

Automatisierte Rückfragen für eine dynamische Klassifikation von Telefonanrufen

Diplomarbeit in Wirtschaftsinformatik
Institut für Informatik der Universität Zürich
Prof. Abraham Bernstein

Simon Bleher
St. Gallen, Schweiz
00-908-327

Betreuer
Peter Vorburger

22. September 2006

Zusammenfassung

Die Ablenkung durch negative externe Einflüsse ist im Arbeitsalltag nicht zu unterschätzen und kann zu Produktivitätseinbußen führen. Je nach Stärke der Unterbrechbarkeit können bereits kleinste Störungen den Arbeitsfluss stoppen. Eine bis heute nur schwer kontrollierbare Quelle von solchen Störungen ist der Telefonanschluss.

Ein System zur Filterung von Telefonanrufen nach ihrem aktuellen Bezug zur Unterbrechbarkeit des Angerufenen kann dieses Problem lösen. Diese Diplomarbeit versucht deshalb den Inhalt eines Telefonanrufes in eine solche Klassifikation zu integrieren. Durch automatisierte Rückfragen an den Anrufer wird eine weitere Verbesserung des Systems erwartet.

Abstract

Distractions by negative external influences in the daily business routines are not to be underestimated since they can lead to severe productivity losses. Depending on the interruptability level, already minor disturbances may discontinue the workflow. One of today's scarcely controlled origins of such disturbances is the telephone.

A system that filters incoming phone calls according to their actual context to the interruptability could solve this problem. This diploma thesis suggests a system for classification of phone calls with automatic inquiry to increase the system's performance.

Danksagung

Ganz herzlich möchte ich mich bei allen Personen bedanken, die mich während den letzten sechs Monaten unterstützt haben. Ein besonderer Dank geht an meine wundervolle Freundin Anne-Käthi und meine Eltern, welche mir dieses Studium ermöglichten. Auch danke ich Peter Vorbürger für die äusserst angenehme und kompetente Betreuung sowie Matthias für die vielen interessanten Diskussionen.

Inhaltsverzeichnis

1 Einleitung	6
1.1 Ausgangslage.....	6
1.2 Motivation.....	7
1.3 Konzept.....	7
1.4 Beschreibung der Kapitel.....	8
2 Einordnung in die aktuelle Forschung	9
2.1 Sprachdialogsysteme.....	9
2.2 Telefonspamfilter.....	10
2.3 Spracherkennung.....	11
2.4 Textmining.....	11
2.5 Unterbrechbarkeit.....	12
3 Grundlagen	14
3.1 Spracherkennung.....	14
3.1.1 Vorverarbeitung.....	14
3.1.2 Erkennung.....	15
3.2 Textmining.....	16
3.2.1 Vorgehensweise.....	16
3.2.2 Vorverarbeitung.....	17
3.2.3 Vektordarstellung.....	19
3.2.4 Klassifikationsmodelle.....	20
3.3 Evaluation.....	21
3.3.1 Kreuzvalidierungsverfahren.....	21
3.3.2 Confusion Matrix.....	22
4 Experiment	23
4.1 Problemstellung.....	23
4.2 Übersicht.....	24
4.3 Datenerfassung.....	25
4.3.1 Interview.....	25
4.3.2 Simulation der Anrufe.....	25
4.3.3 Klassierung.....	26
4.4 Spracherkennung.....	27

4.5 Machine-Learning-Framework.....	27
4.5.1 Klassendiagramm.....	27
4.5.2 Morphologie.....	29
4.5.3 WordVectorTool.....	29
4.5.4 Weka.....	29
4.6 Textmining.....	30
4.6.1 Vorverarbeitung.....	30
4.6.2 Einfache Klassifikation.....	31
4.6.3 Klassifikation mit Hintergrundwissen.....	31
4.6.4 Klassifikation mit Rückfragen.....	32
4.7 Evaluation.....	35
4.7.1 Datensatz (deskriptive Statistik).....	35
4.7.2 Einfache Klassifikation.....	38
4.7.3 Klassifikation mit Hintergrundwissen.....	40
4.7.4 Klassifikation mit Rückfragen.....	41
4.8 Einschränkungen.....	46
5 Schluss	47
5.1 Diskussion.....	47
5.2 Ausblick.....	48
Literaturverzeichnis	50
Abbildungsverzeichnis	53
Tabellenverzeichnis	54
Anhang A: Vorverarbeitung dreier Mustersätze	55
Anhang B: Aufstellung der Anrufergruppen	57
Anhang C: Anweisung für Simulationsanruf	58
Anhang D: Korrekturen der Anrufe	59
Anhang E: Einfache Klassifikation	61
Anhang F: Klassifikation mit Hintergrundwissen	64
Anhang G: Klassifikation mit Rückfragen	65

1 Einleitung

1.1 Ausgangslage

Die Kommunikation ist eine der bedeutendsten kulturellen Errungenschaften, die der Mensch in seinem täglichen Leben benutzt. Ohne sie wäre ein Zusammenleben kaum vorstellbar. Nebenst meist positiven Aspekten stört sie uns in nicht wenigen Fällen. Von Werbespots im Radio, über Massen von Spammails bis zu unerwünschten Telefonanrufen. Diese allgegenwärtigen Störungen kann während der Arbeitszeit zu Produktivitätseinbussen führen.

Ein System, das Telefonanrufe in Abhängigkeit von deren Wichtigkeit filtert, ähnlich einem Spamfilter beim Email, wäre deshalb für viele Anwendungen äusserst nützlich. Als Filterkriterien können verschiedene Merkmale eines Telefonats dienen. Vorstellbar wären in erster Linie die Identität des Anrufers, die Anrufszeit oder die Dauer des Wartens in der Leitung. Der Inhalt eines Gespräches ist jedoch, im Gegensatz zum Inhalt eines Emails, bei der Filterung von Spammail nicht vorgängig verfügbar und kann deshalb zum heutigen Zeitpunkt nicht einbezogen werden. Es scheint deshalb sinnvoll, eine Möglichkeit zu suchen, mit welcher der Inhalt eines bevorstehenden Telefongesprächs analysiert werden kann.

Auf verschiedenen Wegen kann versucht werden von einem Anrufer sein Anliegen oder wenigstens Teile davon zu erfahren, bevor das Telefonat an seinen Bestimmungsort durchgestellt wird. In vielen Callcentern werden heutzutage Eingaben über die Telefontastatur benutzt. Je vielfältiger jedoch das in Frage kommende Anwendungsgebiet ist, desto unübersichtlicher und mühsamer wird die Navigation durch die verschiedenen Menüpunkte für den Anrufer [Suhm et al. 2002]. Einen anderen Ansatz wird im Kapitel 2.1 vorgestellt. Statt mühsamen Tastatureingaben nimmt das Callcentersystem gesprochene Anweisungen des Anrufers entgegen. So kann interaktiv versucht werden den Inhalt möglichst gut zu erfassen. Dies gelingt bis jetzt jedoch meistens nur für eine wohldefinierte Domäne.

Diese Diplomarbeit versucht ein System zu entwickeln, welches die Klasse eingehender Telefonanrufe vorhersagt. Die Klasse eines Telefonanrufes ist in dieser Diplomarbeit gleichbedeutend mit seiner Wichtigkeit für den Angerufenen. Explizit soll bei der

Klassifikation der Inhalt in die Bewertung einfließen und gegebenenfalls durch eine Rückfrage an den Anrufer präzisiert werden.

1.2 Motivation

Eingebettet ist diese Arbeit in den Kontext eines intelligenten Büros, das versucht die Unterbrechbarkeit seines Benutzers festzustellen und ihn möglichst gut vor negativen, externen Einflüssen zu schützen. Folgende Arbeiten wurden dazu bereits an der Universität Zürich verfasst: [Vorburger et al. 2005] [Donner 2005].

Das Telefon ist eine Störquelle im Arbeitsalltag, die nur schwer kontrollierbar ist. Bereits das Klingeln reisst uns meistens aus dem Arbeitsfluss, was die Produktivität und Arbeitszufriedenheit mindert. Wenn Anrufe hingegen das aktuelle Arbeitsgebiet betreffen, können sie auch fördernd sein. Eine Untersuchung dazu wird in Kapitel 2.5 vorgestellt.

Das Abschalten des Telefons ist im Normalfall keine zufrieden stellende Alternative, da wichtige Anrufe während dieser Zeit eintreffen könnten. Auch sind manuelle Wechsel des Telefonstatus mühsam und fehleranfällig. Die Abnahme der Entscheidung durch ein automatisches System, ob ein Telefonanruf durchgestellt oder abgewiesen werden soll, kann deshalb äusserst nützlich sein.

1.3 Konzept

Folgend wird in groben Zügen das Konzept des Systems zur Klassifikation von Telefonanrufen skizziert, das in dieser Diplomarbeit entwickelt wird. Mit einem Experiment soll die Wirkungsweise dieses Systems schliesslich analysiert werden.

Der Weg, um die Klasse eines Anrufes zu ermitteln, führt über mehrere Komponenten. Ein ankommender Anruf muss zuerst vom System entgegengenommen und durch eine Spracherkennung in Text umgewandelt werden. Danach wird mittels Textmining versucht die Klasse vorherzusagen. Durch geschicktes Rückfragen kann das erhaltene Resultat verbessert werden.

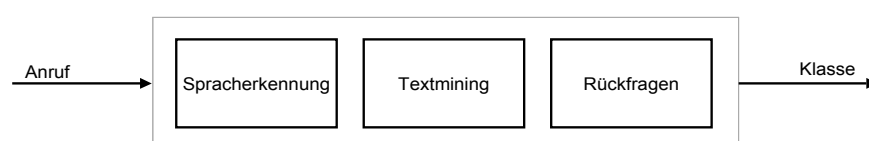


Abbildung 1.1: Konzept des Klassifikationssystems

1.4 Beschreibung der Kapitel

Zuerst wird im Kapitel 2 einen Einblick in die Forschungsgebiete, welche diese Diplomarbeit betreffen, gegeben. Es werden dabei nicht nur allgemeine Konzepte beschrieben, sondern auch einzelne Arbeiten näher erläutert.

Das Kapitel 3 befasst sich danach mit theoretischen Ansätzen. Sowohl die Spracherkennung, das Textmining, als auch Werkzeuge zur Evaluation von Resultaten werden vorgestellt. Es spricht vor allem Leser an, die sich bis jetzt noch nicht mit den erwähnten Themen auseinander gesetzt haben.

