

Lernen lokaler Zusammenhänge in multivariaten Zeitreihen¹

Frank Höppner²
Fachhochschule Braunschweig/Wolfenbüttel
Salzdahlumer Str. 46/48
38302 Wolfenbüttel

1 Einleitung

Zahlreiche Verfahren zur Datenanalyse oder zum „Lernen aus Daten“ sind in den letzten Jahren entwickelt und erfolgreich angewendet worden (etwa Entscheidungsbäume, Bayes'sche Netze, Neuronale Netze, (Fuzzy) Clustering, Assoziationsregeln, siehe bspw. [13, 5, 8]). Fast alle diese Verfahren werden auf Attributvektoren als Eingabedaten angewendet, wobei jeder Vektor ein bestimmtes Objekt in der Realität charakterisiert. Die Ähnlichkeit zwischen zwei Objekten wird dann häufig durch ein Distanzmaß zwischen den Attributvektoren abgebildet. Zeitliche und/oder räumliche Daten stellen besondere Anforderungen an die Lernverfahren, die im allgemeinen von den Standardverfahren nicht erfüllt werden. Das liegt daran, daß zeitliche und/oder räumliche Objekte a priori in bestimmten Beziehungen zueinander stehen (räumlich benachbart oder zeitlich aufeinanderfolgend), und daß dieses Zusatzwissen oft im Lernprozeß berücksichtigt werden muß, um die gewünschten Resultate erzielen zu können.

Als Beispiel sei die Gruppierung von Ereignissen bezüglich ihres Auftrittsortes genannt, etwa die Lage und Schwere von Verkehrsunfällen. Trägt man die Orte auf ein weißes Blatt Papier ein, so suchen die oben genannten Verfahren vielfach nach Punktwolken, also lokalen, oft kreisförmigen Regionen hoher Ereignishäufigkeit. Sicherlich werden sich auf diese Weise interessante Zonen entdecken lassen (gefährliche Kreuzung), andere jedoch werden erst unter Einbeziehung von Zusatzwissen deutlich. Blenden wir geographische Informationen wie Straßenzüge ein, so kann sich bspw. eine „erhöhte Unfallhäufigkeit entlang der B1“ ergeben. Dieses „Muster“ läßt sich nicht durch Verwendung von allgemeinen Abstandsbegriffen (wie dem euklidischen Abstand) ableiten, sondern erfordert eine spezielle Berücksichtigung der Art der Daten.

Auch bei der Analyse von Zeitreihen spielt das Verständnis von Ähnlichkeit für die Mustersuche eine große Rolle. Abbildung 1 zeigt vier Zeitreihen. Auf die Aufforderung diese Zeitreihen nach Ähnlichkeit zu gruppieren, würde wohl jeder Mensch die Aufteilung in $\{a, b, d\}$ und $\{c\}$ vornehmen. Der Mensch unterscheidet die Kurven

¹Diese Arbeit wurde im Rahmen des DFG-Projektes Kl 648/1 erstellt.

²frank.hoepfner@ieee.org, public.rz.fh-wolfenbuettel.de/~hoepfnef
Präsentiert auf dem 5. Göttinger Symposium Soft Computing, S. 113–125, Juni 2002.

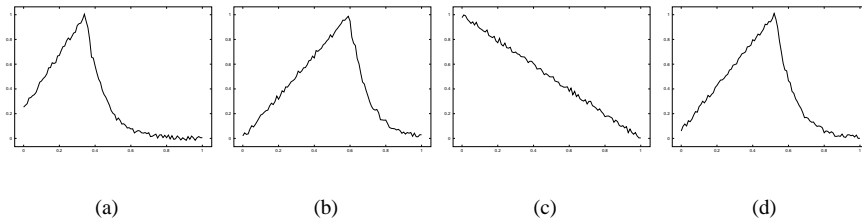


Abbildung 1: Ähnlichkeit von Zeitreihen: Unter Verwendung des euklidischen Abstands ist (a) dichter an (c) als an (b) oder (d).

offenbar nach einigen Grundformen, aus denen die Kurven zusammengesetzt sind, und entscheidet die Ähnlichkeit dann über diese Grundformen: Letztere Gruppe besteht aus einem linear fallenden Segment, während in der ersten Gruppe Kurven enthalten sind, die aus einem linearen Anstieg und einem exponentiellen Abfall zusammengesetzt sind. Dabei spielt es zunächst keine Rolle, daß die Lage des lokalen Maximums und die Steigung des linearen Segments jeweils verschieden ist. Zum einen mag dies aus der Art der menschlichen Wahrnehmung herrühren (Zeitreihen werden nicht als Zahlenfolgen sondern als Bildelemente wahrgenommen), zum anderen trägt es der Tatsache Rechnung, daß sich in der Realität zeitliche Abäufe selten identisch wiederholen, sondern zeitlich leicht verschoben oder etwas verlangsamt/beschleunigt ablaufen. Derartige Effekte führen dann in den beobachteten Variablen zu der zeitlichen Verschiebung von Extrema und zur Streckung/Stauchung von Segmenten, wie es bei den Teilabbildungen (a), (b) und (d) der Fall ist. Würden die Zeitreihen als Wertevektor aufgefaßt und über die euklidische Norm verglichen, wäre (a) dichter an (c) als an (b) oder (d) und die resultierende Einteilung $\{\{a, c\}, \{b, d\}\}$ entspräche nicht der Vorstellung des Menschen.

