

# Weak Signals

Anwendungen 1 Ausarbeitung

*Marcel Schöneberg*

[marcel.schoeneberg@haw-hamburg.de](mailto:marcel.schoeneberg@haw-hamburg.de)

**Hochschule für Angewandte Wissenschaften Hamburg (HAW)  
Fakultät für Technik und Informatik  
Department Informatik**

## **Zusammenfassung**

Dieses Dokument stellt eine Einführung in die Thematik der „Weak Signals“, welche als Hilfsmittel für Zukunftsprognosen verwendet werden können, dar. Es werden sowohl grundlegende Definitionen, Herangehensweisen und Ansatzpunkte für die Informatik vorgestellt. Darüber hinaus wird auch das Forschungsumfeld der Weak Signals, sowie auf einen Grundbaustein bei der Analyse von schwachen Signalen – das Data Mining – eingegangen. Abschließend werden zukünftige Ziele und Herausforderungen, mit welchen sich der Autor beschäftigen wird, vorgestellt.

## 1 Einführung

Das folgende Dokument stellt eine Einführung in das Thema „Weak Signals“ dar. Zu diesem Zweck werden verschiedene Themenbereiche aufgegriffen und auf weiterführende Literatur verwiesen.

### 1.1 Definition

„Weak Signal“ (Schwachtes Signal) ist ein Begriff aus der Datenanalyse (oft bezogen auf sehr große Datenmengen – Big Data). Man kann sie u.A. beschreiben als „first symptoms of strategic discontinuities, i.e. symptoms of possible change in the future, acting as warning signs of new possibilities“ [FoS13, Seite 8, nach Holopainen et al]. Sie dienen also häufig zur Vorhersage von möglichen Ereignissen in der Zukunft, da sie erste Anzeichen von möglichen Veränderungen in der Zukunft anzeigen können. Aufgrund ihrer schwachen Ausprägung sind sie häufig auch „too incomplete to permit an accurate estimation“ [FoS13, Seite 9, nach Ansoff]. Also oft sehr vage und daher als Hinweise, nicht aber als Beweise für ein aufkommenden Trend zu verstehen. Im Laufe der Zeit kann sich die Entwicklung eines Weak Signals bzw. der zugehörigen Indikatoren verstärken, so dass es sich gegebenenfalls zu einem Trend oder sogar zum „Next Big Thing“ entwickelt.

Betrachtet man ein schwaches Signal in seiner Eigenschaft als Anzeichen für eine mögliche Entwicklung dann stellt sich u.A. die Frage woher diese Signale kommen und wie man sie entdecken kann. Da eine Entwicklung im Anfangsstadium schwer zu entdecken ist – eben weil sie noch nicht offensichtlich ist – liegt es nahe eine möglichst große Datenbasis für die Suche zu verwenden. Die Art dieser Datenbasis könnten zum Beispiel die Verkaufsdaten eines Händlers, politische Reden, Inhalte sozialer Netzwerke oder jegliche andere Art von Messdaten sein. Diese Menge an Daten ist aber zugleich ein Störfaktor, welcher das Auffinden von Weak Signals verhindert. Da man nicht genau weiß wonach man sucht behindern nicht relevante Daten das Finden von Ergebnissen. Grafik 1 verdeutlicht noch einmal die grundsätzliche Problemstellung eines Weak Signals: Das Auffinden einer unbekanntnen Information in einer großen Datenmenge, welche durch nicht relevante Daten das Signal verdeckt. Hierbei stellt der rote Plot das Weak Signal bzw. dessen Indikator dar, während das blaue Rauschen die Datenbasis zeigt welche als Quelle dient aber gleichzeitig das Signal verdeckt. Aus diesem Grund tritt der Indikator nur selten in den Vordergrund und ist daher schwer zu entdecken.

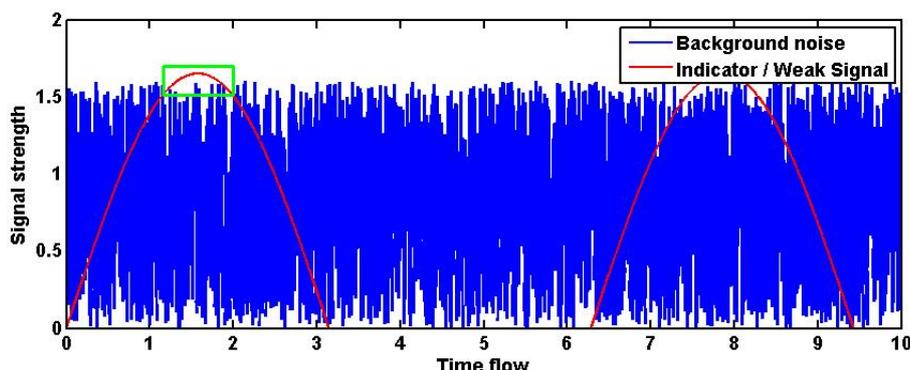


Fig. 1: Illustration eines Weak Signals

## 1.2 Anwendungsfälle

Nachdem die grundsätzliche Problemstellung im Zusammenhang der Weak Signals erläutert wurde stellt sich nun die Frage wofür man diese Signale konkret verwenden kann.