Die Durchführung des Experiments wird schliesslich im Kapitel 4 beschrieben. Angefangen mit der Problemstellung, über die Beschaffung der Daten und die Beschreibung ihrer Verarbeitung, werden auch die erhaltenen Ergebnisse ausführlich evaluiert. Der Schluss des Kapitels befasst sich mit den Einschränkungen, denen dieses Experiment unterliegt.

Zu guter Letzt wird im Kapitel 5 versucht, die erhaltenen Ergebnisse einzuordnen und Schlüsse aus dem Experiment zu ziehen. Auch wird ein Ausblick auf mögliche Weiterentwicklungen gegeben.

2 Einordnung in die aktuelle Forschung

In diesem Kapitel werden die für diese Arbeit wichtigen Forschungsgebiete beschrieben. Weiterführende Informationen zu den jeweiligen Themen können aus den angegebenen Quellen bezogen werden.

Bei den ersten beiden Themen Sprachdialogsysteme und Telefonspamfilter handelt es sich um Anwendungen. Vor allem das erst kürzlich publizierte System von Microsoft zur Filterung von Spam bei einem Telefonanschluss, zeigt interessante Ansätze. Sie können jedoch durch die anders gelagerten Problemstellungen nicht direkt mit dieser Diplomarbeit verglichen werden.

Die nachfolgend beschriebenen Techniken und Forschungsgebiete werden für den Bau des in dieser Diplomarbeit konzipierten Systems verwendet. Konkret sind dies die Spracherkennung und das Textmining. Als ökonomischer Hintergrund wird auch eine Studie zur Unterbrechbarkeit im Arbeitsalltag vorgestellt.

2.1 Sprachdialogsysteme

Fast jedes grössere Unternehmen setzt für seinen Kundensupport telefonbasierte Systeme ein. Um die anrufenden Kunden möglichst automatisiert der richtigen Stelle zuordnen zu können, werden weitere Informationen über die Art seines Anliegens gebraucht. Vielfach werden dazu Eingaben per Telefontastatur verwendet.

Immer häufiger kommen auch Sprachdialogssysteme zum Einsatz. Sie können natürlichsprachliche Dialoge führen und die Antworten des Anrufers in Echtzeit auswerten. Dank intelligenter Dialoge kann auf das mühselige Durchnavigieren per Telefontastatur verzichtet werden. Bereits sind erste Standards wie VoiceXML¹ zur Vereinheitlichung der Programmierung solcher Systeme verabschiedet worden. Weiterentwicklungen auf diesem Gebiet führen in Richtung Dialog Design, der Konzipierung effizienter und natürlicher Dialoge, und Persona-Design, der Erschaffung virtueller Charakteren.

¹ <http://www.voicexml.org>, 25.8.2006

Als Beispiele von Sprachdialogssystemen dienen das im Jahre 2006 eingeführte Kundenportal der Swisscom¹ oder die, nur von 1996 bis 2000, eingesetzte Fahrplanauskunft der Schweizerischen Bundesbahnen.

J. Chu-Carroll und B. Carpenter stellen in [Chu-Carroll und Carpenter 1999] ein System zur automatischen Weiterleitung von Anrufen in einem Callcenter eines grossen Finanzdienstleisters in den Vereinigten Staaten vor. Auf die Frage “Wie kann ich Sie verbinden?” versucht das Systems den richtigen Bestimmungsort des Telefonanrufes zu eruieren. Falls sich kein eindeutiges Resultat ergibt, ist auch mehrmaliges Rückfragen zur Eingrenzung des Ortes möglich.

Folgend wird die Funktionsweise dieses Systems stark vereinfacht erläutert. Einige der benutzten Techniken werden im Kapitel 3.2 näher beschrieben. Vorgängig aufgezeichnete Anrufe, die den Bestimmungsorten eindeutig zuweisbar sind, werden jeweils in einem Dokument für jeden Ort zusammengefasst. Bei einem neuen Anruf wird nun das ihm ähnlichste Dokument gesucht und daraus den Bestimmungsort abgeleitet. Falls mehrere Dokumente zur Auswahl stehen, wird mit Rückfragen versucht diese auf ein einziges zu reduzieren.

Das der Arbeit [Chu-Carroll und Carpenter 1999] vorangegangene Experiment [Lee et al. 1998] umfasste 4497 Anrufe, die an 23 verschiedene Orte weitergeleitet werden konnten. 3753 Anrufe wurden zum Anlernen des Systems benutzt. Die restlichen wurden für das Testen verwendet. Rund 36% waren nicht eindeutig klassifizierbar und benötigten eine Rückfrage. Durch Rückfragen wurden 57% dieser im ersten Schritt nicht zuordbaren Anrufe richtig klassiert. Insgesamt leitete das System gut 89% der Anrufe korrekt weiter.

2.2 Telefonspamfilter

Spam bei Telefonanschlüssen ist zur Zeit, im Gegensatz zu Emails, ein eher recht selten vorkommendes Problem. Durch die immer grössere Verbreitung von Voice over IP² und damit äusserst geringen Verbindungskosten, wird die Zahl der unerwünschten Anrufe jedoch unweigerlich steigen.

Eric Horvitz, Forscher bei Microsoft, ist an der Entwicklung eines Systems, genannt V-Priorities, zur Filterung von Voice Spam [Graham-Rowe 2006]. V-Priorities analysiert Merkmale dreier verschiedener Ebenen eines Anrufs. Zum einen wird der Satzrhythmus,

¹ <http://www.swisscom-fixnet.ch/fx/privatkunden/dienste/Sprachportal/index.htm>, 25.8.2006

² <http://www.fcc.gov/voip>, 17.9.2006

die Tonhöhe und Pausenlänge einbezogen. Zum anderen werden einzelne Zielwörter erkannt. Als letztes fließen auch Metadaten wie die Uhrzeit des Anrufes oder seine Dauer in die Entscheidung ein. In einem Versuch mit 207 Anrufen auf einen Telefonbeantworter eines Mitarbeiters wurde 90% Genauigkeit bei der Unterscheidung zwischen Spam und erwünschten Anrufen erreicht. Zu 75% konnte zwischen privaten und geschäftlichen Anrufen unterschieden werden.

2.3 Spracherkennung

Auf dem Gebiet der Spracherkennung wird seit bald 50 Jahren geforscht. Der erste Durchbruch gelang IBM 1984 [Nadas 1984]. Das System konnte 5'000 englische Einzelwörter erkennen, brauchte dazu aber mehrere Minuten Rechenzeit auf einem Grossrechner. Heutzutage sind zahlreiche Systeme verschiedener Anbieter verfügbar.

Es wird zwischen sprecherunabhängiger und sprecherabhängiger Spracherkennung unterschieden. Erstere ist ohne Training des Systems von unterschiedlichen Personen nutzbar, weist jedoch nur einen geringen Wortschatz von einigen tausend Worten auf. Sprecherabhängige Systeme müssen zuerst auf den jeweiligen Benutzer eingestellt werden. Dies kostet einige Minuten Einarbeitungszeit, wird jedoch durch eine höhere Genauigkeit und einen grösseren Wortschatz belohnt.

Die Hauptprobleme der Spracherkennung liegen in der Kontinuität und der Vielfalt der Sprache. Das Erkennen von Wortanfang und Wortende, so wie der einzelnen Laute, ist für einen Computer äusserst schwierig. Auch stellen die vielen Wortformen und die unterschiedliche Schreibweise ähnlich klingender Worte ein Problem dar. Dazu kommen Störgeräusche aus der Umwelt und technische Einschränkungen, wie der beschnittene Frequenzbereich der Signalübertragung beim Telefon dazu.

2.4 Textmining

Textmining ist ein Teilgebiet des Data Mining [Witten und Frank 2005]. Mit Hilfe von Data Mining wird versucht in grossen Datenmengen sich wiederholende Muster zu finden. Warenhausketten erforschen beispielsweise mit anhand von Kundenkarten gesammelten Daten das Einkaufsverhalten ihrer Kunden und optimieren so die Werbung oder steuern Aktionsangebote¹.

¹ <http://www.golem.de/0208/21218.html>, 18.9.2006

Während es beim Data Mining um das Analysieren von Datenreihen jeglicher Art geht, befasst sich Textmining, wie aus dem Namen bereits ersichtlich ist, speziell mit Texten [Witten 2004]. Einzuordnen ist dieses Gebiet in die Forschungsbereiche Computerlinguistik und maschinelles Lernen. Die wohl bekannteste Anwendung des Textminings ist das Herausfiltern von nicht erwünschten Email-Nachrichten, dem Spam. Aufgrund von bereits eingegangenen und bewerteten Nachrichten versucht das System Regeln herzuleiten, um neue Nachrichten möglichst gut filtern zu können.

In [Zelikovitz und Hirsh 2000] wird eine Methode zur Verbesserung der Klassifikation von Texten vorgestellt. Dies geschieht mittels einer grossen Zahl unklassierter Dokumente – dem sogenannten Hintergrundwissen. Statt die zu klassierenden Dokumente direkt mit jenen im Trainingsdatensatz zu vergleichen, wird das Hintergrundwissen dazwischen geschaltet. Dieses Verfahren erzielt vor allem bei kleinen Trainingsdatensätzen oder sehr kurzen Texten erstaunliche Verbesserungen.

2.5 Unterbrechbarkeit

Eine häufige Unterbrechung des Arbeitsflusses kann unangenehme ökonomische Folgen haben. Genaue Zahlen aus der Wirtschaft sind dazu noch nicht erhoben worden. In Unternehmen wird jedoch grossen Wert auf eine produktive Gestaltung der Arbeitsprozesse und vor allem auch des Arbeitsumfeldes gelegt. Der Arbeitnehmer soll möglichst ungestört seiner Arbeit nachgehen können.