Natürlich läßt sich der euklidische Abstand dennoch bei der Suche nach Abhängigkeiten in Zeitreihen einsetzen (und das wird auch oft getan, etwa in [6]), aber wie bei dem Beispiel mit räumlichen Daten ist die Menge der erkennbaren Muster dann stark eingeschränkt. Die mangelhafte Kompensation von Dilatations- und Translationseffekten macht eine Extraktion vieler in der Realität auftretender Muster unmöglich. Hinzu kommt, daß zur Beurteilung der Generalisierbarkeit extrahierter Regeln, ihre Interpretierbarkeit entscheidend ist. Wenn ein Mensch die gefundenen Regeln mit seinem Wissen hinterfragen und schließlich bestätigen kann, nur dann kann man wirklich von einem Beitrag zur „Wissensentdeckung“ sprechen. Zur Beurteilung der Regeln muß der Mensch mit der gewählten Darstellung jedoch vertraut sein, sie sollte so nah wie möglich am „Musterbegriff“ des Menschen liegen.

In diesem Papier wird ein Verfahren zum Lernen von lokalen Abhängigkeiten in Zeitreihen skizziert (für die Details wird auf weitere Literatur verwiesen). Mit der Einschränkung auf lokale Phänomene werden die Einschränkungen klassischer Verfahren überwunden, die bei nicht-stationären oder chaotischen Zeitreihen scheitern, weil sie ein globales Modell für die Zeitreihen als Ganzes zu entwickeln versuchen.

Im Vordergrund steht außerdem eine Modellierung zeitlicher Phänomene, die der Wahrnehmung eines Menschen sehr nahe kommt (Abschnitt 2). Dazu werden die Zeitreihen in eine andere Repräsentation überführt (Abschnitt 3), nämlich in eine Intervall-Sequenz. Aus dieser Sequenz werden dann häufige Muster (Abschnitt 4) und Regeln (Abschnitt 5) extrahiert. Abschnitt 6 gibt einen Eindruck über das Laufzeitverhalten und zeigt einige Beispiele.

2 Lokale Muster in multivariaten Zeitreihen

Die am häufigsten auftretende Definition lokaler Muster in Zeitreihen ist die Definition eines Zeitfensters und eine Transformation des Fensterinhalts in einen Attributvektor (um dann eine der eingangs erwähnten statischen Verfahren anwenden zu können). Der Attributvektor kann durch die direkte Übernahme der (möglicherweise transformierten) Zeitreihenwerte im Fenster gebildet werden [6] oder auch die ersten Fourier-Koeffizienten des Fensterinhalts darstellen [7]. In diesen Fällen erfolgt ein Vergleich der extrahierten Attributvektoren (mit dem Ziel der Identifikation von ähnlichen Verläufen) dann mit Hilfe des euklidischen Abstands.

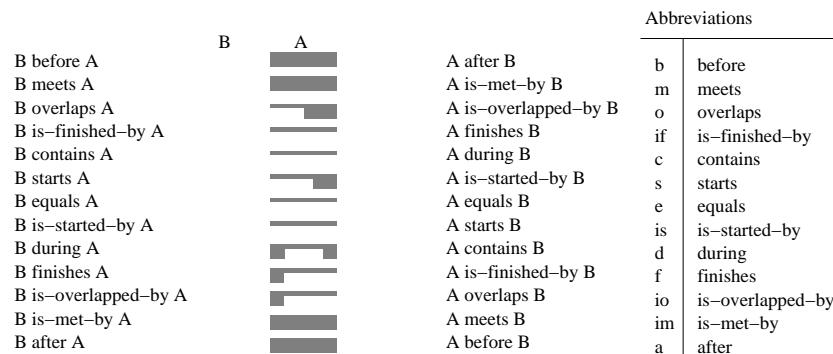


Abbildung 2: Allen’s qualitative Intervall-Beziehungen.