Die Erkennung und Verarbeitung von schwachen Signalen ist in vielen Domänen interessant. Ein populäres Beispiel ist der Katastrophenschutz [QER10, Seite 5], also beispielsweise die Vorhersage des (weiteren) Verlaufs einer Flut. Oder des Verhaltens einer betroffenen Personengruppe nach einem Sturm, welcher zu Beeinträchtigungen im Alltagsleben der Menschen führt. In dieser Domäne wird auch oft der Terminus „(strategic) early warning“ statt Weak Signals benutzt. Ein anderes Anwendungsgebiet ist die Bekämpfung von organisierter Kriminalität [CIS08]. In diesem Umfeld ist es z.B. wichtig vorherzusehen wie sich rivalisierende Banden in der Zukunft entwickeln werden. Interessante Aspekte welche man frühzeitig erkennen will sind zukünftige Felder in denen sich Banden betätigen werden. Ebenso interessant ist ob sich Streitigkeiten zwischen Gruppen entwickeln oder diese zusammenwachsen um ihre Kräfte zu vereinen. Im Bereich der Kriminalitätsbekämpfung wird der Prozess der Weak Signals-Analyse auch als „indicator and warning analysis“ bezeichnet. Ein weiteres Anwendungsbeispiel für die Verwendung von Weak Signals ist das Entdecken von ökonomischen Entwicklungen [Swi13]. Dieses ist wertvoll um Börsenkurse vorherzusagen oder um das Geschäftsmodell eines Unternehmens auf zukünftige Ereignisse abzustimmen. Ein Beispiel dafür ist eine Versicherung welche ihre Policen an der Häufigkeit von Erkrankungen und Unfällen ausrichtet. Kann das Versicherungsunternehmen Trends für Ereignisse vorzeitig ausmachen kann es die Versicherungsbeiträge der Kunden anpassen um den Finanzhaushalt positiv zu beeinflussen. In diesem Umfeld wird die Analyse von weak signals auch als „issue/risk management“ bezeichnet. Ein generalisierter Anwendungsfall von schwachen Signalen ist die Identifizierung und Vorhersage von Trends im Allgemeinen. Dieses ist u.A. eine Zielsetzung des iKnow Projektes [iKn]. Generell wird die Erforschung von Weak Signals daher auch oft unter dem Begriff „future studies“ angesiedelt.

## 2 Kontextbestimmung von Weak Signals

Befasst man sich nun aus Sicht der Informatik mit dem Thema der Weak Signals gilt es zunächst den Kontext in welchem man sich bewegt zu ermitteln.

Das Umfeld der Weak Signals Analyse umfasst u.A das Gebiet „Big Data“ – Ein Schlagwort welches auf die Verarbeitung von großen Datenmengen hinweist. Diese sind oft hilfreich für das Auffinden schwacher Signale, da sich entwickelnde Trends nicht in einzelnen Ereignissen sondern in der Masse manifestieren. Daher ist es erst ab einer entsprechenden Datenmenge sinnvoll möglich Weak Signals bzw. deren Indikatoren wahrzunehmen. Zudem kann eine Analyse auf wenigen Ausgangsdaten zu falschen Schlüssen führen da statistische Ausreißer der Indikatoren falsch gewertet werden [FoS13, vgl. Seite 18]. Gleichzeitig sind die benötigten großen Datenmengen natürlich auch ein Störfaktor bei der Auffindung von Signalen [TEL11, vgl. Seite 7], da die Daten natürlich auch gespeichert und verarbeitet werden müssen. Darüber hinaus enthält eine große Datenbasis eine entsprechende Menge an Informationen die im speziellen Kontext wertlos sein können aber dennoch verarbeitet werden müssen, da es sich um potenzielle Indikatoren handeln könnte [TEL11, vgl. Seite 10]. Diese stellt die Informatik vor Herausforderungen wie die Speicherung und Datenextraktion aus großen Datensets. Ein Beispiel wäre das Auffinden neuer Produkttrends auf der social media Plattform Twitter. Ein Indikator für ein Weak Signal kann u.A. die vermehrte Nutzung eines Hashtags sein. Wenn die vorhandene Datenbasis nur wenige Dutzend Tweets enthält sagt ein vermehrtes Vorkommen wenig

aus in Bezug auf einen wirklichen Trend. Es könnte sich z.B. um befreundete Personen handeln, welche sich über ihren eigenen Einkauf unterhalten, diese Unterhaltung wird höchstwahrscheinlich zu keinem Kauftrend des Produktes führen. Wenn die Anomalie des Vorkommens des Stichwortes sich allerdings auf eine entsprechend große Datenbasis stützt steigt die Wahrscheinlichkeit eines potentiellen Trends, zusammen mit der zu analysierenden Datenmenge. Ein weiteres Gebiet welches die Thematik der Weak Signals anschneidet ist die Künstliche Intelligenz. Die Methoden dieser Disziplin können hilfreich sein um zu ermitteln welche Informationen aus vorliegenden Daten relevant sind. Daher können Fragen wie „Nach was genau sucht man?“ und „Wie kann man nützliche Informationen extrahieren?“ oft mit Methoden der künstlichen Intelligenz beantwortet werden. Konkret bieten die Ansätze der künstlichen Intelligenz Hilfestellungen wenn es darum geht Daten zu kategorisieren und Anomalien zu finden. Diese Herangehensweisen bieten daher eine gute Basis um Trends – also Veränderungen in Bezug auf eine Basis – zu entdecken [TEL11, vgl. Seite 19 ff. ].