Anhand der Beobachtung von 24 Informatikern während gut 700 Stunden ihrer Arbeit versucht die Studie [Mark et al. 2005], die Beschaffenheiten der Fragmentierung der Arbeit genauer zu bestimmen. Für die Berechnung dieser Fragmentierung sind für sie die Dauer der vorhergehenden Arbeitsphase und die Häufigkeit der Unterbrechungen relevant. Insgesamt werden 57% der Arbeitsphasen unterbrochen. Die Gründe dafür können sowohl externer als auch interner Natur sein. Nicht jeder Unterbruch ist jedoch mit einem negativen Einfluss verbunden. Hat die Unterbrechung einen Zusammenhang mit dem aktuellen Arbeitsthema, kann sie durchaus auch die Produktivität steigern. Erstaunlicherweise können Personen in einem Grossraumbüro länger an einem Stück arbeiten, werden jedoch häufiger gestört. Im Gegensatz dazu dauern die Unterbrechungen bei einzeln arbeitenden Personen länger.

Ein Forschungsprojekt des MIT [Marti und Schmandt 2005] befasst sich mit einer speziellen Möglichkeit, um einen Telefonanruf möglichst störungsfrei in den Arbeitsalltag zu integrieren. Statt des üblichen Klingelns nimmt ein mit Sensoren und Motoren ausgestattetes Stofftierchen, dem Squirell¹, Kontakt mit dem Angerufenen auf. Dazu werden der Natur entnommene und deshalb für den Menschen gewohnte Bewegungsmuster verwendet. Es nimmt in der Nähe stattfindende Konversationen wahr und versucht die Wichtigkeit des Anrufes anhand früherer Ereignisse abzuschätzen. Aus den gewonnenen Erkenntnissen versucht es, die für den Angerufenen angenehmste Möglichkeit zur Abwicklung des Anrufes zu finden. Es kann beispielsweise den Anrufenden in ein Gespräch verwickeln, ihm Informationen geben oder ihn auffordern eine Nachricht zu hinterlassen. Der Angerufene kann durch Interaktion, z.B. Berührung eines Ohrs des Stofftierchens, diese Tätigkeiten beeinflussen.

¹ <http://web.media.mit.edu/~stefanm/phd/cellularsquirrel>, 20.9.2006

3 Grundlagen

Dieses Kapitel widmet sich den Theorien und Mechanismen, die in dieser Diplomarbeit verwendet werden. Es spricht vor allem Leser an, die sich bis jetzt noch nicht mit Spracherkennung oder Textmining befasst haben. Als Literatur ist für die Spracherkennung [Cole et al. 1997] und für das Textmining [Witten und Frank 2005] [Witten 2004] zu empfehlen.

3.1 Spracherkennung

Wie im Kapitel 2.3 bereits erwähnt, ist die Spracherkennung ein äusserst schwieriger Prozess, der erst seit wenigen Jahren ansatzweise beherrscht wird. Grundsätzlich besteht eine Spracherkennung aus zwei Schritten: Der Vorverarbeitung und der eigentlichen Erkennung. Diese beiden Prozesse werden in der Abbildung 3.1 dargestellt und in den folgenden zwei Kapiteln näher vorgestellt.

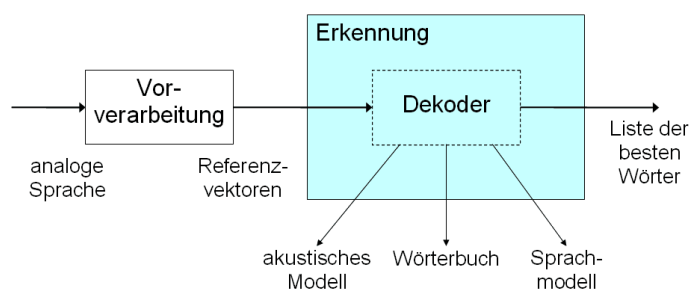


Abbildung 3.1: Spracherkennung I
[<http://de.wikipedia.org/wiki/Spracherkennung>, 11.09.2006]

3.1.1 Vorverarbeitung

Erst durch eine gute Vorverarbeitung des Sprachsignals wird die Erkennung gesprochener Sprache für den Computer überhaupt möglich. Es wird dabei versucht, eine Vielzahl an redundanter und irrelevanter Information herauszufiltern und die Daten möglichst kompakt zu beschreiben.

Das Sprachsignal muss zuerst abgetastet und quantisiert, das heisst von der analogen Form in eine digitale übersetzt werden. In einem nächsten Schritt werden die Passagen des Sprechens von jenen des Schweigens getrennt. Andernfalls versucht die Erkennung auch inhaltslose Abschnitte in Text umzuwandeln. Vor allem bei Aufnahmen in lauter Umgebung mit zum Beispiel gleichzeitig laufenden Gesprächen ist dies besonders wichtig

und auch schwierig. Danach wird das nun digitalisierte Signal durch einen Filter geglättet. Dadurch wird das Signal leistungsmässig gleichmässiger und ist für die Weiterverarbeitung einfacher zu handhaben.

Für die Erkennung werden die relevanten Merkmale extrahiert und als Vektoren dargestellt. Ein einzelner Vektor stellt dabei noch keinen Laut oder ein Wort dar, sondern ist nur ein sehr kleiner Zeitabschnitt einer Aufnahme.

3.1.2 Erkennung

Für die Erkennung des Textes braucht es verschiedene Methoden bis aus den Merkmalsvektoren Wörter und ganze Sätze entstehen. Zuerst müssen aus den einzelnen Zuständen Phoneme gebildet werden. Phoneme sind die kleinsten bedeutungsunterscheidenden, aber nicht bedeutungstragenden Einheiten einer Sprache. Es stehen verschiedene Methoden für die Bildung von Phonemen zur Verfügung. Die bekannteste ist die Verwendung von Hidden-Markov-Modellen [Rabiner 1989]. Der Grund ist vor allem ihre grosse Flexibilität und die effiziente Implementierbarkeit. Andere Möglichkeiten sind Dynamic-Time-Wrapping [Bouvard und Dupond 1996] oder Time-Delay-Neural-Networks [Waibel et al. 1989]. Mit Hilfe von Wörterbüchern und Sprachmodellen können aus den aneinander gereihten Lauten Wörter und Sätze erkannt werden.

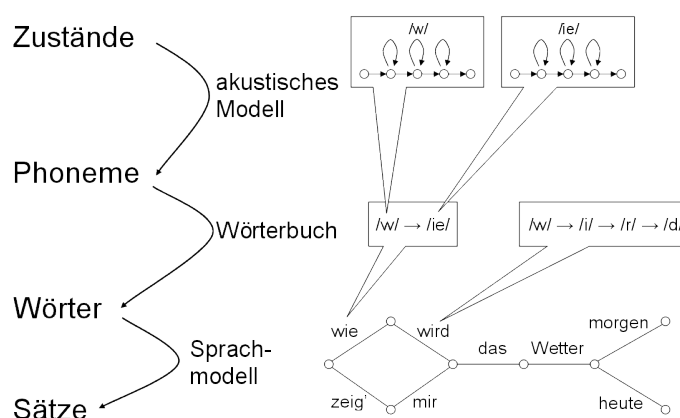


Abbildung 3.2: Spracherkennung II
 [http://de.wikipedia.org/wiki/Spracherkennung, 11.09.2006]

3.2 Textmining

Anders als beim Data Mining sind beim Textmining die zu findenden Informationen für den Menschen grundsätzlich ersichtlich [Witten und Frank 2005]. Die meisten Verfasser von Dokumenten tragen ausreichend Sorge zur Verständlichkeit ihrer Texte. Leider genügt dies nicht, um den Inhalt einem Computer zugänglich zu machen. Ihm fehlt das allgemeine Wissen über die Sprache und er besitzt auch kein Kontextverständnis. Beim Textmining werden für die digitale Verarbeitung und Auswertung statistische und linguistische Werkzeuge gebraucht.

In dieser Diplomarbeit wird aus dem Gebiet Textmining hauptsächlich die Textklassifikation verwendet. Sie wird deshalb in den folgenden Kapiteln näher beschrieben. Man versucht dabei die gegebenen Texte in vorgegebene Klassen einzuteilen.

3.2.1 Vorgehensweise

Bei der Textklassifikation werden, wie bei fast allen maschinellen Lernverfahren, zwei Phasen unterschieden: Die Lernphase und die Klassifikationsphase. Diese Phasen setzen sich wiederum aus meist gleichen Teilschritten zusammen.

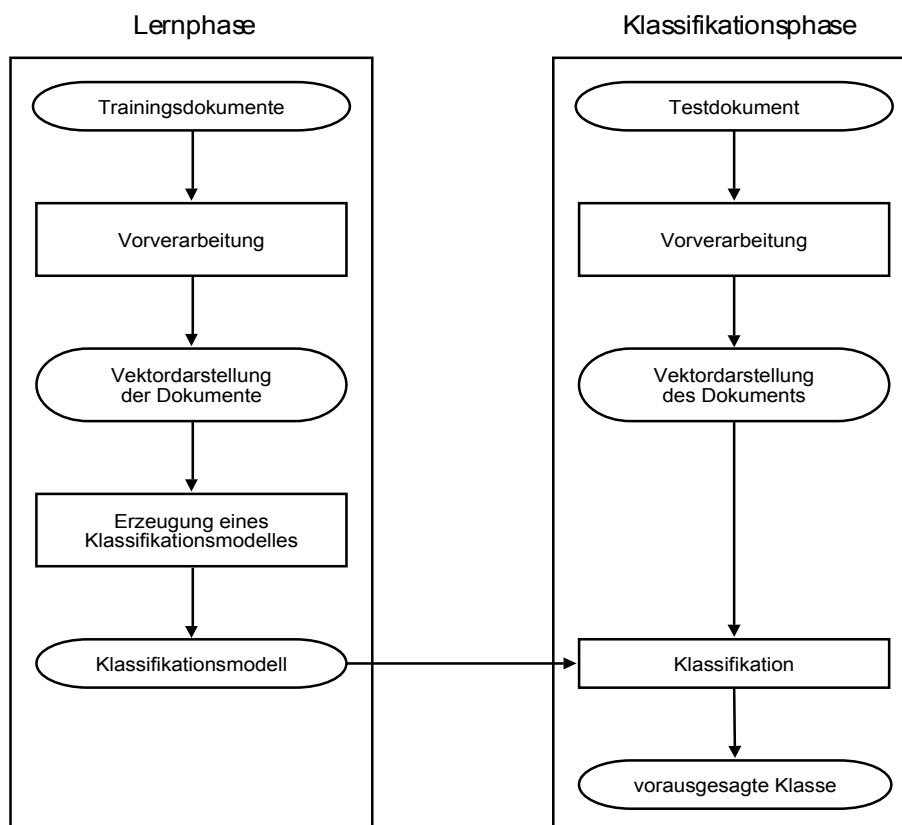


Abbildung 3.3: Phasenmodell einer Textklassifikation

Lernphase

Während der Lernphase wird mit Hilfe eines sogenannten Trainingsdatensatzes von Textdokumenten ein Klassifikationsmodell gebildet. Die Klassenzugehörigkeit der Trainingsdaten muss jeweils bekannt sein, damit das System überhaupt etwas lernen kann.