Hier wollen wir im Gegensatz dazu unter einem Muster eine Menge von Eigenschaften in den Zeitreihen verstehen, die über ein bestimmtes Zeitintervall gelten, zusammen mit den zeitlichen Beziehungen der Intervalle untereinander. Derartige Eigenschaften sind qualitativer Natur (zum Beispiel „linear ansteigend“, „konvex fallend“), können aber mit quantitativen Eigenschaften gepaart werden (zum Beispiel „für mindestens 4 Stunden linear ansteigend“, „konvex fallend mit einem Gradienten kleiner als -1“). Die qualitative Lage zweier beliebiger Intervalle in solch einem Muster läßt sich durch eine von 13 möglichen Intervallbeziehungen angeben (vgl. Abbildung 2, [2]). Unterliegen zwei Intervalle A und B Verschiebungs- oder Streckungseffekten, so bleibt ein Muster „Intervall A überlappt Intervall B“ bis zu einem gewissen Grad erhalten. Ein Muster aus k Intervallen ist durch die k Eigenschaften (Label) der Intervalle und eine $k \times k$ -Matrix aus Intervallbeziehungen eindeutig definiert.

Wie läßt sich die Häufigkeit des Auftretens eines bestimmten Musters charakterisieren? Da die tatsächliche *Anzahl* der Muster-Vorkommen nicht zu einer intuitiven Definition des Häufigkeitsbegriffes führt [9], wird der Support eines Musters P ($\text{supp}(P)$) wie folgt definiert: Ein Zeitfenster einer bestimmten Breite wird über die Sequenz geschoben. Die Supportmenge S eines Musters ist die Menge der Fensterpositionen t , an denen das Muster im Fenster sichtbar ist. Der Support eines Musters ist dann das Integral $\int_t \chi_S(t) dt$ (wobei χ_S die charakteristische Funktion von S ist, also $\chi_S(t) = 1 \Leftrightarrow t \in S, 0$ sonst), also die Gesamtzeit, in der das Muster sichtbar ist. Dividiert man den Support durch die Gesamtlänge der Sequenz, so erhält man die Wahrscheinlichkeit für die Beobachtung des Musters bei einer zufällig ausgewählten Fensterposition.

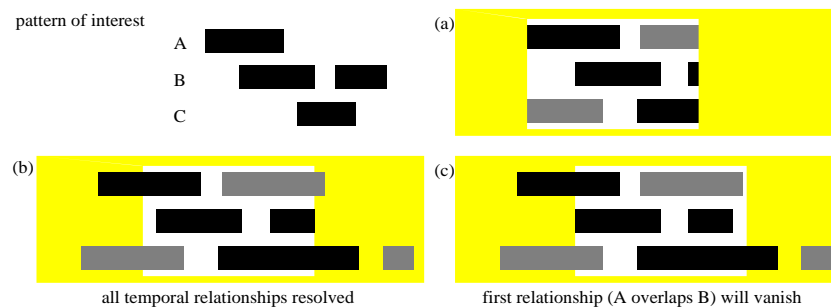


Abbildung 3: Sichtbarkeit von Intervall-Mustern.

Abbildung 3(a) zeigt ein Beispielmuster, bestehend aus 4 Intervallen, Teilbild (a) zeigt den Teil einer Intervallsequenz, der im aktuellen Fenster sichtbar ist. Läßt sich das Muster im Fenster beobachten? Augenscheinlich ja, die ersten drei Intervalle sind bereits sichtbar, das vierte Intervall ist soeben ins Fenster gekommen. In den Teilbildern (b) und (c) sind auch die verdeckten Teile der Sequenz sichtbar und es wird deutlich, daß es sich in (a) nicht notwendigerweise um eine Instanz des Musters handeln muß. Erst wenn alle Intervallbeziehungen aufgelöst wurden (ab Position (b)), kann entschieden werden, ob es sich bei den letzten beiden Intervallen um eine Überlappung oder Inklusion handelt. Entsprechend kann das Muster nicht länger beobachtet werden, sobald die Beziehung zwischen den ersten beiden Intervallen nicht länger beobachtet werden kann (ab Position (c)). Die Zeit, die zwischen (c) und (d) verstreicht, addiert sich gemäß Definition zum Support des Musters.

Bei richtiger Umsetzung der Abstraktion von Zeitreihen in Intervallsequenzen, hat dieser Musterbegriff einige Vorzüge:

- Verglichen mit der Vektordarstellung und dem euklidischen Abstand korrespondiert ein Muster besser mit der menschlichen Wahrnehmung, d.h., von Menschen als ähnlich wahrgenommene Teilsequenzen werden jetzt auf dasselbe Muster abgebildet (Translations- und Dilatationseffekte werden bis zu einem gewissen Grad kompensiert).
- Nachdem eine Zeitreihe in eine Intervallsequenz transformiert wurde, ist im

Vergleich mit anderen Techniken wie Dynamic Time Warping (DTW) [4] die Mustererkennung einfacher und effizienter.