Weiterhin ist auch domänenspezifisches Wissen im Umfeld der Analyse von Weak Signals interessant. Betrachtet man das einführende Beispiel der Kauftranderkennung auf Twitter können Experten u.a. behilflich sein bei der Beantwortung der Frage welche Art von Veränderungen normal innerhalb des social media Kontextes sind und welche dagegen eine Anomalie darstellen. Ebenfalls können Marketingexperten darüber informieren welche Zielgruppe für sie interessant ist und welche Eigenschaften diese Zielgruppe hat. Gegebenenfalls sind Experten auch in der Lage Indikatoren zu benennen welche priorisiert verfolgt werden sollten, da diese bereits in der Vergangenheit in anderen Märkten (z.B. dem US-Markt) von Bedeutung waren.

Darüber hinaus ist auch das Feld der Soziologie von Interesse. Dieses betrifft zwar nicht die direkte Verarbeitung von Daten, wohl aber die Voranalyse der Situation, also die Aufstellung von ersten Indikatoren. So leiden Experten ggf. an einem Tunnelblick und sehen nur das was sie erwarten zu sehen [FoS13, vgl. Seite 18]. Dieser Aspekt engt die ursprünglich möglichst neutrale Suche ein und erschwert es Entscheidungsträger von möglichen Funden der Analyse zu überzeugen.

### 3 Grundlegende Herangehensweise

Der nächste Abschnitt widmet sich dem grundsätzlichen Vorgehen zur Erkennungen und Verarbeitung von Weak Signals. Hierzu wird eines von vielen Vorgehensmodellen [CIS08] vorgestellt, welches die Phasen Erkennung, Bewertung und Einschätzung beinhaltet. Das gewählte Modell wurde vom Criminal Intelligence Service Canada (CISC) entwickelt und ist in Abbildung 2 zu sehen. Die Hervorhebungen stammen vom Autor und stellen Bereiche da, welche seiner Meinung nach zumindest teilweise automatisiert angegangen werden können.

Das Modell beginnt mit der „Thread Perception“, also dem Versuch eine Bedrohung zu erkennen. Hierzu wird zunächst die Umwelt (also alle denkbaren Datenquellen wie z.B. Soziale Medien, Meinungsäußerungen, Reden, Interviews etc.) festgehalten um eine Datenbasis zu schaffen. Der nächste Schritt ist die Entwicklung von denkbaren Bedrohungsszenarien auf Basis der Daten und vorhandenem Domänenwissen Ein Beispiel für ein solches Szenario kann im Umfeld der organisierten Kriminalität u.A. das Verhalten rivalisierender Banden sein untereinander sein.

Die zweite Phase welche das Vorgehensmodell vorsieht ist „Evaluation and Monitoring“, diese führt eine grundsätzliche Analyse der Daten aus. Hierbei wird versucht die „Sentinel Watchlist“ aufzubauen, welche festlegt welche möglichen Ereignisse tiefer ge-

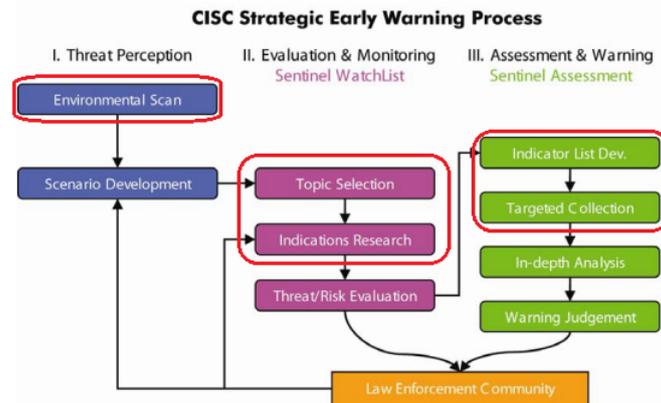


Fig. 2: CISC Methodology [CIS08] mit vom Autor hervorgehobenen Informationsansätzen

