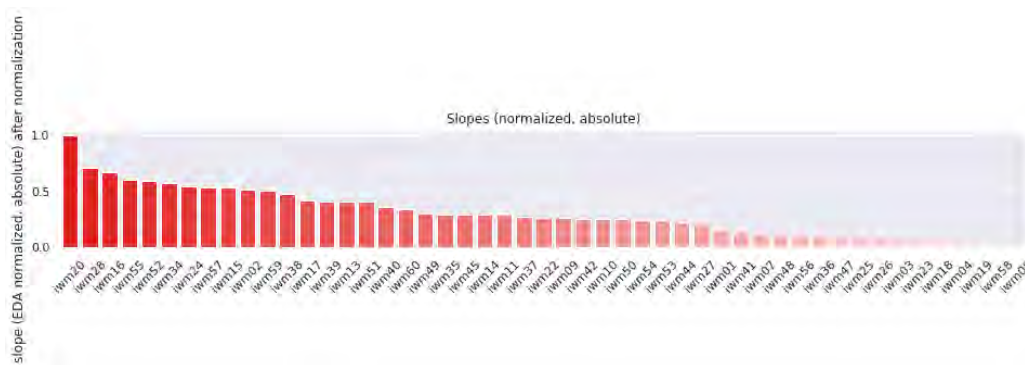


Abbildung 3.21: Absolutwert der Steigung aller Aufnahmen, normalisiert über alle Aufnahmen

Da sich die Steigungskoeffizienten über die gesamte Aufnahme in sehr niedrigen Wertebereichen nahe Null befinden wurde für die Ergebnisse eine weitere MinMax-Normalisierung über alle Aufnahmen durchgeführt um bei einer späteren Weiterverarbeitung innerhalb eines Wertebereiches zwischen Null und Eins arbeiten zu können (Abbildung 3.21).

Kapitel 4

Bewertung und Ausblick

4.1 Datenmodellierung von Lern- und Sensordaten

In den folgenden Unterabschnitten werden jeweils beide Teile des Forschungsprojektes B zusammengefasst und bewertet sowie ein Ausblick auf mögliche Fortführungen der Arbeit gegeben.

Zusammenfassung

Aufbauend auf dem Entwurf eines dokumentenbasierten Datenmodells im Teil A des Forschungsprojekts wurde dieses Datenmodell erweitert, mit einer MongoDB-Datenbank umgesetzt und mit Lerndaten von Datenbankabbilds von LEMO2 sowie mit Sensordaten des Emotional Picture Experiments befüllt. Zum Datenzugriff wurde eine schlanke Webanwendung zur Umsetzung eines eigenständigen und unabhängigen Datenbackends entwickelt, die in einem Docker-Container virtualisiert wurde.

Bewertung

Mit einem dokumentenorientierten Datenmodell ist nun

- eine einfache Änderung des Datenmodells wie die Ergänzungen von Metainformationen und neuen Entitäten möglich
- eine Nutzung der horizontale Skalierbarkeit dokumentenbasierter Datenbanken möglich
- existiert ein eigenständiges Datenbackend, das auch unabhängig von LEMO2 genutzt werden kann
- die Aufteilung von Sensordaten auf mehrere Dokumente möglich

Damit wurde der Entwurf des in Teil A des Forschungsprojektes entwickelten Datenmodells umgesetzt.

Ausblick

Noch nicht realisiert wurde eine Integration in LEMO2, die neben dem Datenzugriff auch eine gemeinsame Authentifizierung beinhaltet. Durch die Entwicklung als eigenständige Anwendung ist ein Einsatz auch unabhängig von LEMO2 zur Persistierung von Sensordaten möglich, die leichte Anpassbarkeit des Datenmodells erlaubt die schnelle Erweiterung für individuelle Einsatzzwecke.

4.2 Anomalieerkennung

Zusammenfassung

Ziel dieses Teils war die automatisierte Identifikation von EDA-Aufnahmen mit anomalen Merkmalen des EDA. Dazu wurde nach einer explorativen Datenanalyse aufgrund der Eigenschaften dieser Aufgaben zunächst eine manuelle Identifikation solcher Merkmale und der von ihnen betroffenen Aufnahmen durchgeführt (Abschnitt 3.2).

Im Anschluss wurden diese Merkmale statistisch modelliert um eine automatisierte Erkennung von sie enthaltenden Aufnahmen ermöglichen. Dadurch konnten weitere Aufnahmen, die bei einer manuellen Identifikation aufgrund der visuell weniger auffälligen Ausprägung der Merkmale nicht entdeckt wurden, gefunden werden (Merkmal der gekappten Maximalwerte, siehe Abschnitt 3.3.2).

