

Analyse & Klassifikation von Blickmustern

Auf Basis einer Datenerhebung der Universität Hamburg

David Klehr 757914, Jörn Malich 752312

8. Juni 2013

Inhaltsverzeichnis

1	Gegenstand	3
2	Dyadenklassifikation anhand der Aufenthaltswahrscheinlichkeit	3
2.1	Grundlagen	3
2.2	Methode der Hierarchischen Clusterung	3
2.3	Anwendung	5
2.4	Klassifikation mithilfe der SVM Methode	8
3	Untersuchung der Zustandsänderungsraten vor und nach Stresseinwirkung	9
4	Untersuchung der Korrelationskoeffizienten vor und nach Stresseinwirkung	12
5	Simulation des Verhaltens einer Dyade mit Markov Modellierung	12
6	Auswertungsautomatisierung	15

1 Gegenstand

In dieser Arbeit werden von Psychologen der Universität Hamburg erhobene Daten ausgewertet. In den Untersuchungen wurden die Blickmuster von Mutter Kind Paaren jeweils für eine stressfreie als auch für eine Situation unter äußerem Stresseinfluss dokumentiert. Die Blicke zu einen bestimmten Zeitpunkt wurden bereits in jeweils vier Zustände für die Mutter und das Kind eingeteilt. Bei den Müttern lag jeweils ein uns nicht bekanntes Krankheitsbild vor. Zur Einteilung der Dyaden in Gruppen wurden Kenngrößen wie Aufenthaltswahrscheinlichkeit, Korrelation und Zustandsraten für unterschiedliche Zeitreihen untersucht. Hierzu wurde unter anderem die Methode der hierarchischen Clustering eingeführt.

2 Dyadenklassifikation anhand der Aufenthaltswahrscheinlichkeit

2.1 Grundlagen

Kombiniert man die jeweils vier möglichen Mutter und Kind Zustände, so erhält man 16 (4×4) mögliche Zustände, die den Gesamtzustand des Systems zu einem bestimmten Zeitpunkt beschreiben.

$$Z(t) \in \{z_1, z_2, \dots, z_{15}, z_{16}\}$$

Anhand einer Zeitreihe lässt sich für jede Dyade eine empirische Aufenthaltswahrscheinlichkeit $P(z_i)$ für den Gesamtzustand bestimmen. Die Aufenthaltswahrscheinlichkeiten werden auf 1 normiert. $\sum_{i=1}^{16} P(z_i) = 1$

Die Aufenthaltswahrscheinlichkeiten einer Zeitreihe können einen Ortsvektor \vec{P} im 16-dimesnionalen P-Raum definieren $P_i = P(z_i)$. Nun wird die Annahme aufgestellt, dass der \vec{P} Vektor eine relevante Aussage über die Dyade trifft.

2.2 Methode der Hierarchischen Clustering

Ziel der hier definierten Methode ist es, die Dyaden in hierarchisch ineinander gestaffelte Gruppen einzuteilen, d.h. eine Gruppe kann jeweils eine oder mehrere Untergruppen enthalten. Um die Komplexität des Ergebnisses zu begrenzen, soll eine Überlappung verschiedener Gruppen vermieden werden.

Als Kriterium für die Ähnlichkeit zweier \vec{P} -Vektoren wurde ihre Nähe im P-Raum gewählt. Der Abstand d zweier Vektoren $\vec{P}(x)$ und $\vec{P}(y)$ wird hier mit der euklidischen Norm definiert.

$$d = \sqrt{\sum_i [P_i(x) - P_i(y)]^2}$$

Nun wird für den \vec{P} -Vektor jeder Dyade eine nach dem Abstand sortierte Liste der Nachbarn aufgestellt. Jetzt werden die Dyaden nach folgendem Verfahren in Gruppen

eingeteilt: Zunächst wird eine beliebige Dyade (hier als x bezeichnet) gewählt. Der Abstand dieser Dyade zu ihrem nächsten Nachbarn (hier als y bezeichnet) wird als Maßstab m zur Gruppeneinteilung verwendet.

$$m = d(x, y)$$

Die hier gebildete Gruppe besteht zunächst nur aus den Dyaden x und y . Nun wird überprüft, ob die Dyade y Nachbarn mit einem geringeren Abstand als m aufweist. Sollte dies der Fall sein, so werden diese Nachbarn mit in die Gruppe aufgenommen. Auch die neu aufgenommenen Gruppenmitglieder werden auf Nachbarn mit einem Abstand kleiner m überprüft. Kann kein Nachbar eines der Gruppenmitglieder mit einem Abstand kleiner m ermittelt werden, so wird die Gruppe geschlossen. Dieses Procedere muss für jede Dyade durchgeführt werden.