Klassifikationphase

Das in der Lernphase gebildete Klassifikationsmodell dient dann in der Klassifikationsphase zur Voraussage der Klasse bei neuen Dokumenten. Dies sollte im Normalfall in Echtzeit geschehen und verlangt deshalb eine kluge Architektur des Systems.

3.2.2 Vorverarbeitung

Um die sich vielfach stark unterscheidenden Texte überhaupt vergleichen zu können, müssen sie mit speziellen Techniken vereinfacht und abstrahiert werden. Was für das menschliche Hirn eine selbstverständliche Angelegenheit ist, stellt für den Computer, der ohne angeborenes linguistisches und kontextsensitives Verständnis auskommen muss, eine grosse Hürde dar. Dennoch sind durch eine geschickte Aufteilung des Problems in verschiedene Teilbereiche äusserst akzeptable Resultate erzielbar.

Zur Illustration der einzelnen Schritte findet sich als Beispiel im Anhang A die gesamte Vorverarbeitung dreier Mustersätze.

Wortauswahl

Nicht jedes Wort eines Dokumentes trägt den gleichen Anteil zum Informationsgehalt für die Klassifikation bei. Wörter von geringem Interesse nennt man Stoppwörter. Sie werden vor der weiteren Verarbeitung aus einem Dokument entfernt. So können beispielsweise die Artikel (z.B. *der* oder *eine*) problemlos entfernt werden, ohne dass sich dies auf den Inhalt eines Dokumentes auswirkt.

Es sind verschiedene Verfahren für die Stoppwortfilterung bekannt, die beliebig kombiniert werden können. Hier eine Aufzählung der Gebräuchlichsten:

- **Stoppwortliste:** Alle Wörter, die in einer vorher definierten Liste vorkommen, werden aus dem Dokument entfernt.
- **Anzahl Buchstaben:** Falls ein Wort eine gewisse Länge unterschreitet, wird es entfernt.

- **Wortfrequenzanalyse:** Zuerst wird die Häufigkeit der Wörter in einem Dokument ermittelt. Danach können zwei Kategorien von Wörtern gelöscht werden. Jene, die in einem Dokument oft vorkommen, weisen meist einen sehr geringen Informationsgehalt auf. Auch sehr seltene sind für eine Klassifikation schlecht zu gebrauchen, da sie mit sehr geringer Wahrscheinlichkeit wieder auftauchen werden.

Normalisierung

Eine weitere Verbesserung für die Klassifizierung eines Dokumentes kann durch Eliminierung der verschiedenen Flexionsformen der Wörter erreicht werden. Obwohl die zwei Ausdrücke *gehst* und *geht* für uns Menschen inhaltlich einen sehr ähnlichen Sinn haben, sind sie für den Computer verschieden, da die Zeichenfolge nicht dieselbe ist.

Um dieses Problem zu lösen gibt es mehrere Methoden. Zwei verschieden aufwändige werden folgend vorgestellt. Der morphologische Ansatz ist gegenüber dem Stemming um ein Vielfaches rechen- und speicherintensiver.

- **Morphologie:** Es wird versucht durch Analyse eines ganzen Satzes die darin enthaltenen Wörter auf ihre Grundformen zurückzuführen. Dazu wird ein ausführliches Modell der verwendeten Sprache benötigt. Die oben erwähnten Wörter werden durch diese Methode auf ihre gemeinsame Grundform *gehen* transformiert.
- **Stemming:** Anhand von meist wenigen Regeln werden beim Stemming die Wörter in eine einfachere Form gebracht [Lovins 1968] [Snowball 2006]. Dabei werden die gängigen Präfixe und Suffixe entfernt. Je nach Sprache ist das Resultat unterschiedlich gut. Zum Beispiel schneidet das Deutsche wegen seiner komplizierteren Wortbildung gegenüber dem Englischen klar schlechter ab. Aus den Wörtern *gehst* und *geht* entsteht durch Stemming das Fragment *geh*, aus *gegangen* jedoch nur *gang*.

Generalisierung

Bei einer Generalisierung wird versucht die Bedeutung der Wörter zu erfassen und sie möglichst universell zu beschreiben. Dazu ist die Analyse des Kontexts ihres Vorkommens zwingend notwendig. Sowohl Synonyme wie auch Überbegriffe können für eine Abstrahierung des Dokumentes verwendet werden. Das Wort *Auto* wird beispielsweise durch *Wagen* oder *Fahrzeug* ergänzt oder ersetzt.

Übersetzung

Falls Dokumente unterschiedlicher Sprachen im Datensatz vorhanden sind, könnte eine Übersetzung in eine einzige Sprache sinnvoll sein. Dies vergrößert das Wissen, über das gelernt wird, je nach Vorkommen und Verteilung der Sprachen immens. Wichtig ist jedoch die semantische Richtigkeit, da der verfälschte Inhalt sonst zu Fehlern in der Klassifikation führt. An Systemen für eine korrekte, automatische Übersetzung wird derzeit immer noch mit grossem Aufwand geforscht [Willée et al. 2002].

3.2.3 Vektordarstellung

Für die Weiterverarbeitung müssen die vorverarbeiteten Dokumente in eine geeignete Form gebracht werden. Da einige Klassifikationsalgorithmen nur mit Zahlen und nicht mit Zeichenketten umgehen können, ist die Vektordarstellung für diesen Zweck am besten geeignet [Salton et al. 1975].

Für die Bildung des Vektorraums werden alle in den Dokumenten vorkommenden Worte verwendet. Jedes Wort entspricht dabei einer Achse und gemeinsam spannen sie einen mehrdimensionalen Raum auf.

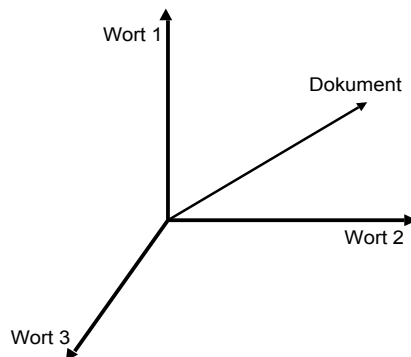


Abbildung 3.4: Vektorraum

Die einzelnen Dokumente können auf mehrere Arten in einen Vektor umgewandelt werden. Eine gebräuchliche und bewährte Methode ist das TF-IDF-Verfahren [Salton et al. 1975]. Es benutzt für die Berechnung eines Vektorwertes folgende Werte:

- **Termfrequenz (TF):** Relatives Vorkommen eines Wortes innerhalb eines Dokumentes.
- **Invertierte Dokumentfrequenz (IDF):** Anzahl aller Dokumente dividiert durch die Anzahl der Dokumente, in denen das betreffende Wort vorkommt.

- **TF-IDF-Wert:** $x = TF \cdot IDF$

Das TF-IDF-Verfahren gewichtet Worte, die in wenigen Dokumenten oft vorkommen, höher als solche, die in vielen Dokumenten wenig erscheinen. Den Sinn dahinter erfasst man bei der Betrachtung eines Extremums schnell. Falls zum Beispiel ein Wort in allen Dokumenten enthalten ist, können die Dokumente nicht anhand dieses Wortes unterschieden werden. Deshalb ist es für eine Klassifikation nutzlos. Richtigerweise ergibt die Gewichtung für dieses Wort nach TF-IDF-Verfahren den Wert null.

3.2.4 Klassifikationsmodelle

Die Forschung im Bereich Data Mining hat viele verschiedene Klassifikationsalgorithmen hervorgebracht. Nur ein Teil davon kann jedoch beim Textmining eingesetzt werden. Ein Klassifikationsalgorithmus wird gebraucht, um aus den vorhandenen Trainingsdokumenten ein Klassifikationsmodell zu erstellen und danach die Klasse des Testdokumentes vorauszusagen.

K-Nächste-Nachbarn

Der K-Nächste-Nachbarn-Algorithmus (engl. k-nearest neighbor) ermittelt die voraussichtliche Klasse eines Dokuments anhand seiner Ähnlichkeit zu bereits bekannten Dokumenten im Trainingsdatensatz [Cover und Hart 1967]. Dazu muss die Anzahl k der für die Klassifikation herangezogenen Dokumente und das Ähnlichkeitsmass festgelegt werden. Im Falle eines Vektorraums kann die Ähnlichkeit entweder über das Skalarprodukt oder den Kosinuswinkel zweier Dokumente bestimmt werden. Im Gegensatz zum Kosinuswinkel wird beim Skalarprodukt auch die Länge der Dokumente in die Berechnung einbezogen.

Naive Bayes-Klassifikator

Die meisten Spamfilter basieren auf dem vor über 250 Jahren aufgestellten Satz von Bayes des gleichnamigen englischen Mathematikers Thomas Bayes [Bayes 1763] [McCallum und Nigam 1998]. Die Klassifikation beruht auf dem Prinzip, dass wenn ein Ereignis eingetroffen ist, daraus die Wahrscheinlichkeit eines anderen abgeleitet werden kann. Dazu trifft Bayes zwei starke Annahmen: Erstens sind die auftretenden Ereignisse untereinander unabhängig, zweitens ist ihr Vorkommen normalverteilt. Obwohl diese Annahmen in der Praxis nur sehr selten zutreffen, funktioniert die naive Bayes-Klassifikation bei vielen Problemen äusserst gut.

3.3.2 Confusion Matrix

In einer Confusion Matrix [Witten und Frank 2005] werden die Ergebnisse einer Klassifikation eingetragen. Anhand eines Beispiels mit den beiden Klassen *positiv* und *negativ* wird folgend die Funktionsweise erklärt. Eine Confusion Matrix kann auch für die Darstellung von Problemen mit mehreren Klassen eingesetzt werden.

Falls ein Dokument in der Realität der Klasse *positiv* zugeteilt ist und auch von der Klassifikation als *positiv* vorausgesagt wird, wird das Resultat als *richtig positiv* bezeichnet. Falls ein *positives* Dokument fälschlicherweise als *negativ* klassiert wird, fällt dieses Resultat unter die Kategorie *falsch positiv*. Analog dazu verläuft die Bezeichnung für die Voraussage der Klasse *negativ*.