hend betrachtet werden sollen. Der Inhalt dieser Liste ist eine Untermenge aller möglichen Ereignisse welche vorstellbar wären innerhalb eines zuvor gewählten Szenarios. Ein Beispielszenario könnte die schon erwähnte Bandendynamik in der organisierten Kriminalität sein. In diesem Kontext müsste man sich zunächst ein konkretes Unterthema suchen, welches weiter untersucht werden soll (beispielsweise das mögliche Zusammenwachsen der erwähnten Gruppen um ihre Kräfte zu vereinen). In dem gewählten Themengebiet kann man sich nun mit Hilfe von Expertenwissen Indikatoren suchen, welche eine Veränderung andeuten können. Der letzte Schwerpunkt in der Evaluierungsphase ist die Einschätzung der Funde. Die grundsätzliche Frage die sich stellt ist ob es die bisherigen Ergebnisse eine weitere Analyse rechtfertigen, daher muss eine Risikoanalyse erstellt werden. Am Beispiel der organisierten Bandenkriminalität im Kontext des möglichen Zusammenwachsens von Gruppen werden bei dieser Analyse die Fragen beantwortet ob das Ereignis starke Auswirkungen hätte welche Gegenmaßnahmen erfordern würden. Eine andere Frage die die Analyse beantworten muss ist der zeitliche Kontext in dem das Ereignis stattfindet. Ereignet sich das Zusammenwachsen der Banden in der näheren Zukunft oder ist es etwas das in einem Zeitrahmen von mehreren Jahren passieren kann und daher momentan nicht näher bewertet werden kann und muss. Diese und ggf. kontextspezifische weitere Fragen müssen im Rahmen der Analyse geklärt werden um das weitere Vorgehen sowie dessen Nutzen zu planen zu evaluieren.

Die dritte Phase des CISC-Vorgehensmodells befasst sich mit der weiteren Bewertung der Funde und der möglichen Erstellung einer Warnung welche an Entscheidungsträger weitergereicht werden kann. Dieser letzte Schritt erstellt aus dem bisherigen Wissen eine gewichtete Liste von Indikatoren welche eine Veränderung anzeigen können. Des Weiteren werden näher zu untersuchende Aspekte (beispielsweise das soziale Netzwerk und Umfeld einer Gruppierung) festgelegt und von erfahrenen Analysten und Domänenexperten auf Anomalien hin geprüft. Sollten sich dabei konkrete Anhaltspunkte für eine Warnung ergeben wird diese erstellt und an Entscheidungsträger (z.B. in der Legislative) weitergegeben. Diese können daraufhin dementsprechende Gesetze und Initiativen entwerfen und so proaktiv auf eine Bedrohung vor ihrem Eintreten reagieren.

### 3.1 Automatisierbare Prozesse

Aufbauend auf dem Vorgehensmodell stellt sich nun die Frage an welchen Stellen die Informatik hilfreich sein kann. Nach Einschätzung des Autors gibt es Prozessteile des Modells welche sich (teilweise) automatisieren lassen und andere die sehr vom Wissen und der Intuition der Domänenexperten abhängig sind. Ebenso sind die als nicht automatisierbar eingestuften Aufgaben sehr sensibel und sollten daher auch aufgrund Sicherheitsgründen sowie ethischen Bedenken im Zweifelsfall besser von Menschen bearbeitet werden. Die in Abbildung 2 hervorgehobenen Gebiete gehören der ersten Kategorie an und werden daher im Folgenden näher beleuchtet.

#### 3.1.1 Aufgaben im Bereich der Bedrohungserfassung

Eine der wichtigsten Aufgaben welche im Analyseprozess zur Auffindung von Weak Signals erfüllt werden muss ist ein Scan der zu untersuchenden Medien (z.B. soziale Netzwerke oder Dokumente). Diese erste Betrachtung legt den Grundstein für alle weiteren Schritte indem Daten erfasst werden. Diese Datenerfassung sollte kontinuierlich fortgesetzt werden um einen möglichst umfassenden Blick auf die aktuelle Lage zu erhalten. Da das Sammeln aller verfügbarer Daten praktisch unmöglich ist kann hier bereits eine erste grobe Einschränkung auf Dokumenten- oder Stichwortbasis vorgenommen werden. Hierbei ist zu beachten das eine weitere Spezifizierung der zu analysierenden Daten auch in den späteren Phasen stattfindet. Daher können aufgrund des Konfliktes zwischen einem theoretischen Modell und der Realität Überschneidungen auftreten, sodass eine erste Einschränkung zu grob ausfällt und damit zu viele Daten aus dem Fokus der Analyse nimmt. Das grundsätzliche Erfassen von Daten kann in vielen Fällen ohne großen Aufwand automatisch geschehen. So kann man Programmierschnittstellen (APIs) sozialer Medien oft direkt dazu nutzen Daten zu beziehen (entsprechende Rechte vorausgesetzt). Ebenso denkbar wäre der Download von Protokollen politischer Diskussionen oder das Erfassen von Standortpositionen verdächtiger Personen um nur einige Beispiele zu nennen. Diese Beispiele stellen nur einen kleinen Teil der technischen Möglichkeiten dar und lassen sich nahezu beliebig erweitern. Allen Methoden zur Datenerfassung ist allerdings gemein, dass sie wesentlich effizienter automatisiert funktionieren als manuell durch Menschen.