In Abbildung 1 ist die hierarchische Clusterung (HC) anhand von Punkten im zweidimensionalen Raum veranschaulicht. Die Clusterbildung findet hier für den Punkt 2 statt. Zur vollständigen Einteilung in Cluster muss das Verfahren auch auf die Punkte 1, 3, und 4 angewandt werden. In Abbildung 1.2 wird der charakteristische Abstand für die Clusterbildung anhand von Punkt 2 abgebildet. In den Abbildungen 1.3 und 1.4 gelangen die Punkte 1 und 3 in den Cluster. Punkt 4 befindet sich zu keinem der Clusterpunkte in einem Abstand geringer dem des charakteristischen Abstandes und ist somit nicht Teil des Clusters. Wendet man die hierarchische Clusterung auch auf die Punkte 1,3 und 4 an, so erhält man folgende Cluster:

$$C_1 = \{1, 3\}$$

$$C_2 = \{1, 2, 3\}$$

$$C_3 = \{1, 3\}$$

$$C_4 = \{1, 2, 3, 4\}$$

Da $C_1 \subset C_3$ entspricht, würde man für die die vier Punkte 3 unterschiedliche Cluster erhalten, wobei C_1 eine Untermenge von C_3 ist, welcher wiederum eine Untermenge von C_4 ist.

Die hierarchische Clusterung im höherdimensionalen Raum verläuft analog.