Bewertung und Ausblick

Die beiden anomalen Merkmale, die von den befragten Forschungsprojektteilnehmern als Ergebnis von Fehlern eingeordnet wurden, treten bei 9 von 49 Datensätzen auf. Damit wären über 18% der Datensätze betroffen, ein so hoher Anteil würde probabilistische Verfahren zur Anomalieerkennung, die selten vorkommenden Anomalien identifizieren, eher ungeeignet erscheinen lassen. Durch die manuelle Identifikation der Merkmale war zudem schon zu Beginn eine eindeutige Beschreibung und damit die Möglichkeit der Kommunikation der Merkmale vorhanden, bei einer automatisierten Suche nach unbekanntem Abweichungen, wäre eine Interpretation der Ergebnisse schwieriger.

Sollte die Ursache der entdeckten Anomalien in Messfehlern liegen, könnten sie zu schlechteren Ergebnissen bei der Vorhersage von akademischen Emotionen führen. Allerdings sind nur Teilbereiche der Aufnahmen betroffen und die Auswirkungen dürften geringer sein, als der hohe Anteil der betroffenen Aufnahmen von über 18% zunächst vermuten lassen könnte.

Mit anderen Forschungsprojektteilnehmern wurde eine Testmessung mit EDA- und EKG-Sensoren durchgeführt, die anomalen Merkmale traten dabei nicht auf, um sie zu reproduzieren und damit mehr über ihre Ursache – und mögliche Behebung – herauszufinden wären weitere Versuche oder die Untersuchung weiterer Aufnahmen notwendig. Da nun statistische Verfahren zur automatisierten Erkennung der anomalen Merkmale existieren ist die Analyse weiterer Aufnahmen deutlich schneller und einfacher möglich.

KAPITEL 4. BEWERTUNG UND AUSBLICK

Die beiden anomalen Merkmale

- starkes Absteigen/Fallen in der gesamten Aufnahme
- hohe/niedrige Varianz in Abschnitten

könnten auf Fehler hindeuten, eine Interpretation erfordert jedoch weiteres Domänenwissen über die Interpretation der EDA; eine Diskussion zwischen den Forschungsprojektteilnehmern kam zu keinem abschließenden Ergebnis. Eine Untersuchung weiterer Aufnahmen auf diese Merkmale könnte weitere hilfreiche Informationen ergeben.

Literaturverzeichnis

- [1] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.
- [2] Jason J Braithwaite, Derrick G Watson, Robert Jones, and Mickey Rowe. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology*, 49(1):1017–1034, 2013.
- [3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.
- [4] T. Dunning and E. Friedman. *Practical Machine Learning: A New Look at Anomaly Detection*. O’Reilly Media, 2014.
- [5] David J Hill and Barbara S Minsker. Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. *Environmental Modelling & Software*, 25(9):1014–1022, 2010.
- [6] P Lang and Margaret M Bradley. The international affective picture system (iaps) in the study of emotion and attention. *Handbook of emotion elicitation and assessment*, 29, 2007.
- [7] Egger Maria, Ley Matthias, and Hanke Sten. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019.
- [8] K.G. Mehrotra, C.K. Mohan, and H.M. Huang. *Anomaly Detection Principles and Algorithms*. Terrorism, Security, and Computation. Springer International Publishing, 2017.
- [9] Ricarda Mertens. *Aussagekraft der elektrodermalen Aktivität in Laborexperimenten mit Schwerpunkt Lärm-Literaturstudie zu wichtigen Einflussfaktoren und gesundheitlichen Implikationen*. PhD thesis, Universitäts- und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf, 2016.

LITERATURVERZEICHNIS

- [10] Fadi Moukayed, Haeseon Yun, Tom Bisson, and Albrecht Fortenbacher. Detecting academic emotions from learners' skin con-ductance and heart rate: A data-driven approach using fuzzy logic. 2018.
- [11] Osman Salem, Alexey Guerassimov, Ahmed Mehaoua, Anthony Marcus, and Borko Furht. Sensor fault and patient anomaly detection and classification in medical wireless sensor networks. In *2013 IEEE international conference on communications (ICC)*, pages 4373–4378. IEEE, 2013.
- [12] Haeseon Yun, Albrecht Fortenbacher, Niels Pinkwart, Tom Bisson, and Fadi Moukayed. A pilot study of emotion detection using sensors in a learning context: Towards an affective learning companion. In *DeLFI/GMW Workshops*, 2017.