#### 3.1.2 Aufgaben im Bereich der Evaluierung

Betrachtet man die zweite Phase des CISC-Modells haben vor allem die Aufgaben der Themenauswahl und der Suche nach Indikatoren Automatisierungspotenzial. Die Themenauswahl fokussiert die folgenden Analyseschritte auf einen konkreten Bereich welcher detaillierter untersucht werden soll. Diese Einschränkung ermöglicht das Aufstellen von Indikatoren als Hinweise auf einen möglicherweise aufkommenden Trend. Für das schon angesprochene Szenario im Rahmen der organisierten Kriminalität kann eine starke Vernetzung in einem Graphen (welcher die in Banden organisierten Personen samt ihrer Kontakte darstellt) ein Anhaltspunkt für ein Zusammenwachsen sein. Zeigt sich in diesem Netzwerk eine starke Verbindung (z.B. durch viele Treffen oder gemeinsame Bekannte) zwischen rivalisierenden Personen könnte dieses auf ein nachlassen der Feindschaft hinweisen. Ebenso kann ein gehäuftes Vorkommen von Begriffen, Orten oder Namen Rückschlüsse auf zukünftige Ereignisse ermöglichen. Diese Anomalien können erste Indikatoren darstellen oder zumindest als Basis dienen. Wenn spätere Auswertungen (auf neueren oder vollständigeren Daten) grob von der Ausgangssituation abweichen kann eine Veränderung (und damit ggf. ein Trend) vermutet werden. Das Erstellen von Influence-Netzwerken (wie den

erwähnten Graphen) sowie das Analysieren von Worthäufigkeiten kann – entsprechende Daten vorausgesetzt – sehr gut automatisiert werden. Ebenso können vorhandene Daten automatisch in verschiedene Grobkategorien eingeordnet werden. Entsprechend könnte auch das ungefähre Thema eines Textes ermittelt werden. All dieses vereinfacht eine tiefer gehende Analyse von Daten und liefert erste Anhaltspunkte, Anomalien und Indikatoren ohne das größerer Aufwand von Menschen benötigt wird.

### 3.1.3 Aufgaben im Bereich der Bewertung und Erstellung von Warnungen

In der letzten Phase des CISC Modells steht die Bewertung der Funde und die ggf. erforderliche Erstellung von Warnungen im Vordergrund. Aufgrund der Funde der vorherigen Phasen können so Entscheidungsträger informiert werden um weitere Schritte einzuleiten. Die eigentliche Bewertung der Funde kann kaum automatisiert werden da hierzu Expertenwissen, Erfahrung sowie ggf. Befragungen von Personen usw. erforderlich ist. Allerdings ist es möglich, dass die in den vorherigen Phasen entstandene Menge an neuen (komprimierten) Informationen (also z.B. Indikatoren) wiederum Anhaltspunkte für weitere Untersuchungen gibt. Hat bei beispielsweise in vorherigen Arbeitsschritten die Themen mehrerer Dokumente grob in Kategorien geordnet kann man darauf basierend den Scan der Umwelt erneut starten um den Blickwinkel zu erweitern. Als Input für diesen Durchlauf können die ermittelten Indikatoren und Stichwörter der Kategorien dienen. Die dritte Phase des Modells kann also auch als Feedbackmechanismus für die ersten Phasen aufgefasst werden, indem Metainformationen die Suche automatisch erweitern.

## 4 Forschungsumfeld

Betrachtet man die Forschungslandschaft der schwachen Signale so ist zunächst Harry Igor Ansoff (1918 in Wladiwostok †2002 in San Diego) zu benennen. Er legte 1975 den Grundstein für die Erforschung von Weak Signals. Der Mathematiker und Ökonom brachte das Thema in seiner Theorie zur „strategischen Reaktionen auf schwache Signale“ [Ans75] auf. Diese beschreibt ein mehrstufiges System zur Bewältigung von Veränderungen. Hierbei werden Schritte von der Erkennung eben dieser bis zur Berechnung ihres Einflusses betrachtet. Dieses Beispiel ist nur eines von vielen welche zeigen, dass die Erforschung von Weak Signals durchaus kein neues Thema ist. Es hat im Gegenteil (für Informatikverhältnisse) alte Wurzeln und wurde bereits damals mit großem Interesse verfolgt. Ein weiteres Beispiel für diese Art der Forschung sind Delphie-Studien [TEL11, vgl. Seite 14] (entwickelt in den 1950 und 1960 Jahren) welche ebenfalls das Ziel verfolgen Trends und Entwicklungen vorherzusagen.

