

Datenbeschreibung

Data Mining Projekt

Eyetracking

Deckblatt:

- SG, THB, FB ...
- Leo
- Seminar.
- Tex?

Gruppenmitglieder:

Herval Bernice Nganya Nana
Mario Kaulmann

Inhaltsverzeichnis

1 Aufgabenstellung.....	3
2 Datenbeschreibung.....	4
3 Arbeitsschritte.....	5
4 Auffälligkeiten.....	7
5 Fragen.....	8
6 Merkmalserzeugung.....	9
6.1 Merkmale innerhalb der Datenströme.....	9
6.2 Merkmale Versuchsperson.....	9
7 Clustern.....	11

1 Aufgabenstellung

In einem Experiment wurden Eyetracking-Daten erhoben, bei denen die Versuchspersonen 3 verschiedene Versuche durchführen sollten. Dabei sollten die Versuchsperson mit den Augen einem Punkt folgen, der eine spezielle Figur zeichnete. Diese Figuren sind ein liegende acht und eine horizontale Linie.

- Der erste Versuch ist die liegende acht langsam (acht Sekunden für einen Durchlauf).
- Der zweite Versuch ist die liegende acht schnell (vier Sekunden für einen Durchlauf).
- Der dritte Versuch ist die horizontale Linie (vier Sekunden für einen Durchlauf).

Für jeden Versuch wurden zwei Messungen gemacht und für die liegende Acht langsam wurde zusätzlich ein Probendurchlauf gemacht.

Die Aufgabe besteht darin die Versuchspersonen zu gruppieren (clustern). Dabei sollen mit Hilfe der erhobenen Daten Merkmale gefunden werden, die es ermöglichen entsprechende Gruppen zu bilden.

?

Aufgabe bearbeitet am 30.11.

1. Cluster
2. concept description
3. Merkmale

2 Datenbeschreibung

Die Daten umfassen für 134 Versuchspersonen jeweils eine Datei mit Daten zu den gemessenen Blickpositionen (Blickdatei) und eine Datei mit den Positionsdaten des Zielpunktes (Targetdatei).

Die Blickdateien enthalten folgende Attribute:

Attribut	Wert
Zeitstempel	Ganze Zahl positiv <i>Zeitreihen</i>
Blick linkes Auge x-Koordinate	Fließkommazahl
Blick linkes Auge y-Koordinate	Fließkommazahl
Pupillengröße linkes Auge	Kann ignoriert werden
Position linkes Auge vor Eyetracker x-Koordinate	Kann ignoriert werden
Position linkes Auge vor Eyetracker y-Koordinate	Kann ignoriert werden
Entfernung linkes Auge vor Eyetracker	Kann ignoriert werden
Blick rechtes Auge x-Koordinate	Fließkommazahl
Blick rechtes Auge y-Koordinate	Fließkommazahl
Pupillengröße rechtes Auge	Kann ignoriert werden
Position rechtes Auge vor Eyetracker x-Koordinate	Kann ignoriert werden
Position rechtes Auge vor Eyetracker y-Koordinate	Kann ignoriert werden
Entfernung rechtes Auge vor Eyetracker	Kann ignoriert werden

Tabelle 1: Attribute Blickdatei

Die Targetdateien haben folgende Attribute:

Attribut	Wert
Zeitstempel	Ganze positive Zahl
t_soll	Kann ignoriert werden
t_ist	Kann ignoriert werden
pix_x	Fließkommazahl
pix_y	Fließkommazahl
deg_x	Kann ignoriert werden
deg_y	Kann ignoriert werden

Tabelle 2: Attribute Targetdatei

Eine Blickdatei enthält zusätzlich zu den erfassten Blickpositionen noch Eventeinträge. Diese Eventeinträge haben auch einen Zeitstempel und unterteilen die Datei in verschiedene Phasen des Experiments.

Event
↓
Sequenz
↓
Protokoll

3 Arbeitsschritte

Um ein besseres Verständnis für die Daten zu bekommen haben wir diese zunächst so zerlegt, dass die Blickdaten in kleineren Dateien gespeichert wurden, die nur noch jeweils die Daten für einen Versuch enthalten. Die Targetdateien wurden ebenfalls so zerlegt, dass kleine Dateien entstanden, die die Daten für einen Versuch pro Versuchsperson enthalten.