- Zeitliche Beziehungen zwischen mehreren Zeitreihen oder zu singulären Ereignissen können ebenfalls modelliert werden (wohingegen viele andere Ähnlichkeitsbegriffe nur den direkten Vergleich zweier Zeitreihen abdecken).
- Wie auch in anderen Ansätzen wird ein Zeitfenster zur Begrenzung der zeitlichen Ausdehnung von Mustern eingesetzt, aber hier kann der Vergleich auf einem Teil des Fensterinhalts durchgeführt werden, während bei den meisten anderen Verfahren jeweils der gesamte Fensterinhalt passen muß. Daher ist die Wahl der Fenstergröße hier weniger kritisch.
- Und schließlich reduziert die Repräsentation der Zeitreihen als Intervallsequenz das Datenvolumen und damit den Aufwand für folgende Analyseschritte.

3 Abstraktion von Zeitreihen zu Intervallsequenzen

Welche Eigenschaften der Zeitreihen sollen bei der Definition der Intervallsequenz herangezogen werden? Grundsätzlich sind hier keine Vorgaben gemacht, allerdings stellt sich die Unterscheidung nach Vorzeichen von Steigung und Krümmung als besonders vorteilhaft heraus, weil wir bereits in der Schule zur Bestimmung des Graphen einer Funktion gelehrt wurden, die Nullstellen der ersten Ableitungen für die Lage der Extrema und Wendepunkte heranzuziehen. Aus den Nullstellen ergeben sich gleichzeitig die Intervallgrenzen steigender und fallender, konkaver und konvexer Zeitreihen-Segmente.

Liegt ein abgetastetes Signal vor, so ist dies meistens verrauscht und die naive Bestimmung der ersten Ableitungen (Differenzbildung) führt zu einer Fülle von Nullstellen und Segmenten. Segmente, die auf diese Weise aus dem Signal abstrahiert werden, entsprechen sicherlich nicht denen, die ein Mensch im Signal identifiziert hätte. Die im letzten Abschnitt als Vorteil aufgeführte Nähe zur menschlichen Auffassung von Mustern setzt aber voraus, daß die abstrahierten Segmente ebenfalls der menschlichen Wahrnehmung entsprechen. Es gibt viele Verfahren zur Rauschunterdrückung, so kann das Signal bspw. geglättet oder lokal approximiert werden. Leider erfordern die meisten der dazu verwendeten Verfahren die Angabe einiger Parameter, die vorab bestimmt werden müssen. Handelt es sich bei den Daten um einen kurzen Zeitabschnitt mit identischen Bedingungen, so mag sich ein geeigneter Parametersatz finden lassen. Wenn die Historiendaten, aus denen wir lernen wollen, aber über mehrere Jahre ermittelt wurden, dann können die Messungen Verschleißerscheinungen unterliegen (schleichende Zunahme des Rauschens) oder die Meßgeräte könnten gar erneuert worden sein (plötzliche Abnahme des Rauschens). In typischen KDD-Anwendungen (Knowledge Discovery in Databases) ist das Datenvolumen so groß, daß eine manuelle Sichtung der Daten nicht möglich ist. Die Vorab-Festlegung eines einzigen Parametersatzes ist daher in solchen Fällen äußerst

fragwürdig, zumal die Qualität der Ergebnisse deutlich von der korrekten Parameterwahl abhängt (vgl. [10]).

Eine Mehr-Skalen-Analyse des Signals (zum Beispiel mit Hilfe der Wavelet-Analyse) kann helfen, diese Probleme zu verhindern, indem das Signal nicht nur in *einer* sondern in einer *Vielzahl* von Glättungsstufen analysiert wird. Die Glättung eines Signals führt zur Auslöschung von einem benachbarten Nullstellen-Paar (Minimum und Maximum), und aus den drei Segmenten um die beiden Nullstellen ergibt sich ein einzelnes, neues Segment. Aus dem Vergleich der unterschiedlich stark geglätteten Signale läßt sich ablesen, wie lange ein bestimmtes Segment der fortschreitenden Glättung „standhält“, bevor es verschwindet. Wenn wir Segmente heranziehen, die bereits nach einem geringen Maß an Glättung verschwinden, so ist die extrahierte Intervallsequenz sehr stark von der durchgeführten Glättung, d.h. den Filterkoeffizienten, abhängig: Eine leichte Abweichung bei den Koeffizienten oder im Rauschen der Kurve und schon würden wir andere Segmente erhalten – die Segmente der Abstraktion und damit die folgenden Schritte wären nicht robust. Sinnvoll erscheint daher die Forderung nach einem Mindestmaß an Glättung, das ein Segment überstehen muß, um später auch unter leicht veränderten Rahmenbedingungen rekonstruiert werden zu können.