Untersucht man die aktuellen Entwicklungen auf dem Gebiet der schwachen Signale so stellt man fest, dass die gesamte Thematik derzeit einen erneuten Aufschwung erfährt. Dieses ist vermutlich auch auf die mittlerweile zur Verfügung stehende Rechenleistung (man bedenke die Entwicklung nach dem Mooreschen Gesetz von 1960 bis heute), die vorhanden Datenmengen (z.B. aus sozialen Netzwerken), sowie neue Methoden der Analyse zurückzuführen. An diesem Boom des Themas, sowie der weiteren Erforschung, sind daher auch verschiedene Organisationen interessiert. Die Bandbreite schließt daher u.A. die europäische Kommission, im Rahmen von Forschungsprogrammen [TEL, z.B.] und Geheimdienste [CIS08], welche möglichen Gefahren früh entgegenwirken wollen ein. Ebenso sind Forscher aus verschiedenen Fachbereichen (z.B. Frühwarnsysteme für Katastrophenschützer), sowie Wirtschaftsunternehmen (beispielsweise. Versicherungen) engagiert um damit Leben sowie Investitionen zu sichern. Diese Vielfältigkeit spiegelt

sich auch zum Teil in den verschiedenen Fachkonferenzen und Workshops wieder. Diese umfassen u.A.:



Die „European Conference on Technology Enhanced Learning“ [EC:]. Diese ist stark mit dem TEL Map Projekt der europäischen Union verbunden, welches sich die Erkennung von aufkommenden Trends als Ziel gesetzt hat.



Ein weiteres Zusammentreffen ist die „Conference on Knowledge Discovery and Data Mining“ (KDD) [ACMa] der „Association of Computing Machinery“ (ACM) bzw. der entsprechenden „Special Interest Group“ (SIGKDD).



Die „International Conference of the System Dynamics Society“ [SDS], welche mit der ACM „Special Interest Group Simulation and Modeling“ (SIGSIM, [ACMb]) verbunden ist deckt einen weiteren Aspekt der Forschung ab und deckt neben Themen der Simulation von komplexen Systemen in diesem Rahmen auch die Erforschung von schwachen Signalen ab.

Darüber hinaus gibt es verschiedene (unregelmäßige) Workshops [FoS13, vgl. z.B.] zum Thema, sowie viele Paper verschiedener Fachbereiche (siehe 1.2) welche das Thema (aus der entsprechenden Sichtweise) behandeln.

## 5 Datamining im Weak Signals Umfeld

Die nächsten Abschnitte betrachten erste Ansätze zur Datenanalyse im Bereich der Weak Signals. Wie bereits angesprochen bieten hier Methoden der Informatik viele Möglichkeiten: Hier wären z.B. Datenbanktechnologien, verteilte Verarbeitung von Daten und Visualisierung von Ergebnissen Ansatzpunkte. Auch können Simulationen (z.B. mit Hilfe von (Multi-)Agentensystemen) wertvolle Hilfsmittel zur Analyse sein. Mit diesen kann man verschiedene Ereignisse in einem System simulieren um die Dynamik eines Systems zu begreifen. Mindestens ebenso wichtig ist die Extraktion von Informationen aus Daten. Dieses kann z.B. mit Hilfe von Techniken des Datamining und der Analyse von (sozialen) Netzwerken passieren. All diese Methoden können dazu beitragen Indikatoren zu finden, Modelle zu bauen und so ein System besser zu verstehen um dadurch Vorhersagen zu ermöglichen und letztendlich das Wissen weiterzugeben.

Konzentriert man sich zunächst nur auf die Extraktion von Informationen aus den vorhandenen Daten (im einfachsten Fall Dokumente) gibt es ebenfalls eine große Spannweite von Möglichkeiten. Eine Herangehensweise wird im folgenden behandelt und soll als vergleichsweise einfacher erster Ansatz dienen – das Textmining [TEL11, vgl. Seite 23 ff.]. Wenn man zunächst davon ausgeht, dass die vorhandenen Daten sich nur auf Text beschränken (also Beispielsweise Reden, Interviews, Beiträge in sozialen Medien, Gesprächsmitschnitte usw.) dann kann man jeden vorhanden Text als eine Anhäufung von Worten auffassen und dabei die Semantik des Inhalts ignorieren. Grundsätzlich kann man im Weiteren eine Vorverarbeitung der Texte durchführen und dabei Elemente welche, im einfachsten Fall, nicht interessant sind entfernen (z.B. Satzzeichen, Konjunktionen etc.) und die Worte in Gruppen von  $n$  Wörtern zusammenfassen (sogenannte  $n$ -Gramme). Darüber hinaus kann man ausgewählte Worte mit Meta-Informationen anreichern (denkbar ist z.B. die Semantik, dass ein Wort eine Negation andeutet). Aufbauend auf einer solcher Wortsammlung ist es möglich Dokumentenstatistiken zu entwickeln um weitere Schlüsse zu ermöglichen. Einige Ideen für Statistiken auf Textbasis und ihre Anwendung sollen im folgenden gegeben werden.

**Steigende und fallende Terme** Diese Methode baut eine Statistik über das Vorkommen jedes n-Gramms innerhalb des analysierten Textes auf und vergleicht diese zu einer vorher ermittelten Basis. Wird nun ein n-Gramm überdurchschnittlich häufig/selten benutzt lässt sich dieses ermitteln und kann somit ein Anzeichen für eine Veränderung bzw. Entwicklung sein.