Die Daten wurden dann entsprechend neu strukturiert gespeichert, sodass es zu jeder Versuchsperson einen Ordner gibt, in dem die Versuche enthalten sind, in denen wiederum die Durchgänge enthalten sind. Dabei entstand folgende Struktur:

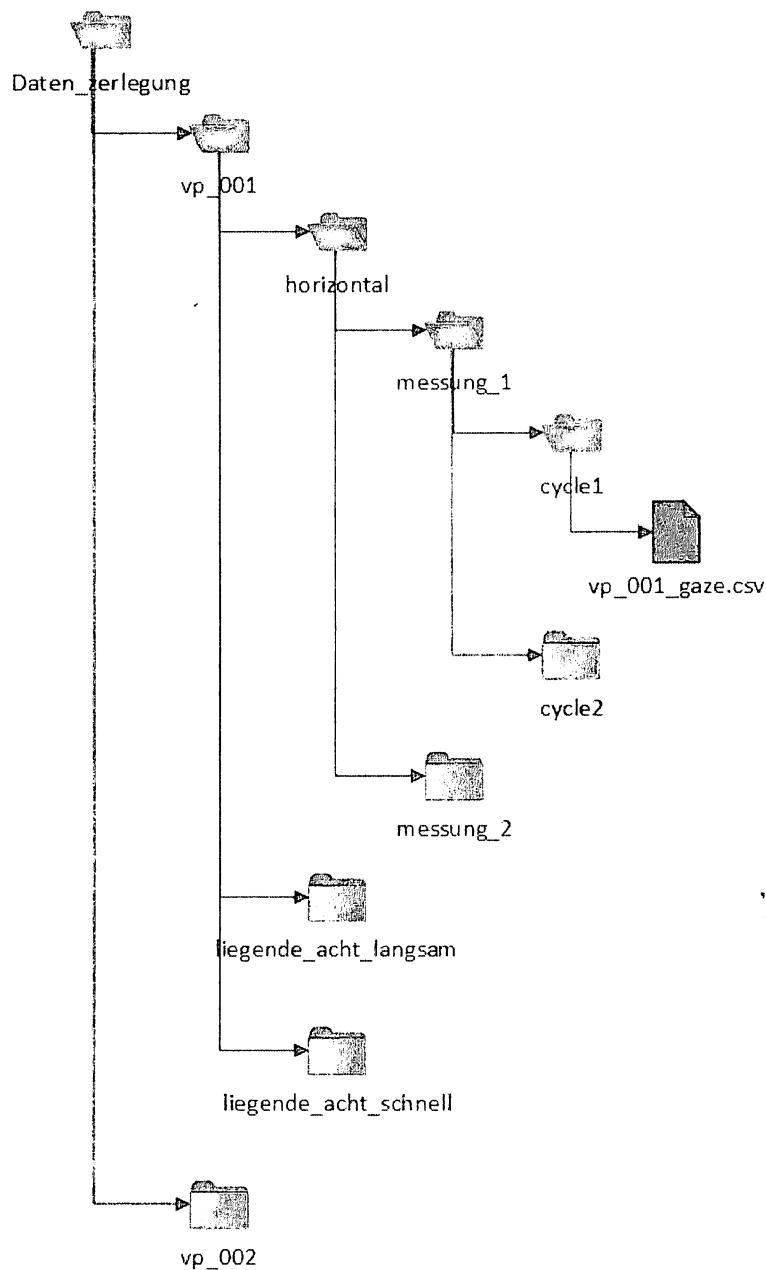


Abbildung 1: Ordnerstruktur der Daten

Die Versuche wurden dann visualisiert.

Durch die Visualisierung konnte man besonders gut erkennen, dass die Targetdateien und die Blickdateien auf unterschiedlichen Koordinatensystemen beruhen.
hier mehrere Fixierpunkte

Das Bild nach der Visualisierung sah wie folgt aus.