Die Analyse des Signals in mehreren Glättungsstufen ist hilfreich bei der Identifizierung von hervorstechenden Merkmalen der Kurve, aber hinderlich beim Vergleich mehrerer Segmente, die aus unterschiedlichen Glättungsstufen gewonnen wurden. Die Glättung von Kurven führt generell zur Verschiebung von Extrema und Wendepunkten, so daß man unterschiedliche Intervallängen bekommt, wenn man in einer Intervallsequenz Segmente aus verschiedenen Glättungsstufen extrahiert. Eine Aufdeckung von Zusammenhängen in Zeitreihen, bei denen Intervallängen eine Rolle spielen, wird also erschwert, weil die tatsächlichen Segmentpositionen verfälscht sind. Ähnlich verhält es sich mit der Extraktion von Gradienteninformationen, da unterschiedlich starke Glättung auch eine unterschiedlich starke Abflachung der Kurve zur Folge hat.

Als Lösung dieser Probleme bei der Abstraktion von Zeitreihen bietet sich das Verfahren aus [16, 3] an, das bisher in der KDD-Literatur weitestgehend unbekannt scheint. Speziell im Vergleich mit den zahlreichen (heuristischen Single-Scale-) Verfahren, die oft in der Literatur herangezogen werden (vgl. [10]), zeichnet sich dieses Verfahren besonders aus, weil es

- den Effekt der Verschiebung von Nullstellen in den Ableitungen bei fortschreitender Glättung [16]
- und den Effekt der Abflachung der Kurve bei fortschreitender Glättung kompensieren kann [3]
- und Aussagen über die Robustheit von Intervallen abgeleitet werden, die zur Selektion von Intervallen herangezogen werden können.

4 Aufzählung häufiger Muster

Nachdem der Musterbegriff definiert ist und die Zeitreihen sinnvoll in eine Intervall-Sequenz überführt worden sind, ist nun die Häufigkeit bestimmter Muster in einer gegebenen Intervall-Sequenz zu bestimmen. Da wir vorher nicht wissen, welche Muster interessant sind, müssen wir alle Muster aufzählen, um nicht einen interessanten Zusammenhang zu verpassen und die Vollständigkeit zu erhalten. Die Anzahl der möglichen Muster steigt exponentiell mit der Größe der Muster (Anzahl der Intervalle). Da wir die Anzahl der Intervalle in einem Muster nicht a priori beschränken wollen, erscheint die Aufzählung potentiell unendlich vieler Muster hoffnungslos. Da die Intervalle in unserer Sequenz aber nicht beliebig klein sind und ein Muster innerhalb eines begrenzten Zeitfensters sichtbar sein muß, ist die Anzahl der sichtbaren Muster endlich. Zur Aufzählung dieser Muster werden Techniken aus dem Bereich der Assoziationsregeln [1] modifiziert.

Im Hinblick auf Wissen in Form von Regeln, das wir im nächsten Abschnitt aus den Mustern erzeugen wollen, macht es Sinn, nur jene Muster zu betrachten, die eine bestimmte Grundhäufigkeit aufweisen. Denn einer Regel, die nur aus ein oder zwei Beobachtungen abgeleitet ist, wird generell wenig Vertrauen geschenkt werden. Wir beschränken uns daher auf *häufige Muster*, deren Support einen minimalen Support überschreitet. (Diese Mindesthäufigkeit mag auch nur 1% sein, wenngleich man dann im normalen Sprachgebrauch nicht unbedingt von „häufigen Mustern“ sprechen würde.)

Die häufigen Muster werden schrittweise aufgezählt, beginnend bei den 1-Mustern aus nur einem Intervall, dann alle 2-Muster mit zwei Intervallen usw. Beim k -ten Schritt, der alle häufigen k -Muster bestimmt, könnte in einem naiven Ansatz jedes mögliche k -Muster erzeugt und sein Support bestimmt werden. Sobald für ein Wert von k keine häufigen Muster mehr gefunden wurden, kann der Algorithmus beendet werden. Das liegt daran, daß für den Support zweier Muster A und B , wobei A Teilmuster von B ist, stets $\text{supp}(A) \geq \text{supp}(B)$ gilt (wann immer wir das größere Muster beobachten können, können wir auch alle Teilmuster beobachten). Daher kann es kein häufiges m -Muster geben, wenn wir schon keine häufigen k -Muster, $k < m$, mehr gefunden haben.