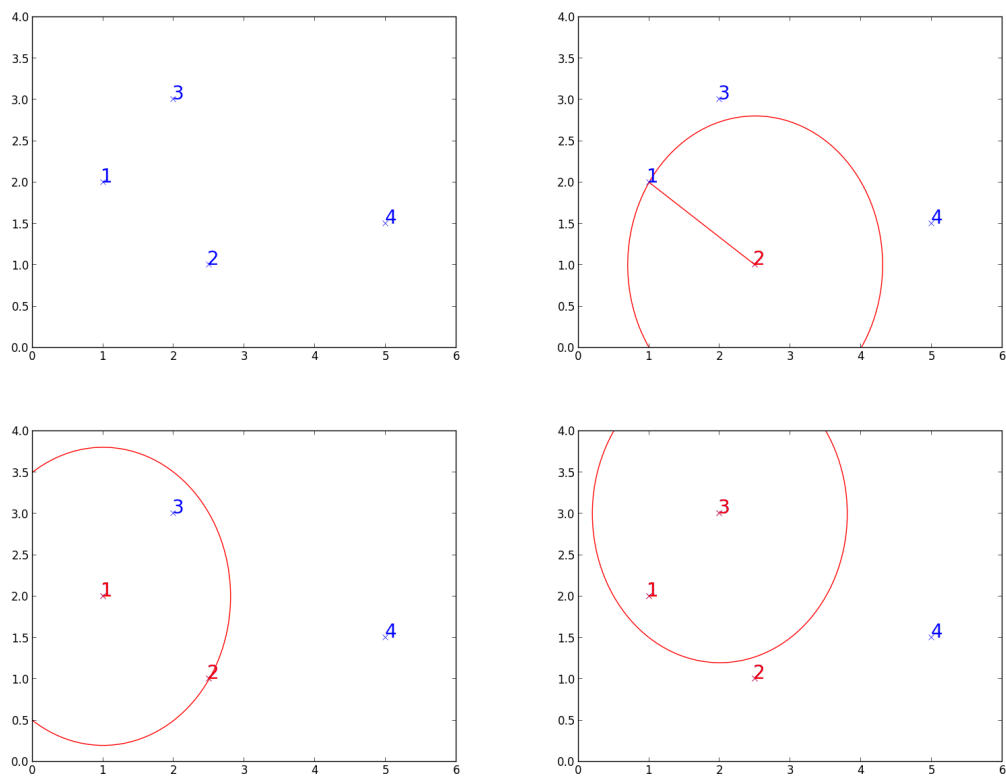


Abbildung 1: Beispiel einer hierarchischen Clusterung für den Punkt 2

2.3 Anwendung

In den folgenden Abbildungen sind die Ergebnisse der hierarchischen Clusterung grafisch dargestellt. Zu unterst sind die großen Obergruppen dargestellt. Oberhalb einer Obergruppe befinden sich jeweils die kleineren Teilgruppen.