		Voraussage	
		positiv	negativ
Realität	positiv	richtig positiv (RP)	falsch positiv (FP)
	negativ	falsch negativ (FN)	richtig negativ (RN)

Tabelle 3.1: Confusion Matrix

Aus den durch mehrere Klassifikationen gewonnenen Daten können nun verschiedene Kennzahlen abgeleitet werden. Die wichtigsten sind Genauigkeit, Precision und Recall.

$$\text{Genauigkeit} = \frac{(RP + RN)}{(RP + FP + FN + RN)}$$

$$\text{Precision} = \frac{RP}{(RP + FP)}$$

$$\text{Recall} = \frac{RP}{(RP + FN)}$$

4 Experiment

Wie bereits im Kapitel 1.1 beschrieben, wird in dieser Diplomarbeit versucht ein System zu entwickeln, das die Klasse von Telefonanrufen vorhersagt. Dieses Kapitel widmet sich der Entwicklung und Evaluation eines solchen Systems durch ein Experiment.

4.1 Problemstellung

Das Anwendungsgebiet ist der Telefonanschluss eines Professors (Subjekt) des Instituts für Informatik der Universität Zürich. Berufsbedingt sind die Thematiken der ankommenden Anrufe sehr vielfältig. Sie reichen von Anrufen des Sekretariates, über solche von befreundeten Professoren bis zu Unternehmen, die an einer Zusammenarbeit interessiert sind. Nicht jeder Anruf wird dabei vom Angerufenen als gleich wichtig eingestuft. Sobald die Wichtigkeit eines Anrufs in Bezug zu seiner Unterbrechbarkeit zu gering ist, sollte dies erkannt und der entsprechende Anruf abgewiesen werden, damit die Zielperson ungestört weiterarbeiten kann.

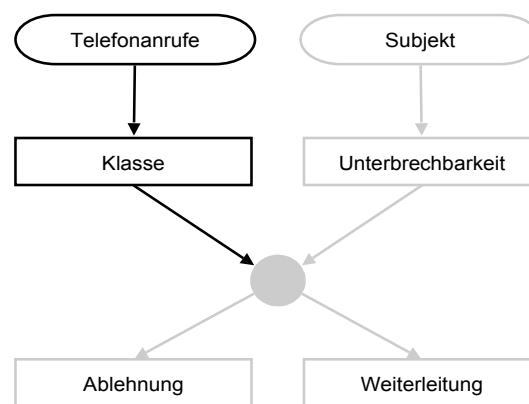


Abbildung 4.1: Problemstellung

Alleine durch die Identität eines Anrufers kann die Wichtigkeit, respektive Klasse, nicht zuverlässig vorausgesagt werden. Zum Beispiel kann die Mitteilung des Sekretariates über die unvorhergesehene Absage des nächsten Termins bedeutsamer sein als ihre wiederholte Frage nach fehlenden Quittungen für die anstehende Spesenabrechnung.

Im Gegensatz zur im Kapitel 2.1 vorgestellten Arbeit [Chu-Carroll und Carpenter 1999] kann auch nicht direkt anhand des Inhaltes auf die Klasse geschlossen werden. Falls ein potentieller Sponsor wegen einer Terminvereinbarung anruft, wird dies normalerweise um

einiges höher gewichtet, als wenn ein Student einen Termin für eine Besprechung vereinbaren will.

Aus Gründen des Persönlichkeitsschutzes werden Anrufe aus dem privaten Umfeld der Versuchsperson nicht in das Experiment miteinbezogen.

4.2 Übersicht

Die Voraussage der Wichtigkeit eines Anrufes ist ein vielschichtiges Problem. Um es möglichst gut zu lösen, bedarf es mehrerer gut aufeinander abgestimmter Schritte. Folgend wird der Aufbau des Experiments beschrieben.

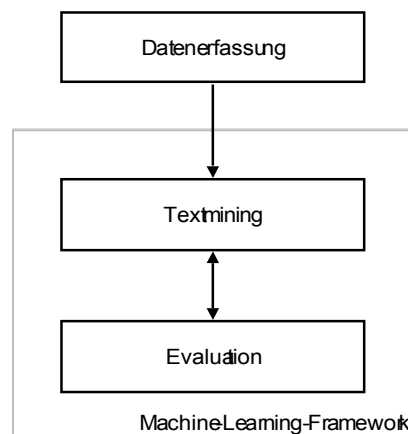


Abbildung 4.2: Experiment

In einem ersten Schritt werden die Anrufe und ihre Klassierung in fünf Wichtigkeitsstufen erfasst. Danach wird das Machine-Learning-Framework vorgestellt, das für die automatische Auswertung der Anrufe zuständig ist. Auch werden die verwendeten Fremdsysteme kurz erläutert. Im Kapitel 4.6 wird auf die praktische Umsetzung der im vorangegangenen Kapitel 3 erarbeiteten und im Machine-Learning-Framework eingesetzten Techniken eingegangen. Insbesondere werden auch zwei neue Methoden – Klassifikation mit Hintergrundwissen und Klassifikation mit Rückfragen – vorgestellt, die die Klassifikation verbessern sollen. Es folgt eine ausführliche Auswertung des Datensatzes und die Evaluation verschiedener Einstellungen im Textminingprozess. Zum Schluss wird auf die Einschränkungen dieses Experiments eingegangen.

4.3 Datenerfassung

In diesem Kapitel wird der Weg der Datenerfassung für das später folgende Textmining beschrieben. Da aus Datenschutz- und Zeitgründen nicht auf reale Anrufe zurückgegriffen werden kann, müssen die potentiell auftretenden Anrufe simuliert werden.

4.3.1 Interview

Durch mehrere Gespräche mit dem direkt betroffenen Professor und Personen aus seinem Arbeitsumfeld wird versucht, die Themen der ankommenden Anrufe und ihre proportionale Verteilung zu erfassen. Um die Genauigkeit der Schätzung zu erhöhen werden Anrufergruppen und Untergruppen gebildet. Die folgende Tabelle zeigt eine vereinfachte Darstellung dieser Gruppen und ihrer Verteilung. Die komplette Aufstellung befindet sich in Anhang B.

Gruppe	Verteilung	Untergruppe	Verteilung
Institut für Informatik	50%	Sekretariat	33%
		Professoren und Assistenten	12%
		Techniker und Studenten	5%
Erweitertes Arbeitsumfeld	45%	Verbände und Konferenzen	16%
		Universität Zürich	11%
		Professoren und Nationalfonds	11%
		Unternehmen	7%
Spam	5%		5%
	100%		100%

Tabelle 4.1: Verteilung der Anrufergruppen

Jede Untergruppe enthält bis zu 11 verschiedene Themengebiete, die versuchen die gesamte Bandbreite der auftretenden Anrufe möglichst gut abzudecken. Auch sie sind im oben genannten Anhang zu finden. Das zweimalige Vorkommen der Untergruppe *Professoren* führt auf die Unterscheidung zwischen Professoren des Instituts für Informatik und auswärtigen Professoren zurück. Als Beispiel für Themengebiete sind jene der Untergruppe *Universität Zürich* erwähnt: *Prorektorate*, *Unitecra* und *Verwaltung*.

4.3.2 Simulation der Anrufe

Um für das Experiment eine genügend grosse Stichprobe an Anrufen zu erhalten, werden als Zielgrösse 200 Anrufe festgelegt. Dadurch ergeben sich pro Themengebiet mindest zwei bis maximal acht Anrufe.

Für die Durchführung der Anrufe werden 13 Personen sorgfältig ausgewählt und instruiert. Sie kommen alle aus dem engeren Arbeitsumfeld des Professors und haben somit ein Grundwissen über seine Tätigkeiten. Dies soll die Qualität der Anrufe erheblich steigern.

Die Testanrufer erhalten einen Briefumschlag mit einer Anweisung und 10 bis 25 Themenkarten. Die Anweisung ist in Anhang C zu finden, eine Themenkarte folgend abgebildet.

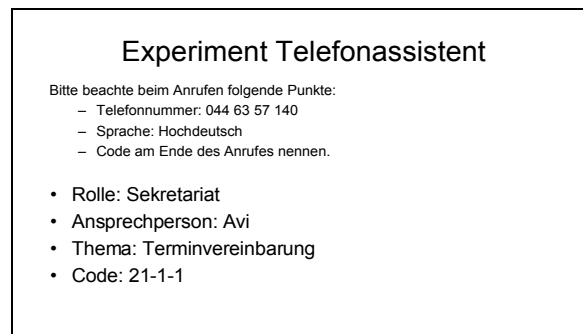


Abbildung 4.3: Themenkarte

Durch die knapp gehaltenen Informationen auf der Karte (Rolle, Ansprechperson und Thema) sind die Anrufer gezwungen die Situation weiter auszuschnüffeln. Dadurch steigt die Vielfalt der Anrufe. Der Code am Ende jedes Anrufes dient zur Identifizierung der Gruppe, der Untergruppe und des Themas und ist für die spätere Auswertung nützlich.

4.3.3 Klassierung

Nach dem Aufnehmen der Anrufe auf einem Telefonbeantworter wird der Inhalt durch die im nächsten Kapitel vorgestellte Spracherkennung in geschriebenen Text konvertiert. Anhand dieser Textfragmente legt der Professor die jeweilige Wichtigkeit der Anrufe fest. Die Wichtigkeit ist analog zur Unterbrechbarkeit im Experiment [Vorburger et al. 2005], in fünf Klassen unterteilt. Die Klassen sind in der Tabelle 4.2 dargestellt.

Klasse	Beschreibung
1	unerwünscht
2	unwichtig
3	neutral
4	wichtig
5	sehr wichtig

Tabelle 4.2: Klassen

4.4 Spracherkennung

Eine gute Spracherkennung ist für die Lösung der Problemstellung unabdingbar. Ganz nach dem Sprichwort “Garbage in – Garbage out” verläuft auch ein Textminingprozess unbefriedigend, wenn das Ausgangsmaterial nicht optimal ist.