**Wahrscheinlichkeitsmodelle zur Ermittlung von Themen** Basierend auf Häufigkeitsstatistiken ist es ebenso denkbar das Thema eines Textes zu bestimmen. Ein Thema kann beispielsweise dadurch eingegrenzt werden, dass Schlüsselwörter häufiger vorkommen. Ein Beispiel hier für wäre ein politischer Kontext indem Namen von Politikern oder Schlagwörter aktueller Themen häufiger verwendet werden. Diese Methode setzt allerdings in vielen Fällen voraus das bereits klassifizierte Texte vorliegen gegen welche verglichen bzw. kategorisiert werden kann.

**Aufkommende Themen** Diese Idee verbindet die Idee der steigenden und fallenden Terme mit den Themenmodellen und versucht aufkommende Themen auf analoge Weise zu entdecken.

**Domänenübergreifende Vergleiche** Dieser Ansatz beruht auf der Idee Indikatoren für Entwicklungen aus anderen Domänen zu übertragen um in der Zieldomäne Änderungen zu erkennen. Hierbei sind zwei Arten zu unterscheiden.

Die erste Herangehensweise nimmt sich eine Domäne im gleichen Umfeld und überträgt die Erkenntnisse. Ein Beispiel hier für kann die Kaufentwicklung auf dem US-Markt sein, da dieser den europäischen Kaufrends häufig einige Zeit voraus ist. Mit den Indikatoren des US-Marktes (z.B. Schlüsselwörter in Kommentaren welche positives Feedback anzeigen) kann man nun den europäischen Markt (bzw. Kommentare) frühzeitig als Trend erkennen und die Entwicklung vergleichen.

Die andere Herangehensweise überschreitet die Domänengrenze und übernimmt Indikatoren aus anderen Bereichen. Ein Beispiel hierfür sind Indikatoren welche sich im Smartphone-Umfeld bewiesen haben. Diese kann man auf den Markt für smart TVs übertragen und so versuchen Entwicklungen vorherzusehen.

Ein komplett anderes Konzept des Datamining ist die Netzwerkanalyse [Pul11, vgl.]. Hierbei wird versucht Einflussnetzwerke zu entwickeln um basierend auf diesen Daten vorauszusagen wie sich z.B. Meinungen verbreiten. Ein anschauliches Beispiel für solche Netzwerke sind soziale Medien. Diese enthalten häufig wertvolle Informationen, beispielsweise wer eine bestimmte Person kennt und wie diese zueinander stehen bzw. welchen Bekanntenkreis sie teilen. Darüber hinaus stehen auch Daten über die Interessen dieser Personen so wie gegebenenfalls ihr Standort etc. zur Verfügung. Da sich Meinungen gegenseitig beeinflussen kann man nun versuchen Schlüsse aufgrund der Verbindung von Menschen zu ziehen, so ist eine stärker vernetzte Person einflussreicher, da sie mehr Leute erreicht. Fügt man dem Modell Meta-Informationen wie z.B. öffentliche Wahrnehmung oder Ähnliches hinzu ließe sich das Modell damit ggf. noch verbessern.

## 6 Ausblick

Im weiteren Verlauf des Studiums wird der Autor sein Wissen über das Thema der schwachen Signale erweitern und Ansätze für einen konkreten Anwendungsfall entwickeln. Hierzu muss einerseits die vorhandene Literatur tiefer durchdrungen werden und andererseits müssen vorhandene Frameworks (z.B. für Datamining) sowie Datenquellen auf ihre Brauchbarkeit geprüft werden.

Das vom Autor gewählte konkrete Thema ist Politik. Gründe hierfür sind vor allem die Allgegenwärtigkeit politischer Themen, sowie wie die Schwierigkeit Aussagen in dieser Domäne zuzuordnen und für jeden zugänglich zu machen. Der Autor ist der Überzeugung, dass mehr Informationen aus den vorliegenden Daten gewonnen werden könnten wenn man diese verbindet und auswertet. Beispiele sind Reden welche nur zu Protokoll gegeben werden, sowie Meinungen welche in verschiedenen Komitees wachsen aber nur partiell nach außen dringen. Die Zielsetzung des Autors ist damit die Ermittlung der Möglichkeiten zum Auffinden von politischen Trends bevor diese offensichtlich werden. Ein konkreter Anwendungsfall kann sich zum Beispiel in der Retroperspektive mit der politischen Entwicklung der Piratenpartei in Deutschland beschäftigen. Dieses hat den Vorteil, dass die gewonnen Informationen vergleichsweise einfach validiert werden können. Allerdings ist dieses nur eines von vielen interessanten Themen welche denkbar sind.