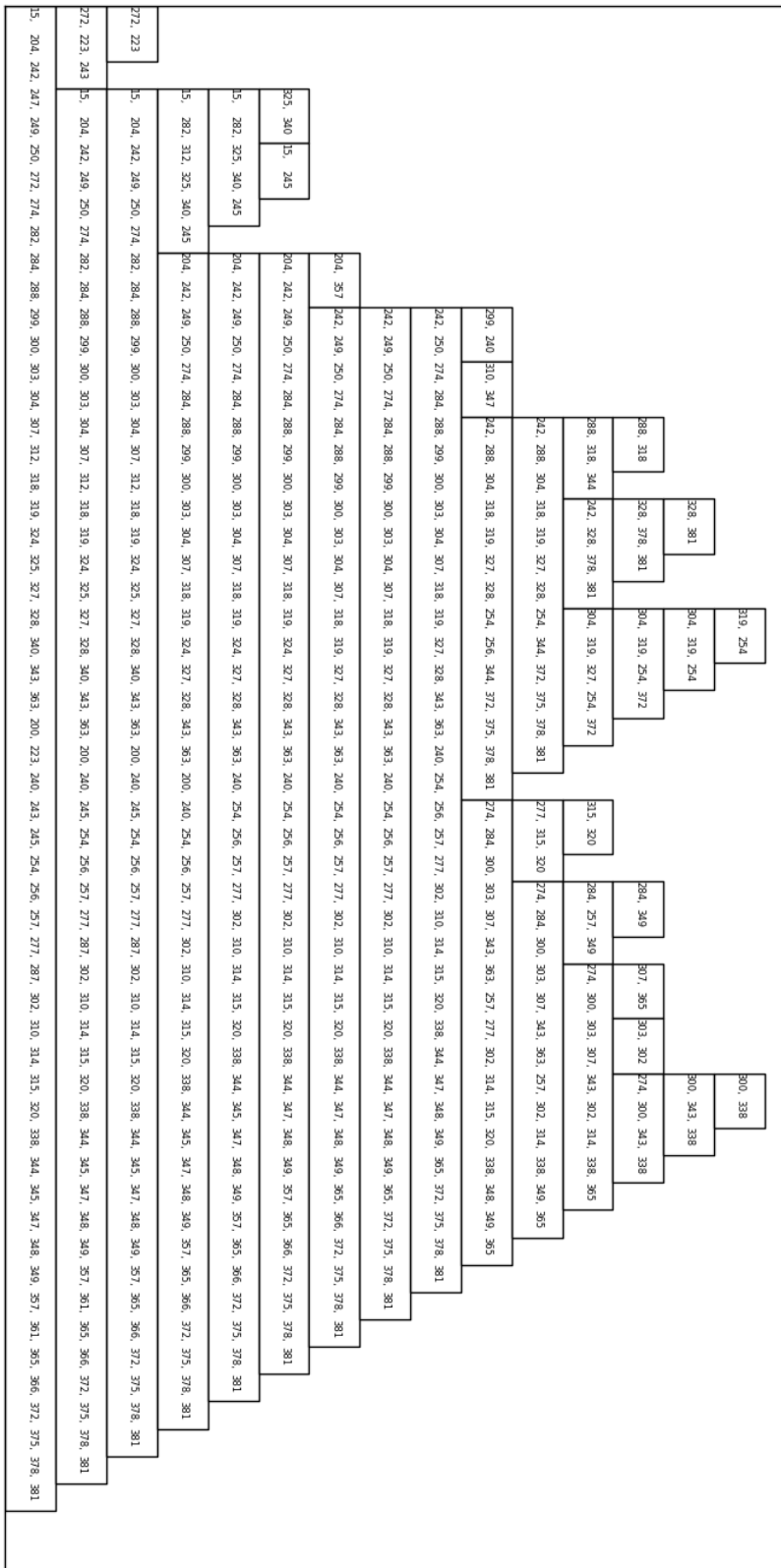


Abbildung 2: Dyaden Gruppen ohne Stress

Support Vector Machines werden für die automatisierte Klassifizierung von Messdaten verwendet. Hierfür werden Testdaten mit bereits vorgenommener Klassifizierung eingelesen. Die Einordnung erfolgt durch Erzeugung von Hyperebenen im Zustandsraum in der Weise, dass der Abstand von Punkten mit unterschiedlichen Klassen maximal wird. Durch Verwendung von entsprechenden Kernelfunktionen im höherdimensionalen Raum als der des Zustandsraums können sogar nichtlineare Trennungsmannigfaltigkeiten erzeugt werden. Für die Umsetzung wurde die Python Bibliothek scikit-learn verwendet. Um sich die Möglichkeiten von SVM vorzustellen, sei hier ein Beispiel dargestellt.



Aus dem obigen Klassifikationsspektrum wurde ein Vorschlag entnommen und die SVM dementsprechend trainiert. Anschließend wurden die Lerndaten von der SVM klassifiziert. Der Klassifikationsvorschlag mit der HC Methode:

Die SVM liefert nach entsprechendem Training:

Man erkennt deutlich, dass die Klassen mit geringer Mächtigkeit teilweise nicht mehr besetzt sind. Dies ließe sich jedoch mit entsprechender Gewichtung der Klassen optimieren. Der Vorteil zur HC Methode besteht vor allem darin, dass sehr große Datenmengen klassifiziert werden können, da der Lernprozess das Resultat durch die eingeführten Hyperebenen vollständig vorgibt.