Da die Anzahl der k -Muster mit k jedoch exponentiell steigt, wäre das Verfahren für zunehmendes k immer ineffektiver. (Dabei erreicht die Anzahl der häufigen Muster in der Praxis oft bei $4 \leq k \leq 6$ ein Maximum und nimmt dann stark ab.) Praktikabel wird das Aufzählungsverfahren erst dadurch, daß sich die Menge der möglichen Kandidaten für häufige k -Muster – bezogen auf die Menge aller möglichen k -Muster – deutlich einschränken läßt. Dazu werden drei verschiedene Pruning-Verfahren eingesetzt. Zunächst einmal läßt sich die Support-Beziehung für Teilmuster aus dem vorigen Absatz noch weiter ausschöpfen. Bevor ein k -Muster B als Kandidat angesehen wird, überprüfen wir, ob jedes der k Teilmuster A_i von B auch häufig ist, wobei sich A_i aus B ergibt, indem das i -te Intervall gestrichen wird. Sollte ein A_i nicht häufig sein, so kann auch B nicht häufig sein, und wir brauchen seinen Support gar nicht erst zu ermitteln. Zweitens läßt sich das Transitivitätsgesetz für Intervallbeziehungen ausnutzen [2], um inkonsistente Muster auszuschließen.

Schließlich lassen sich die Supportmengen der Teilmuster A_i (mengentheoretisch) schneiden, um eine obere Grenze für den Support von B zu erhalten. Sollte diese obere Grenze unterhalb des minimalen Supports sein, kann B nicht häufig sein und wird kein Kandidat.

Für die Menge der Kandidaten ist dann der Support (ebenfalls möglichst effektiv) zu bestimmen, um häufige von den nicht-häufigen Kandidaten trennen zu können. Der Test, ob ein Muster in einen Fensterinhalt eingebettet werden kann, läßt sich nicht in konstanter Zeit durchführen, je nach Komplexität des Musters und Fensterinhalts kann hier Backtracking in unterschiedlichem Maße erforderlich werden. Daher spielt es hier ebenfalls eine große Rolle, die Anzahl dieser Tests möglichst minimal zu halten, um die (immer noch) große Menge der Kandidaten effizient abarbeiten zu können. Details zur Aufzählung der häufigen Muster und zur Bestimmung des Supports finden sich in [12].

5 Interessante Regeln

Im letzten Schritt haben wir alle häufigen Intervall-Muster bestimmt, aus denen es nun gilt, „interessante Muster“ herauszusuchen. Üblicherweise ist die Menge der häufigen Muster sehr groß. Ein einziges interessantes, häufiges k -Muster hat $2^k - 1$ häufige Teilmuster (unterschiedlicher Größe). Viele zufällige Phänomene, die aber die Mindesthäufigkeit erreichen, sorgen für eine weitere Vergrößerung der Menge. Aus jedem Muster B und einem Teilmuster A läßt sich eine Regel $A \rightarrow B$ erzeugen: Wenn wir A im Fenster beobachten können, dann mit der Wahrscheinlichkeit $\text{supp}(B)/\text{supp}(A)$ (Konfidenz) auch das Muster B . Offenbar ist die Menge der möglichen Regeln noch größer als die Menge der häufigen Muster. Die ganze Menge der Regeln kann aufgrund ihrer Größe nicht manuell gesichtet werden, wie können wir aus dieser Regelmenge die interessantesten zur Sichtung auswählen?

Wie bei den Mustern, bei denen ein minimaler Support eingeführt wurde, ist man an Regeln mit geringen Konfidenzwerten nicht besonders interessiert. Traditionell wird auch hier ein Schwellwert für eine minimale Konfidenz eingeführt: Bei hohen Konfidenzwerten ist der Schluß von der Prämisse zur Konklusion besonders sicher. Wie zuvor die Menge der Muster reduzieren wir dadurch nun die Menge der Regeln, allerdings ist der Konfidenzwert einer Regel allein nicht besonders aussagekräftig: Ein vergleichsweise geringer Konfidenzwert von 0.4 kann dennoch sehr interessant sein, wenn das Muster B nur sehr selten auftritt ($\text{supp}(B)$ klein), eine Regel mit Konfidenzwert von 1.0 kann hingegen uninteressant sein, wenn das Muster B sowieso fast immer sichtbar ist ($\text{supp}(B)$ hoch). Besser geeignet ist daher das informations-theoretische J-Maß [14], das beurteilt, wieviel Informationsgewinn die Beobachtung der Prämisse für die Beobachtung der Konklusion im Schnitt bedeutet. Dieses Maß zur Regelbewertung ist besser geeignet als die Konfidenz, weil nicht nur die a posteriori sondern auch die a priori Wahrscheinlichkeit für das Auftreten der Konklusion berücksichtigt wird.