Nach intensivem Suchen und Testen, der über das Internet erhältlichen Spracherkennungstools, ist kein Taugliches gefunden worden. Entweder sind die angebotenen Produkte nicht sprecherunabhängig oder die deutsche Sprache steht für Testzwecke nicht zur Verfügung. Da aber bereits existierende Sprachdialogsysteme zeigen, dass eine genügend hohe Erkennungsrate mit der heutigen Technik möglich ist, ist die Suche nach einer befriedigenden Software eingestellt worden.

Für dieses Experiment werden die Anrufe deshalb von Hand abgeschrieben und durch ein Rechtschreibprogramm korrigiert. Das dadurch erhaltene Resultat kann nun für das Textmining verwendet werden.

Im selben Schritt wie das Abschreiben werden auch gewisse Fehler seitens der Testanrufer eliminiert. So wird zum Beispiel ein in Schweizer Mundart gesprochener Anruf ins Hochdeutsche übersetzt. Auch werden unnatürlich oft vorkommende Personenamen angepasst (z.B. *Hans Muster*). Eine Aufstellung aller Korrekturen ist im Anhang D zu finden.

4.5 Machine-Learning-Framework

Das Machine-Learning-Framework dient der automatisierten Klassifikation der in Textform vorliegenden Telefonanrufe. Es vereint verschiedene bereits bestehende Systeme mit selbst entwickelten Algorithmen, wie beispielsweise dem Rückfragemechanismus. Daneben sind auch Statistik- und Aufzeichnungsfunktionen implementiert, die für die spätere Evaluation notwendig sind.

4.5.1 Klassendiagramm

In der Abbildung 4.4 ist das Klassendiagramm der für die Klassifikation und ihre Auswertung wichtigen Java-Klassen dargestellt.

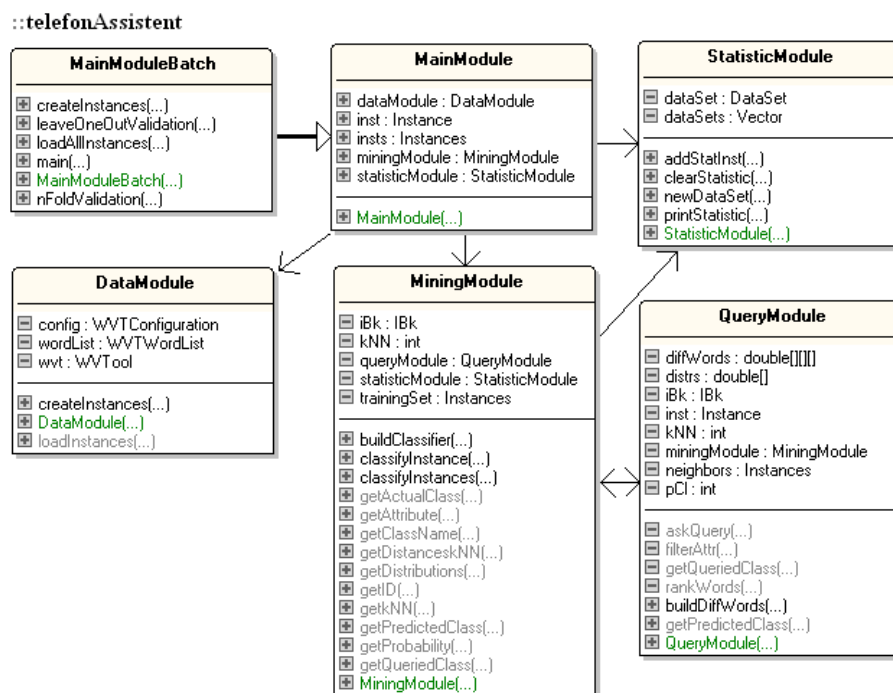


Abbildung 4.4: Klassendiagramm des Machine-Learning-Framework

Folgend werden ihre Funktionsweisen kurz vorgestellt:

- **MainModuleBatch:** Diese Klasse dient zum Starten der Applikation. Sie ist für die automatische Auswertung bereits vorhandener Datensätze mittels Methoden wie der Kreuzvalidation ausgelegt und eignet sich nicht für ein Liveexperiment.
- **DataModule:** Die Klasse DataModule bereitet die einzelnen Daten mit Hilfe des im Kapitel 4.5.3 vorgestellten WordVectorTool auf.
- **MiningModule:** In dieser Klasse wird die eigentliche Klassifikation durchgeführt. Es werden dazu Teile aus der im Kapitel 4.5.4 vorgestellten Klassenbibliothek Weka verwendet.
- **QueryModule:** Diese Klasse ist eng mit dem MiningModule verbunden. Sie wird für die Klassifikation mit Rückfragen gebraucht.
- **StatisticModule:** Das StatisticModule zeichnet die gewonnen Daten auf und führt bereits erste Auswertungen durch. Die am Ende erstellte Datei kann in einem Tabellenkalkulationsprogramm weiterverarbeitet werden.

4.5.2 Morphologie

Für die morphologische Analyse werden zwei Produkte verwendet und kombiniert. Das kommerziell verfügbare GERTWOL [Haapalainen und Majorin 1994] berechnet aus den flektierten Wortformen alle möglichen Grundformen unabhängig von deren Wortarten. Der TreeTagger [Schmid 1995] berechnet mit probabilistischen Methoden (Entscheidungsbäume) die wahrscheinlichste Sequenz von Wortartentags im Satz, indem er mit Hilfe eines Lexikons von bekannten Wortformen und ihren möglichen Wortartentags [Schiller et al. 1999], sowie empirischen Daten über die Verteilung von N-Grammen von Wortartentags für jedes Wort im Satz die wahrscheinlichste Wortart im Kontext bestimmt. Aus den teilweise mehrdeutigen Ergebnissen von GERTWOL werden dann nur diejenigen mit der übereinstimmenden Wortart selektiert. Als Resultat werden neben den Grundformen der Wörter auch die dazugehörigen Wortartentags zurückgegeben.

4.5.3 WordVectorTool

Mit dem WordVectorTool¹ kann die im Kapitel 3.2.3 vorgestellte Vektordarstellung der Dokumente erzeugt werden. Es wird am Lehrstuhl für künstliche Intelligenz der Universität Dortmund entwickelt. Verschiedene Möglichkeiten zur Erstellungen der Vektordarstellung sind möglich. Auch ist ein Stemmingalgorithmus für die deutsche Sprache enthalten.

4.5.4 Weka

Weka² ist eine freiverfügbare Sammlung von Data Mining-Algorithmen. Es wird in [Witten und Frank 2005] näher vorgestellt. Mit Hilfe von ihm wird die Klassifikation der Dokumente durchgeführt. Auch Methoden wie das Kreuzvalidierungsverfahren sind darin integriert.

1 <http://www-ai.cs.uni-dortmund.de/SOFTWARE/WVTOOL/>, 27.08.2006

2 <http://www.cs.waikato.ac.nz/~ml/weka/>, 27.08.2006

4.6 Textmining

In diesem Kapitel wird die Verarbeitung der gewonnenen Daten des Experiments mit Hilfe des Machine-Learning-Frameworks genauer vorgestellt. Insbesondere wird Einblick in die neu entwickelten Methoden gegeben.

Zuerst wird die Funktionsweise der Vorverarbeitung der Daten erläutert, die bei den drei nachfolgenden Klassifikationsmethoden gleich bleibt. Dies sind die einfache Klassifikation, die Klassifikation mit Hintergrundwissen und die Klassifikation mit Rückfragen.

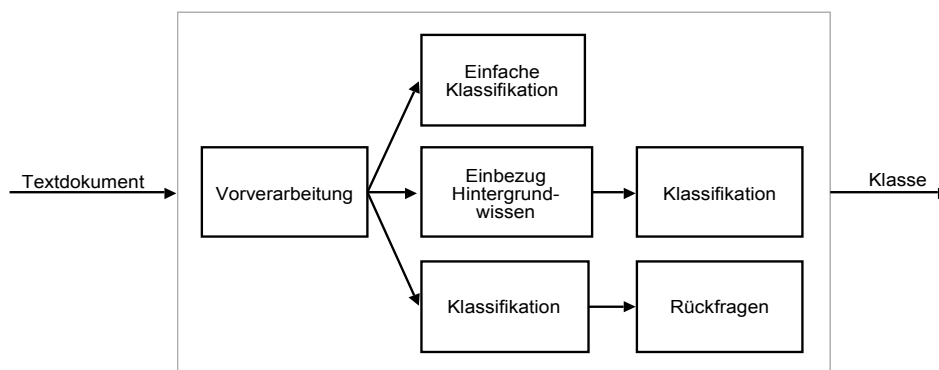


Abbildung 4.5: Methoden der Klassifikation

4.6.1 Vorverarbeitung

Die Vorverarbeitung wird mit den im vorangegangenen Kapitel vorgestellten Werkzeugen durchgeführt. Die in Fliesstext vorliegenden Dokumente werden in eine Vektordarstellung umgewandelt.

Die Sätze werden als ganzes analysiert und die enthaltenen Worte auf ihre Grundform zurückgeführt. Dies entspricht der im Kapitel 3.2.2 vorgestellten morphologischen Normalisierung. Danach werden die Wortarten Nomen und Hauptverben gefiltert und für die Weiterverarbeitung bereitgestellt. Sie besitzen laut [Yang und Wilbur 1996] den grössten Informationsgehalt in Texten.

Die gewonnenen Daten werden nun mit dem TF-IDF-Verfahren in eine Vektordarstellung gebracht. Vorgängig werden Worte, die nur einmal über alle Dokumente gesehen vorkommen, auf Grund ihrer Nutzlosigkeit bei einer Klassifikation gelöscht¹. Dies verkleinert den Vektorraum und beschleunigt die spätere Klassifikation.

¹ Kapitel 3.2.2

4.6.2 Einfache Klassifikation

Für die Klassifikation wird grundsätzlich der K-Nächste-Nachbarn-Algorithmus¹ aus Weka² verwendet. Er eignet sich sehr gut für die Textklassifikation. Auch ist mit ihm die spätere Implementierung der Klassifikation mit Hintergrundwissen und der Klassifikation mit Rückfragen einfach möglich.

Aufgrund der recht kleinen Datenmenge wird ein Spezialfall des Kreuzvalidierungsverfahrens angewendet. Für die Bildung des Trainingsdatensatzes werden alle Dokumente mit Ausnahme des zu klassierenden verwendet. Diese Methode ist unter dem englischen Begriff "Leave-one-out Cross-validation" bekannt.