3 Untersuchung der Zustandsänderungsraten vor und nach Stresseinwirkung

Das Programm `staytime.py` stellt für jede Dyade die Zustandsänderungsraten der Mutter vor und nach Einwirkung von Stress dar. Diese sind in Abbildung 6 übereinander abgetragen. Weiterhin wurde ein Histogramm [Abb. 5] für die absolute Änderung der Änderungsrate erzeugt.

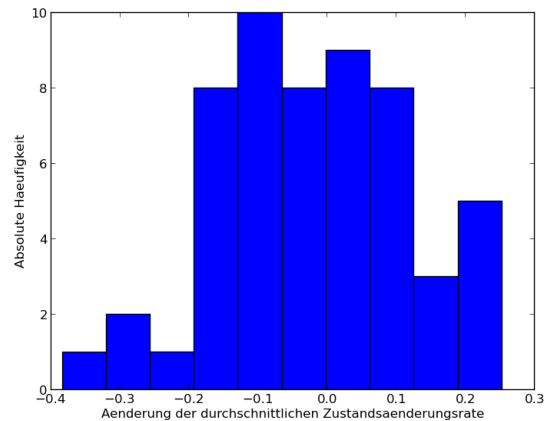


Abbildung 5: Histogramm der Änderung der Zustandsänderungsrate

Erstaunlicherweise hat dies einen Mittelwert bei $-0,02$, sodass die Änderungsrate sich im Mittel nicht merklich ändert. Das erkennt man auch in der oberen Abbildung, da die Anzahl der Messpunkte im oberen Bereich >0 der im unteren Bereich entspricht. Um zu überprüfen, ob die Clusterentstehung zufällig oder begründet ist, werden zufällig generierte Daten hinzugefügt. Diese sind durch Kreuze im Diagramm repräsentiert. Weiterhin wurde die Methode der Hierarchischen Clusterung auf die Messdaten angewandt. Die Anzahl der Klassen kann im Python Skript frei eingestellt werden. Ein Klassifikationsvorschlag ist in Abbildung 7 dargestellt. Die gefundenen Klassen werden auch auf der Konsole ausgegeben.