Ein weiteres Problem ist die Tatsache, daß Muster mit vielen Intervallen stets seltener zu beobachten sind, als Muster mit wenigen Intervallen. Da die Muster in der

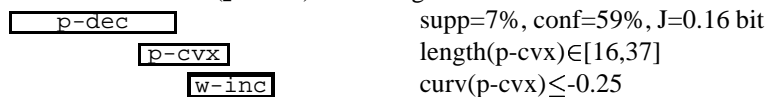
Konklusion notwendig mehr Intervalle besitzen, wird auch die numerische Bewertung der Regeln davon beeinflusst, so fallen etwa Konfidenzwerte dadurch deutlich geringer aus und oft ist ein Konfidenzwert von 1.0 überhaupt nicht zu erreichen. Da dieser Effekt allein auf der Beobachtbarkeit von Muster beruht, aber keine Aussage über die Güte von Regeln zuläßt, ist der Einfluß auf Größen wie dem Konfidenzwert unerwünscht. In [11] wird beschrieben, wie sich dieser Effekt kompensieren läßt. Oftmals sind lokale Zusammenhänge nicht rein qualitativer Natur, zum Beispiel könnte ein Anstieg in Kurve 1 sich in Kurve 2 nur dann bemerkbar machen, wenn der Anstieg in Kurve 1 „deutlich“ ist, d.h. die Steigung größer als ein bestimmter Schwellwert ist. Die automatische Bestimmung solcher Schwellwerte ist ebenfalls in [11] beschrieben. Diese Schwellwerte können dann dazu benutzt werden, um die qualitativen Eigenschaften (wie „steigend“) um quantitative Eigenschaften zu erweitern, also neue Intervalle aus den Zeitreihen zu extrahieren (wie „steigend mit Steigung > 1 “). Der ganze Prozeß wird dann mit dem „erweiterten Vokabular“ wiederholt, um weitere Zusammenhänge aufzudecken.

6 Ergebnisse

Zahlreiche Tests des Verfahrens bestätigen das theoretisch erwartete Verhalten der Implementation: Die Aufzählung der Muster ist linear in the Anzahl der Intervalle, skaliert also gut mit der Länge der Sequenz. Der Algorithmus ist log-linear in der Anzahl der häufigen Muster, die ihrerseits abhängig von der Sequenz selbst und der Anzahl der verschiedenen Label sind. Mit steigender Anzahl von Variablen/Eigenschaften steigt die Anzahl der häufigen Muster rasch überproportional, es sollten daher immer nur Abhängigkeiten zwischen einigen wenigen ausgewählten Variablen gesucht werden. Die Spezialisierung von Regeln ist im schlimmsten Fall quadratisch in der Anzahl der Muster-Instanzen einer Regel. Da die Spezialisierung als nachgelagerter Schritt jedoch auf eine kleiner Menge von Regeln angewendet wird, erscheint die höhere Zeitkomplexität nicht weiter problematisch.

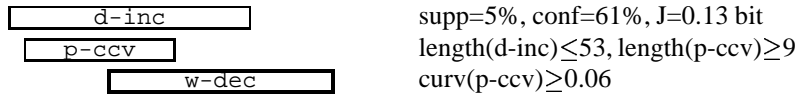
Neben künstlichen Daten wurden auch Experimente mit über 7.5 Jahre stündlich gemessenen Luftdruck-, Windgeschwindigkeits- und Windrichtungs-Daten durchgeführt. Die Zeitreihen-Abstraktion führte zu einer erheblichen Reduktion des Datenvolumens auf etwa 10000 Intervalle. Bei einer Fensterbreite von 72 Stunden und einer minimalen Sichtbarkeit von 650 Stunden für häufige Muster brauchte der Algorithmus bei 8 Labeln für die 3 Variablen nur eine Minute zur Aufzählung der über 19000 häufigen Muster mit bis zu 7 Intervallen.

Ein Beispiel für eine gefundene Regel ist der Schluß von fallendem (p-dec), konvexen Luftdruck (p-cvx) zu ansteigender Windstärke:



Auf das fallende Segment folgt immer ein steigendes, woraus sich ein lokales Minimum ableiten läßt. Die Bedingung an die Krümmung deutet an, daß es sich

um ein ausgeprägtes, steiles Minimum handeln muß, damit die Regel verlässlich ist. Ein anderes Beispiel für eine spezialisierte Regel ist der Zusammenhang zwischen drehender Windrichtung (positive Drehrichtung, *d-inc*) und konkaver Luftdruckkurve (*p-ccv*), der auf eine abnehmende Windgeschwindigkeit hindeutet (*w-dec*):



Diese Beispiele zeigen, daß die erzeugten Regeln leicht vom Menschen interpretierbar sind. Ähnliche Schwellwerte in mehreren Regeln regen zur Einführung von neuen Labeln an. So korrespondieren Schwellwerte, die in einigen Regeln bezüglich der Steigung der Luftdruckkurve gefunden wurden, mit Werten, die in der Literatur angegeben sind (etwa in [15]).