4.6.3 Klassifikation mit Hintergrundwissen

Ein Arbeit [Zelikovitz und Hirsh 2000] zu Textmining mit Hintergrundwissen ist bereits im Kapitel 2.4 vorgestellt worden.

Das Hintergrundwissen besteht in diesem Experiment aus gesammelten Emails der Testperson. Sie decken sich ziemlich gut mit den Themengebieten der Anrufe. In den meisten Fällen sind ihre Inhalte um einiges länger und besitzen deshalb einen grösseren Informationsgehalt. Jedoch muss mit Redundanzen gerechnet werden, da bei der Beantwortung von Emails die Anfrage vielfach wieder zurückgeschickt wird.

Der Einbau in das Klassifikationsmodell erfolgt passiv. Das Hintergrundwissen wird nur benutzt um die TF-IDF-Werte des Trainingsdatensatzes während der Vorverarbeitung zu verbessern. Die Schlüsselwörter im Datensatz sollen dadurch noch stärkeres Gewicht erhalten.

In einem ersten Schritt werden alle den Emails angehängte Dateien entfernt. Danach durchlaufen die Emails die gleiche Vorverarbeitung wie die Telefonanrufe. Bei der Bildung des Vektorraums werden jedoch nur die in den Testanrufen vorkommenden Wörter verwendet. Für die Klassifikation wird, wie bereits bei der einfachen Klassifikation, der K-Nächste-Nachbarn-Algorithmus verwendet. Die Emails selbst werden dabei nicht miteinbezogen, da sie keiner Klasse zugeordnet sind.

1 Kapitel 3.2.4

2 Kapitel 4.5.4

4.6.4 Klassifikation mit Rückfragen

Bei einer Klassifikation mit Rückfragen wird versucht, die einfache Klassifikation aus dem Kapitel 4.6.2 durch eine Rückfrage an den Anrufer zu verbessern. Dazu wird die ermittelte Klasse als Basis verwendet und durch die Antwort auf die gestellte Frage gestützt oder verworfen. Der Inhalt der Frage besteht aus einem Wort und kann vom Anrufer bejaht oder verneint werden (z.B. “Geht es bei ihrem Anliegen um eine *Konferenz*?”).

Grundsätzlich müssen bei der Erzeugung einer Rückfrage zwei Problemstellungen gelöst werden. Als erstes muss eine alternative Klasse zur bereits vom Klassifikator bestimmten Klasse ermittelt werden. Danach kann jenes Fragewort gesucht werden, das die beiden Klassen am besten voneinander differenziert. Anhand der Antwort wird nun die Klasse des Anrufes festgelegt.

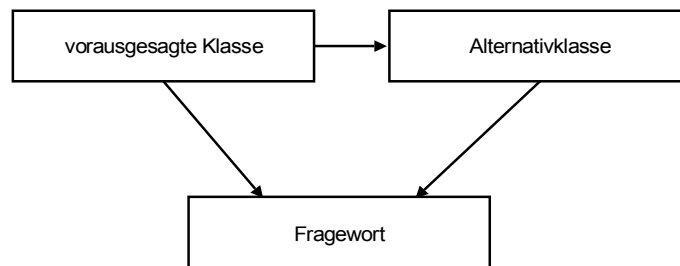


Abbildung 4.6: Erstellung einer Rückfrage

Alternativklasse

Eine gute Bestimmung der alternativen Klasse ist wichtiger, als es auf den ersten Blick scheint. Sie setzt jedoch die Messlatte für die maximal zu erreichende Genauigkeit der Klassifikation mit Rückfragen. Falls weder die durch die einfache Klassifikation bestimmte Klasse noch die Alternativklasse der wirklichen Klasse des Anrufes entspricht, nützt auch eine perfekte Frage nichts. Der Anruf kann unmöglich der korrekten Klasse zugeordnet werden.

Die später in der Evaluation verwendeten Ansätze werden folgend kurz beschrieben. Bei der Auswahl der Alternativklasse wird die bereits durch die einfache Klassifikation vorausgesagte Klasse sinnigerweise nicht miteinbezogen.

- **Zufall:** Eine der Klassen wird durch “Würfeln” ermittelt.
- **Klasse mit zweithöchster, vorausgesagter Wahrscheinlichkeit:** Jene Klasse, die in der einfachen Klassifikation die zweithöchste Wahrscheinlichkeit aufweist, wird

ausgewählt. Als Alternative kann die Wahrscheinlichkeit auch mit der Anzahl Instanzen normiert werden.

- **Klasse mit grösster Anzahl Instanzen:** Die Klasse mit den meisten Instanzen im Trainingsdatensatz wird ausgewählt.
- **Precision:** Die Klasse mit der schlechtesten Precision¹ wird verwendet.
- **Recall:** Die Klasse mit dem schlechtestem Recall¹ wird verwendet.
- **Summe der falschen Positiven und Negativen:** Mit diesem Ansatz werden beide Fehler, die falschen Positiven wie auch die falschen Negativen, berücksichtigt.

Fragewort

Ein Fragewort soll eine Klasse möglichst gut von einer anderen unterscheiden. Eine solche Unterscheidung kann sowohl global über alle vorhandenen Instanzen im Datensatz, wie auch lokal für eine gewisse Anzahl nächster Nachbarn betrachtet werden. Je grösser die Themenvielfalt der Anrufe ist, desto lokaler sollte der gewählte Ansatz sein. Sonst läuft man Gefahr, dass die Fragen zu undifferenziert und deshalb unbrauchbar sind.

In einem ersten Schritt wird eine Differenzmatrix zwischen allen Klassen und den darin vorkommenden Wörtern im Datensatz gebildet. Zu diesem Zweck wird der Satz der Entropie verwendet [Clausius 1865]. Anhand der Differenzmatrix lässt sich bestimmen, wie fest ein Wort die verschiedenen Klassen untereinander trennt. Taucht beispielsweise das Wort *Umfrage* häufig in der Klasse *Unerwünscht* und selten in der Klasse *wichtig* auf, so ist der Wert dieses Wortes in der Differenzmatrix an der Stelle *wichtig-unerwünscht* hoch, respektive tief an der Stelle *unerwünscht-wichtig*.

Nachdem nun die Werte aller Wörter berechnet sind, gilt es das für die aktuelle Klassifikation beste Fragewort zu finden. Zur Formulierung der Frage kann entweder die ausgewählte Alternativklasse oder aber auch die durch die einfache Klassifikation erhaltene Klasse benutzt werden. Auf jeden Fall müssen die entsprechenden Differenzwerte nach ihrem lokalen Bezug geordnet werden.

Dies kann anhand der folgend vorgestellten Methoden geschehen:

- **Anzahl Nachbarn:** Eine bestimmte Anzahl nächster Nachbarn, die der Fragewortklasse zugehören, wird für die Ermittlung des Fragewortes verwendet.

¹ Kapitel 3.3.2

- **Maximale Distanz:** Es werden nur Instanzen der Fragewortklasse innerhalb eines bestimmten Radius berücksichtigt.
- **Gewichtung:** Alle Instanzen der Fragewortklasse werden einbezogen jedoch mit ihrem Abstand gewichtet.

Neben dieser Rangierung müssen die Worte auch auf ihre Zweckmässigkeit im Zusammenhang mit der Erzeugung der zu stellenden Frage überprüft werden. So macht das Rückfragen von Wörtern der Wortarten Artikel oder Hilfsverben wenig Sinn, weshalb in diesem Experiment nur Nomen in die Auswahl einbezogen werden. Auch Eigennamen, Begrüssungs- und Verabschiedungsformeln werden gefiltert. Schlussendlich wird das beste verbleibende Wort für die Rückfrage verwendet.

4.7 Evaluation

Die Resultate des vorgängig vorgestellten Machine-Learning-Frameworks werden in diesem Kapitel ausführlich evaluiert. Zuerst werden die für das Experiment gewonnenen Daten, danach die verschiedenen Methoden der Klassifikation analysiert.

4.7.1 Datensatz (deskriptive Statistik)

Insgesamt können für das Experiment 191 Anrufe genutzt werden. Um ein gewisses Verständnis für den Inhalt dieser Anrufe zu vermitteln, wird je ein Satz pro Klasse folgend vorgestellt. Die Auswahl ist durch die grosse Themenvielfalt auch innerhalb einer Klasse nicht repräsentativ.

- **Unerwünscht:** “Guten Tag Herr Bernstein. Ich bin Vertreter der Nixgut AG und würde ihnen gerne unser neues Produkt vorstellen. Hätten sie dafür Zeit, dann würde ich gerne einen Termin mit ihnen abmachen. Vielen Dank. Auf Wiederhören.”
- **Unwichtig:** “Guten Tag Herr Bernstein. Hier ist Frau Müller von der Konferenz für Informatik im Unternehmen. Für das Tagungsprogramm brauchen wir noch ihr Paper. Bitte schicken sie das bis spätestens in zwei Wochen an uns. Wiederhören.”
- **Neutral:** “Hallo Avi. Hier ist die Lotti. Es geht um die Organisation des Kolloquiums und zwar haben wir noch einen freien Termin, an dem wir einen Gastredner einladen können. Können wir das in nächster Zeit zusammen besprechen? Vielen Dank.”
- **Wichtig:** “Guten Tag Herr Bernstein. Hier ist Frau Cornelia Bösch von der Unitectra. Es geht um den Forschungsvertrag der dringend bis zum 24. Juli unterschrieben werden muss. Bitte rufen sie mich diesbezüglich zurück.“
- **Sehr wichtig:** “Guten Tag Avi. Hier ist Haym. Wir haben heute abgemacht um über neue Textminingansätze zu diskutieren.“

Mit Abstand am meisten Anrufe kommen aus der Gruppe *Sekretariat des Instituts für Informatik* (Sekt). Dahinter folgen die Gruppen *Verbände und Konferenzen* (VerbKonf), *auswärtige Professoren und Nationalfonds* (ProfNFS), *Universität Zürich* (UniZH), *Professoren und Assistenten des IfI* (ProfAssis), *Unternehmen* (Untern), *Techniker des IfI und Studenten* (TechStud) und dem *Spam*.