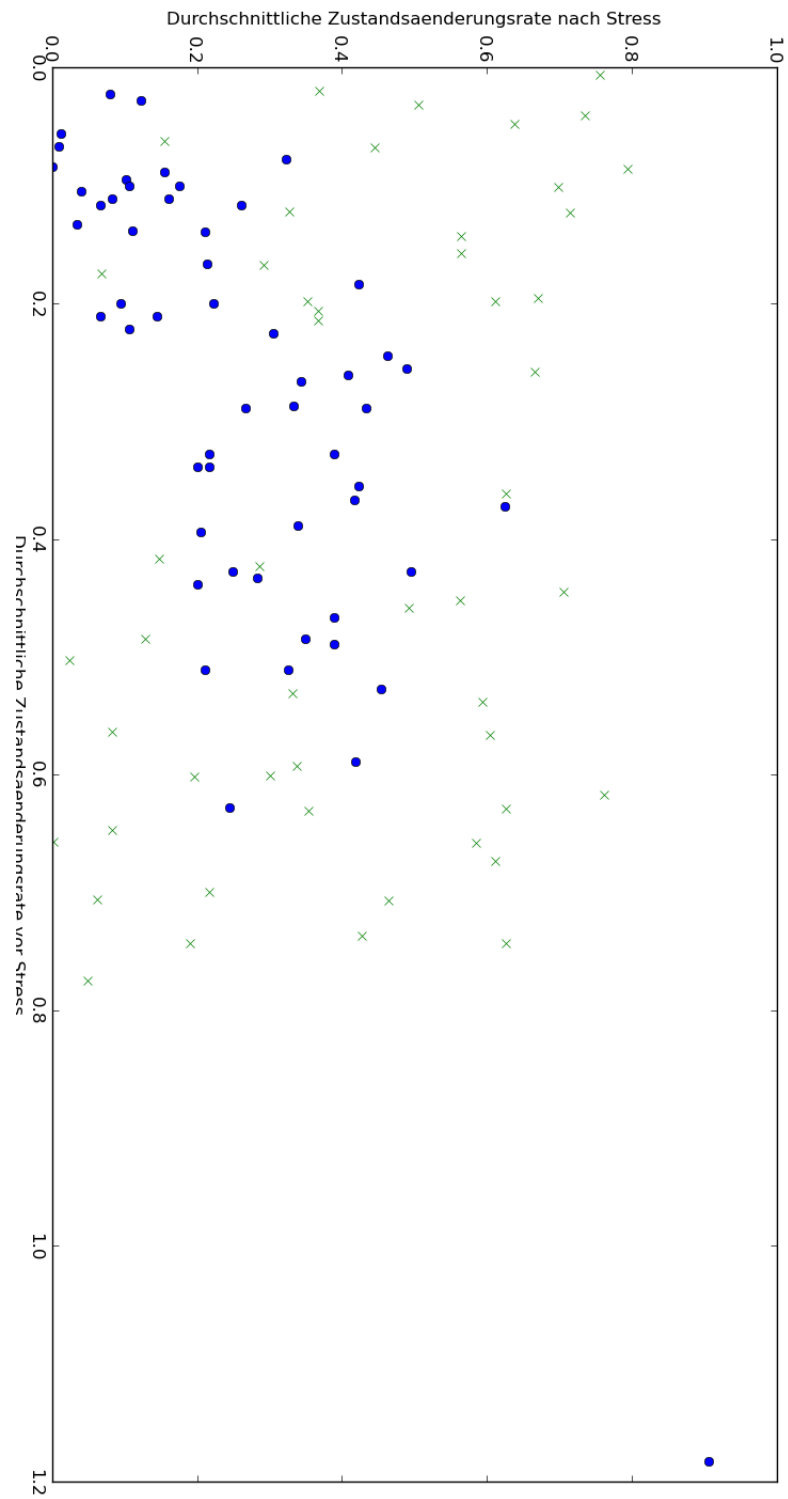


Abbildung 6: Zustandsänderungsrate der Mutter vor und nach Stresseinwirkung

4 Untersuchung der Korrelationskoeffizienten vor und nach Stresseinwirkung

Für jede Dyade wurden die Korrelationskoeffizienten zwischen Mutter und Kind bestimmt und in einem Diagramm abgetragen. Der Korrelationskoeffizient stellt dabei ein Maß für den Synchronisationsgrad einer Dyade dar. Das Ergebnis ist in Abbildung 9 zu sehen. Weiterhin wurde ein Histogramm für die Änderung der Korr.koeff. erstellt (Abb. 8). Der Mittelwert befindet sich bei 0,004. Es zeigt sich damit wieder keine deutliche Veränderung der Messgröße. Ein Klassifikationsvorschlag ist nun die Dyaden in 2 Gruppen einzuteilen, je nach dem Vorzeichen des Korr.koeffizienten.

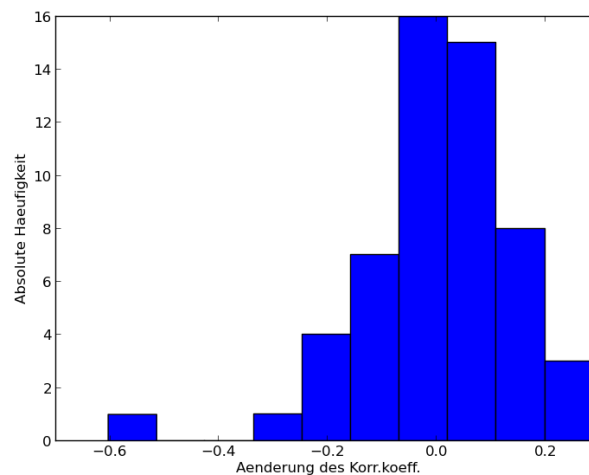


Abbildung 8: Histogramm der Korr.koeff.änderung.

5 Simulation des Verhaltens einer Dyade mit Markov Modellierung

Im Markov Modell wird die Annahme getroffen, dass ein Zustand eines Zufallsprozesses X_t jeweils nur von dem vorherigen Zustand abhängig ist. Mithilfe des Markov Modells lassen sich Dyaden lediglich durch Anfangsverteilung $\mu_i = P(X_0 = s_i)$ und Übergangsmatrix (p_{ij}) beschrieben und weitere Zeitreihen für die Auswertung generieren. Dem Programm wird die Zeitreihe einer Mutter Kind Dyade übergeben. Die Anfangsverteilung der Markov Kette wird aus den relativen Aufenthaltswahrscheinlichkeiten gebildet, die Übergangsmatrix durch Abzählen und Einordnen der Zustandsübergänge. Die Markovkette besteht aus 16 Zuständen $S = \{s_1, \dots, s_{16}\}$. Durch Erzeugung von Zufallszahlen kann damit der Zufallsprozess simuliert werden. Das Ergebnis für die Dyade 10 ist in Abbildung 10 und 11 dargestellt.