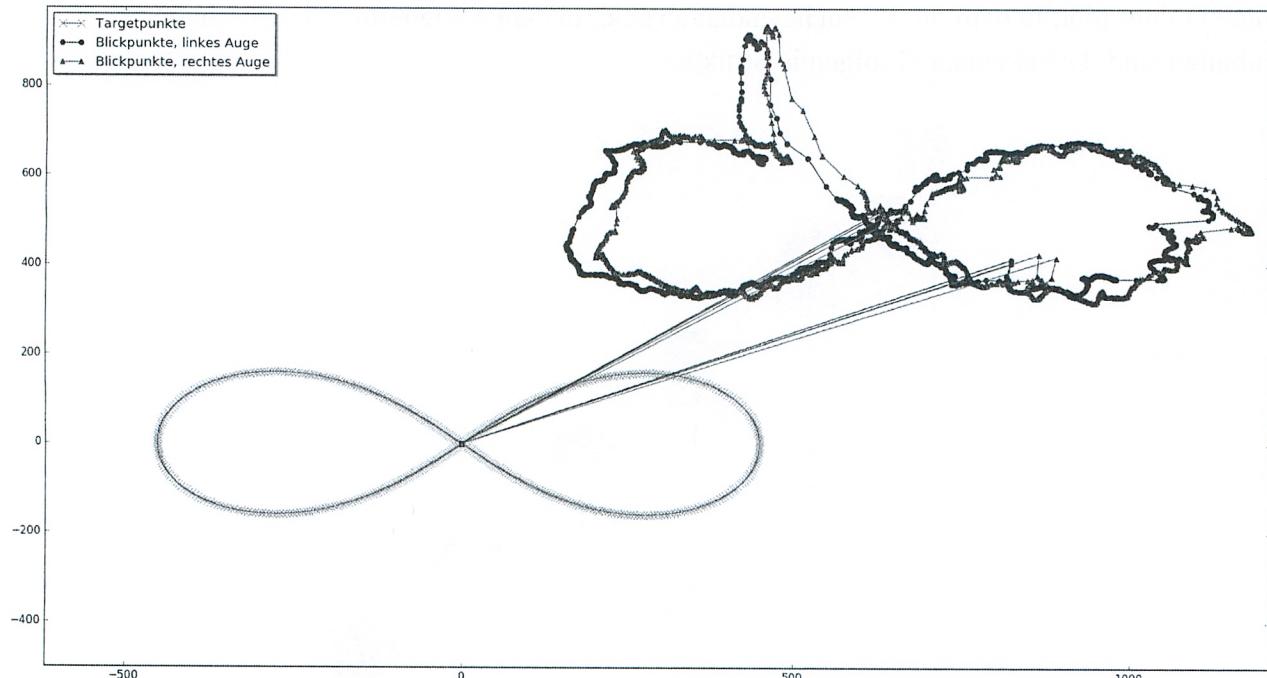


Abbildung 2: Visualisierung der Blickpunkte

Als nächste Arbeitsschritte folgen die Merkmalserzeugung und die Clusterung. Für diese Schritte gibt es jeweils eigene Kapitel.

Exploration!

4 Auffälligkeiten

- Bei den Daten ist bei Versuchsperson 131 aufgefallen, dass in der Targetdatei die Einträge

MSG:UserQuit:ProgramPaused

MSG:UserQuit:ProgramRestarted

in der Spalte für den Zeitstempel vorkommen.

jewahr

- Die Blickpunkte wurden mit einer höheren Frequenz gemessen, als die Targetpunkte. Dadurch gibt es zwischen den Targetpunktmessungen mehrere Blickpunktmessungen.
- Es gibt in den Blickpunktdateien keinen Zeitstempel, der identisch ist mit dem aus einer dazugehörigen Targetpunktdatei. Dadurch erschwert sich die Zuordnung der Blickpunkte zu den Targetpunkten.
- In den Blickpunktdateien tritt oft der Fall auf, dass die x-Werte und die y-Werte für die Augen 0 sind.
- Bei der Visualisierung hat sich gezeigt, dass der Koordinatenursprung bei den Blickdateien am linken oberen Rand des Bildschirms zu sein scheint, mit aufsteigenden Werten nach unten auf der Y-Achse und aufsteigenden Werten nach rechts auf der X-Achse. Das Koordinatensystem der Targetdateien liegt vermutlich im Mittelpunkt des Bildschirms mit aufsteigenden Werten nach oben auf der Y-Achse und aufsteigenden Werten nach rechts auf der X-Achse. Das lässt sich daraus ableiten, dass die Blickpunkte bei der Animation auf der Y-Achse immer in die entgegengesetzte Richtung zu der der Targetpunkte ging. Auf der X-Achse stimmte die Richtung immer überein.