7 Zusammenfassung und Ausblick

In diesem Papier wurde ein Ansatz zur Erkennung von lokalen Zusammenhängen in zeitlichen Daten vorgestellt. Gegenüber anderen Ansätzen (einschließlich klassischen) ist der hier gewählte Musterbegriff so gewählt, daß er der menschlichen Auffassung eines Musters in multivariaten Zeitreihen sehr nahe kommt. Dadurch wird das Verständnis und die Verifikation durch den Menschen vereinfacht, aber auch seine Intuition angesprochen. Der Wissensentdeckungs-Prozeß kann von der Erfahrung des Menschen bei Schlußfolgerungen und Hypothesen-Aufstellung profitieren, weil die gefundenen Zusammenhänge nicht erst vom Menschen transformiert werden müssen, sondern direkt interpretiert werden können. Der gewählte Musterbegriff ist darüber hinaus aber auch aus fachlicher Sicht vorteilhaft, weil er in der Realität auftretende Phänomene wie Dilatation und Translation berücksichtigt. Was die algorithmische Seite angeht, so lassen sich auch längere Zeitreihen in akzeptabler Zeit verarbeiten (Laufzeit linear in der Länge der Intervallsequenz).

Ein weiterer wichtiger Punkt ist die Mehrdeutigkeit bei der Interpretation von Zeitreihen. Abhängig vom Kontext kann ein kleines Zeitintervall als fallend interpretiert oder einem längeren, steigenden Segment zugeordnet werden. Das fallende Zwischenstück kann (nicht-Gauß'sches) Rauschen als Ursache haben, mag aber auch einen Defekt oder ein wichtiges Phänomen andeuten. Ohne Hintergrundwissen läßt sich eine eindeutige Zuordnung nicht treffen, daher muß dieser Mehrdeutigkeit im Prozeß Rechnung getragen werden; die aufgedeckten Zusammenhänge geben dann Aufschluß über die (Un-) Wichtigkeit dieser Unterscheidung. Eine entsprechende Erweiterung des bisherigen Ansatzes ist ohne weiteres möglich und Bestandteil zukünftiger Arbeiten.

Literatur

- [1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo. Fast discovery of association rules. In [8], chapter 12, pages 307–328. MIT Press, 1996.
- [2] J. F. Allen. Maintaining knowledge about temporal intervals. *Comm. ACM*, 26(11):832–843, 1983.
- [3] B. R. Bakshi and G. Stephanopoulos. Reasoning in time: Modelling, analysis, and pattern recognition of temporal process trends. In *Advances in Chemical Engineering*, volume 22, pages 485–548. Academic Press, Inc., 1995.
- [4] D. J. Berndt and J. Clifford. Finding patterns in time series: A dynamic programming approach. In [8], chapter 9, pages 229–248. MIT Press, 1996.
- [5] M. Berthold and D. J. Hand, editors. *Intelligent Data Analysis*. Springer, 1999.
- [6] G. Das, K.-I. Lin, H. Mannila, G. Renganathan, and P. Smyth. Rule discovery from time series. In *Proc. of the 4th Int. Conf. on Knowl. Discovery and Data Mining*, pages 16–22. AAAI Press, 1998.
- [7] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos. Fast subsequence matching in time-series databases. In *Proc. of ACM SIGMOD Int. Conf. on Data Management*, May 1994.
- [8] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining*. MIT Press, 1996.
- [9] F. Höppner. Discovery of temporal patterns – learning rules about the qualitative behaviour of time series. In *Proc. of the 5th Europ. Conf. on Principles of Data Mining and Knowl. Discovery*, number 2168 in LNAI, pages 192–203, Freiburg, Germany, Sept. 2001. Springer.
- [10] F. Höppner. Time series abstraction methods – a survey. In K. Morik, editor, *GI Workshop on Knowl. Discovery in Databases*, Dortmund, Germany, Sept. 2002.
- [11] F. Höppner and F. Klawonn. Finding informative rules in interval sequences. In *Proc. of the 4th Int. Symp. on Intelligent Data Analysis*, volume 2189 of LNCS, pages 123–132, Lissabon, Portugal, Sept. 2001. Springer.
- [12] F. Höppner and F. Klawonn. Learning rules about the development of variables over time. In C. T. Leondes, editor, *Intelligent Systems: Technology and Applications*, volume IV, chapter 9, pages 201–228. CRC Press, 2003. To appear.
- [13] T. M. Mitchell. *Machine Learning*. McGraw Hill, 1997.

- [14] P. Smyth and R. M. Goodman. An information theoretic approach to rule induction from databases. *IEEE Trans. on Knowledge and Data Engineering*, 4(4):301–316, Aug. 1992.
- [15] Sprecher Energie, Linz. *MeteoLiner Kurz-Information*, 1990. User manual for electronic barometer.
- [16] A. P. Witkin. Scale space filtering. In *Proc. of the 8th Int. Joint Conf. on Artificial Intelligence*, pages 1019–1022, Karlsruhe, Germany, 1983.