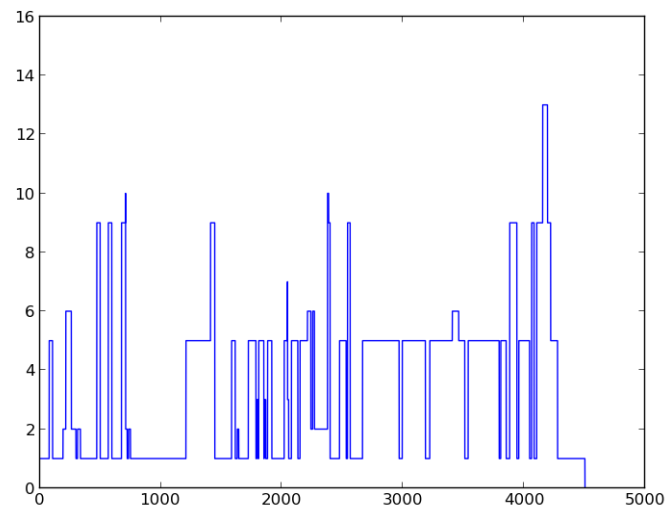


Abbildung 10: Darstellung gemessene Zeitreihe von Dyade 10

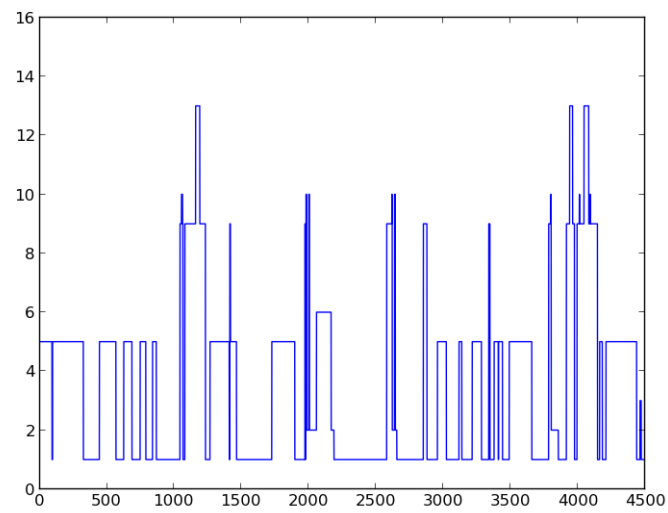


Abbildung 11: Darstellung simulierte Zeitreihe von Dyade 10

6 Auswertungsautomatisierung

Weiterhin wurde ein Tool erstellt, mit dem sich mit Übergabe der Dyadennummer die Zeitreihen von Mutter und Kind sowohl für vor als auch nach Stresseinfluss darstellen lassen. Ein Beispiel ist in Abbildung 12 dargestellt.

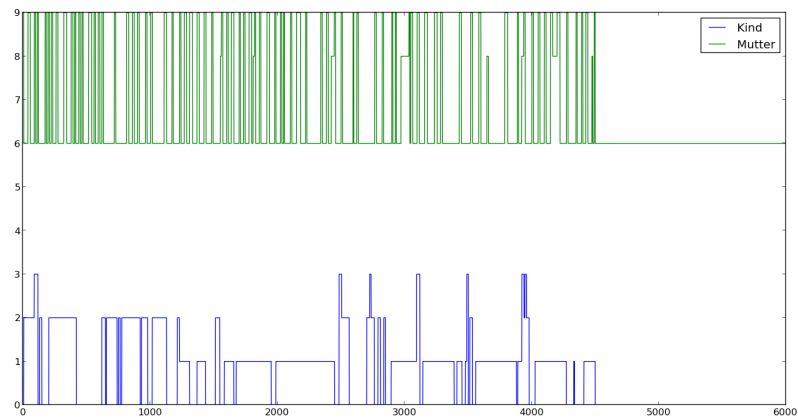


Abbildung 12: Darstellung der Zeitreihe für Dyade 361

Außerdem wird eine Kreuzkorrelations Funktion erstellt und abgebildet (Abb. 13)

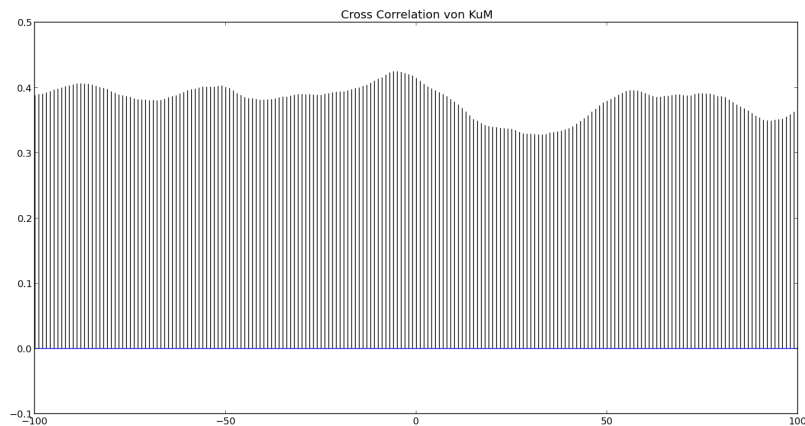


Abbildung 13: Cross Correlation für Dyade 361

Mit dieser lassen sich Aussagen über den Beziehungstyp von Mutter und Kind treffen, d.h. ob das Verhalten der Mutter dem des Kindes folgt oder umgekehrt, indem das Maximum der Funktion in der Nähe des Nullpunkts bestimmt wird.