## 7 Zusammenfassung

Diese Ausarbeitung diente als Einführung in die Analyse von Weak Signals (im Big Data Umfeld). Hierzu wurde zunächst eine grundlegende Einführung samt Definition und Anwendungsfällen gegeben. Daraufhin wurde die Sicht der Informatik auf Weak Signals dargelegt um den Kontext des Dokumentes abzugrenzen. Im Folgenden wurde exemplarisch ein Vorgehensmodell zur Analyse von schwachen Signalen vorgestellt. Hierbei wurde die Idee zur Entdeckung von Anomalien/Indikatoren und der damit verbundenen Erstellung von Prognosen erläutert. Auch wurde auf den Aspekt der Automatisierbarkeit sowie auf Problemstellungen solcher Prozesse eingegangen. Daraufhin wurde das Forschungsumfeld der schwachen Signale vorgestellt. Hierbei wurde die Geschichte, sowie aktuell am Thema interessierte Organisationen vorgestellt. Im Weiteren wurde einer der Grundbausteine zur Analyse von Weak Signals – das Datamining – in aller Kürze eingeführt um eine mögliche Herangehensweise vorstellen. Zum Schluss wurde ein Ausblick auf die weiteren Ziele des Autors gegeben.

Zusammenfassend stellt sich das Forschungsgebiet der schwachen Signale als interessantes Thema dar welches viele Einsatzgebiete für die Informatik bietet. Hierbei kann sowohl auf aktuelle Forschungsthemen (z.B die Erfassung und Verarbeitung von großen Datenmengen) sowie auf bereits etablierte aber immer noch forschungsintensive Themen wie Datamining Bezug genommen werden. Ebenfalls ist die Interdisziplinarität des Gebietes durchaus interessant, allerdings erschwert es dadurch die Arbeit, da häufig Domänenwissen von Vorteil ist. Gleiches gilt für die Themenvielfalt im Bereich der Weak Signals. Diese führt zwar zu einigen interessanten Beiträgen auf dem Gebiet, allerdings sind diese häufig stark auf einen bestimmten Anwendungsfall zugeschnitten und daher nur sehr eingeschränkt zur Wiederverwendung geeignet. Dieser Sachverhalt zeigt sich auch sehr konkret in den fachlichen Beiträgen der Informatikdomäne. So erstreckt sich das Forschungsfeld auf viele Bereiche sowie Konferenzen und ist damit nur schwer zu überblicken. Auch der derzeitige Boom der schwachen Signale führt häufig zu negativen Folgen wie populärwissenschaftlichen Veröffentlichungen und damit ähnlichen aber verschiedenen Definitionen sowie einer Reihe von Ansätzen welche aus Informatik Sicht wenig erfolgversprechend sind. Dieses zeigt sich auch in der Vielzahl der Begrifflichkeiten welche mehr oder minder synonym für schwache Signale verwendet werden. Ein weiterer Effekt dieser Vielfältigkeit sind Begriffsüberdeckungen mit anderen Gebieten wie Elektrotechnik oder der BWL.

Trotzdem stellt sich die Thematik der Weak Signals trotz ihrer Nachteile als lohnenswertes Gebiet dar in welchem durchaus interessante Forschung betrieben werden

kann.

## Literatur

- [ACMa] Association for Computing Machinery SIGKDD: **Conference on Knowledge Discovery and Data Mining**. <http://www.kdd.org>, Abruf: 2013-10-24
- [ACMb] Association for Computing Machinery SIGSIM: **Special Interest Group on Simulation and Modeling**. <http://www.acm-sigsim-mskr.org/>, Abruf: 2013-10-24
- [ACM12] Association for Computing Machinery: **Computing Classification System**. <http://www.acm.org/about/class/2012?pageIndex=0>.  
Version: 2012, Abruf: 2013-10-20
- [Ans75] ANSOFF, Harry I.: Managing Strategic surprise by response to weak signals. In: **California Management Review** 18 (1975), Nr. 2, S. 21–33
- [CIS08] Criminal Intelligence Agency Canada: **Strategic Early Warning for Criminal Intelligence**. 2008
- [EC:] **European Conference on Technology Enhanced Learning**. <http://www.ec-tel.eu/>, Abruf: 2013-10-24
- [FoS13] Forschungsforum Öffentliche Sicherheit: **Dokumentation: Intensiv Workshop Weak Signals in der Sicherheitsforschung**. 2013
- [iKn] European Commission FP7: **iKnow Project**. <http://www.iknowfutures.eu/>, Abruf: 2013-10-26
- [Pul11] PULWITT, Jan-Martin: **Dynamic influence network models and weak signals in social media**. <http://dbis.rwth-aachen.de/cms/theses/dynamic-influence-network-models-and-weak-signals-in-social-media>.  
Version: 2011, Abruf: 2013-10-26
- [QER10] QUANSAH, Joseph E. ; ENGEL, Bernard ; ROCHON, Gilbert L.: Early Warning Systems. In: **Journal of Terrestrial Observation** 2 (2010), Nr. 2
- [SDS] System Dynamics Society: **International Conference of the System Dynamics Society**. <http://www.systemdynamics.org/>, Abruf: 2013-10-24
- [Swi13] Swiss RE: **Emerging risk insights**. 2013
- [TEL] European Commission: **Technology-Enhanced Learning MAP**. <http://www.telmap.org>, Abruf: 2013-10-24
- [TEL11] TEL Map: **D4.1 Report on Weak signals collection**. 2011
- [Wah13] WAHL, Sonja von: **Früherkennung unbekannter Gefahren durch „weak signals“**. 2013