- Fehlende Werte bei Blick, weil nichts, was nichts
- versch. Zeilenumbrüche \rightarrow kodiert mit \emptyset
- Blick ohne Headervile
- Zeitstempel nicht aquivalent mit vielen
-

5 Fragen

Aus den Auffälligkeiten ergeben sich folgende Fragen:

1. Welche Bedeutung haben die 0 Werte in den Blickdateien?
Bedeuten diese, dass die Versuchsperson die Augen geschlossen hatte?
2. Um welchen Wert sind die Koordinatensysteme verschoben?
Gibt es eine Streckung des Koordinatensystems der Blickwerte zu dem der Targetwerte?
3. Liefen die Messprogramme auf dem selben Rechner?
4. Können wir davon ausgehen, dass die Zeitstempel der Blickdateien und der Targetdateien von synchron laufenden Uhren erstellt wurden?
5. Können wir voraussetzen, dass der Fokus beim Sehen auf dem Mittelpunkt von der Blickposition des rechten und des linken Auges liegt?

Plan/Konzeption

6 Merkmalserzeugung

Bei der Merkmalserzeugung muss zwischen Merkmalen unterschieden werden, die innerhalb der Datenströme liegen und Merkmalen, die für die Gesamtbeschreibung der Versuchsperson genutzt werden.

Wertreihe oder genauere Teilreihe oder Signal

Abgeleitete Wertreihen

6.1 Merkmale innerhalb der Datenströme

Ein Merkmal innerhalb der Datenströme ist zum Beispiel die Mitte zwischen Blickposition des linken Auges und Blickposition des rechten Auges, die in Kapitel 3 bereits beschrieben wurde. Dabei wird für jeden Zeitpunkt im Datenstrom ein jeweiliger Wert berechnet.

Nachfolgend sind die Merkmale, die erzeugt werden können aufgelistet.

Merkmal	Berechnung
Mitte Augenpositionen	$x = \frac{xLinks + xRechts}{2}; y = \frac{yLinks + yRechts}{2}$
Abweichung Augenposition linkes Auge zu Targetpunkt	$x = \frac{xLinks + xTarget}{2}; y = \frac{yLinks + yTarget}{2}$
Abweichung Augenposition rechtes Auge zu Targetpunkt	$x = \frac{xRechts + xTarget}{2}; y = \frac{yRechts + yTarget}{2}$
Abweichung Mitte Augenposition zu Targetpunkt	$x = \frac{xMitte + xTarget}{2}; y = \frac{yMitte + yTarget}{2}$

Tabelle 3: Merkmale innerhalb des Datenstroms

*- Ableitung \rightarrow Geschwindigkeit
- Korrelation*

6.2 Merkmale Versuchsperson

Ein Merkmal zu einer Versuchsperson ist ein statistischer Wert über den gesamten Zeitraum der Messung. Diese Werte sind aus den Merkmale innerhalb der Datenströme abgeleitet. In der Regel handelt es sich um die Werte Maximum, Minimum, Durchschnitt, Median, Varianz, Standardabweichung. Diese Werden für die Merkmale innerhalb der Datenströme berechnet und spiegeln dadurch eine Versuchsperson wider.

Um die Daten aussagekräftiger zu bekommen, kann man die statistischen Merkmale nochmal bezogen auf den einzelnen Versuch pro Versuchsperson berechnen.

Nachfolgend sind beispielhaft ein paar Merkmale aufgelistet.

- Durchschnittliche Abweichung liegende Acht langsam
- Durchschnittliche Abweichung liegende Acht schnell
- Durchschnittliche Abweichung horizontale

1) Statistische Kenngrößen der Werteverteilung einer Reihe

*2) dargestellt für Phasen / Versuche / Figuren
↳ auch Pausen als Segment*

*Merkmale nur aus der Augenbewegung
" aus Augenbew + TARGET*

- Standardabweichung liegende Acht langsam
- Maximale Abweichung liegende Acht langsam
- Minimale Abweichung rechtes Auge liegende Acht langsam
- Maximale Abweichung linkes Auge liegende Acht langsam

7 Clustern

Beim Clustern werden die Versuchspersonen mittels der Merkmale, die für die Versuchspersonen erzeugt wurden geclustert (gruppiert). Dafür wird das Verfahren DBScann verwendet.

Um die entstandenen Cluster nutzbar zu machen, werden diese beschrieben. Dadurch soll es möglich werden neue Versuchspersonen entsprechend den Clustern zuordnen zu können und die Zuordnung der aktuell vorhandenen Versuchspersonen wiederholbar zu machen. Das Ergebnis der Clusterbeschreibung soll ein Entscheidungsbaum sein.

wegen Transparenz