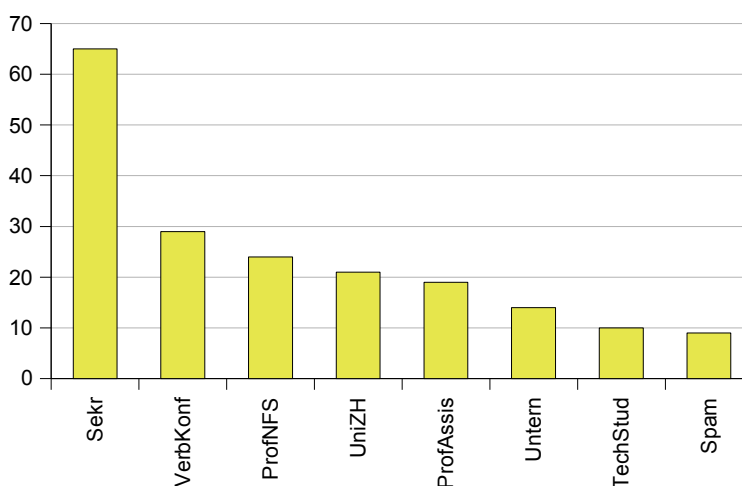


Abbildung 4.7: Anzahl Anrufe pro Gruppe

Gut 70% der Anrufe haben die Wichtigkeitsklasse *neutral* (3) oder *wichtig* (4). Der Anteil an *sehr wichtigen Anrufen* (5) ist ebenso gering wie der an *Spam* (1). Zusammen mit jenen der Klasse *unwichtig* (2) sind somit fast 25% der Anrufe für die Testperson nicht relevant.

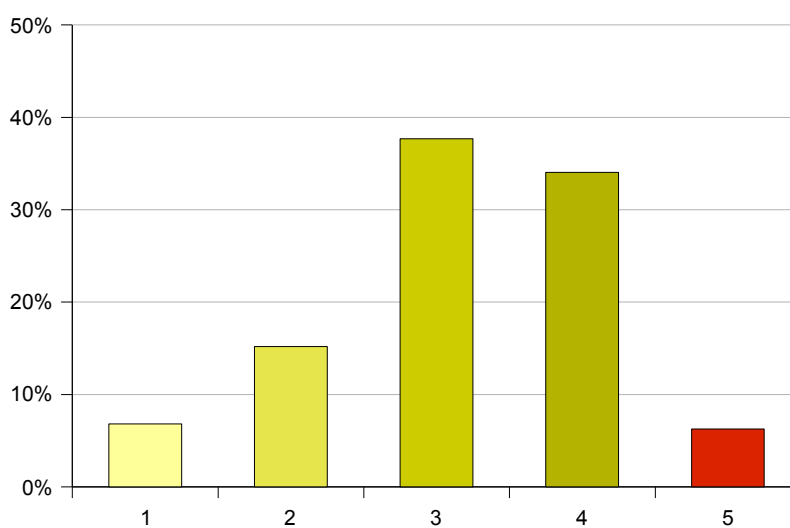


Abbildung 4.8: Relative Verteilung der Wichtigkeitsklassen

Die Verteilung der Wichtigkeitsklassen auf die verschiedenen Gruppen fällt sehr unterschiedlich aus. Am meisten wichtige Anrufe kommen aus den Gruppen *Professoren und Assistenten*, *auswärtige Professoren und Nationalfonds*, *Unternehmen* und dem *Sekretariat*. Wie erwartet kommen aus der Gruppe *Spam* keine relevanten Anrufe.

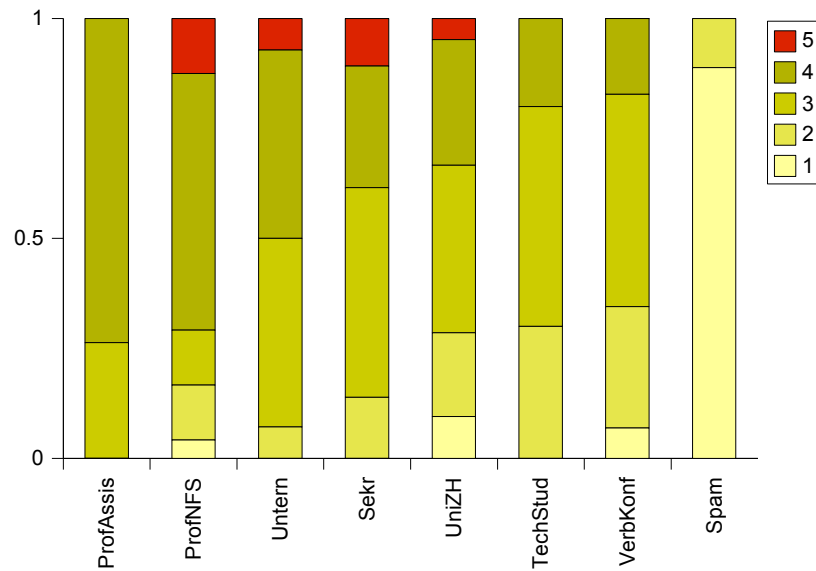


Abbildung 4.9: Relative Verteilung der Wichtigkeitsklassen pro Gruppe

4.7.2 Einfache Klassifikation

Zur Bestimmung der Anzahl k Nächste-Nachbarn (k NN) wird die Klassifikation mit verschiedenen Konfigurationen wiederholt durchgeführt. Als Indikator wird die Genauigkeit verwendet. Die den Berechnungen zugrunde liegenden Daten sind im Anhang E zu finden.

In 4.10 ist die Genauigkeit zweier verschiedener Probleme aufgezeichnet. Beim 5-Klassen-Problem werden alle vorkommenden Klassen einzeln betrachtet. Im 3-Klassen-Problem werden die Klassen *Spam* und *unwichtig*, sowie *wichtig* und *sehr wichtig* zusammengefasst. Die Klasse *neutral* wird weiterhin einzeln betrachtet. Als Grundwert für die Genauigkeit der Klassifikation kann das relative Vorkommen der häufigsten Klasse im Datensatz verwendet werden. Dies entspricht in diesem Experiment 37.7%¹ für das 5-Klassen-Problem und 40.3%² für das 3-Klassen-Problem. Das Mass “Anzahl vorhergesagter Klassen” zeigt, wie viele verschiedene Klassen vom jeweiligen Modell beim 5-Klassen-Problem überhaupt vorhergesagt werden³. Optimal wären natürlich alle vorhandenen Klassen – in unserem Fall fünf.

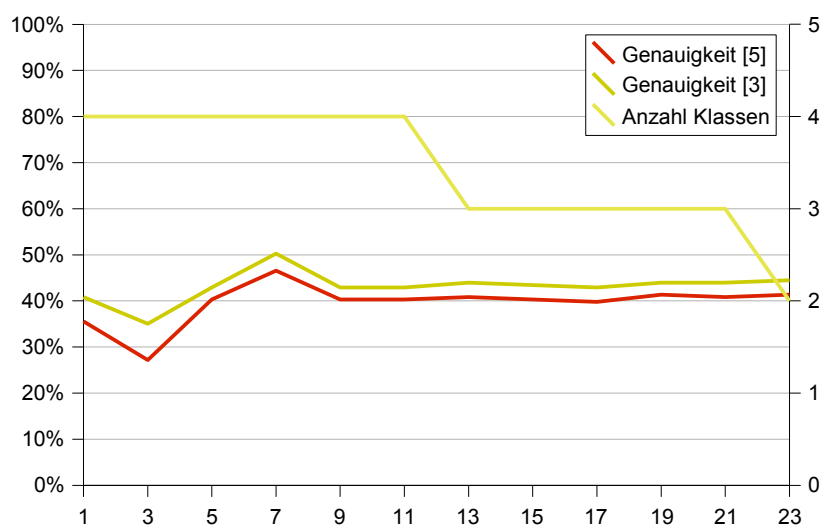


Abbildung 4.10: Genauigkeit 5-Klassen-Problem, Genauigkeit 3-Klassen-Problem und Anzahl vorausgesagter Klassen bei verschiedenen KNN

Es zeigt sich, dass bei allen Massen mit sieben Nächste-Nachbarn das beste Resultat erzielt wird. Die beim 5-Klassen-Problem erreichte Genauigkeit von 46.6% liegt über dem Grundwert von 37.7%. Das 3-Klassen-Problem wird ähnlich gut gelöst.

¹ relatives Vorkommen der Klasse 3, Tabelle 4.8

² Summe der relativen Vorkommen der Klassen 4 und 5

³ Zu geringe Aussagekraft beim 3-Klassen-Problem

Die Tabelle 4.3 zeigt die durch eine Klassifikation mit sieben Nächste-Nachbarn (7NN) erhaltene Confusion Matrix¹. Insgesamt werden vier verschiedene Klassen vorhergesagt. Die Klasse *sehr wichtig* taucht in der Voraussage nicht auf. Die Klasse *neutral* wird am besten bestimmt. Dies kann mit der aus Abbildung 4.8 ersichtlichen hohen relativen Verteilung dieser Klasse erklärt werden. Die Matrizen der anderen Klassifikationen befinden sich in Anhang E.

		Voraussage				
		Klasse	1	2	3	4
Realität	1	7	1	4	1	0
	2	4	1	19	5	0
	3	4	3	52	13	0
	4	2	1	33	29	0
	5	1	1	8	2	0

Tabelle 4.3: Confusion Matrix 7NN

Verglichen mit den im Kapitel 3.2.4 vorgestellten Mechanismen schneidet die 7NN-Klassifikation für das 5- und 3-Klassen-Problem leicht besser ab. Die genauen Werte können der Tabelle 4.4 entnommen werden.

Methode	Genauigkeit [5]	Genauigkeit [3]	Anzahl Klassen
Grundwert	37.7%	40.3%	1
7NN	46.6%	50.3%	4
Support-Vektor-Maschine	45.6%	49.7%	5
Naive Bayes	44.5%	47.6%	2

Tabelle 4.4: Vergleich der Klassifikationsmethoden

¹ Kapitel 3.3.2

4.7.3 Klassifikation mit Hintergrundwissen

Das Hintergrundwissen besteht aus insgesamt 112 Emails der Testperson. Darunter befinden sich auch einige wenige Fremdsprachige. Wie im Kapitel 4.6.3 beschrieben, werden durch sie die TF-IDF-Werte des Trainingsdatensatzes beeinflusst.

In der Abbildung 4.11 werden wieder die im vorhergehenden Kapitel vorgestellten Masse verwendet. Dargestellt ist die einfache Klassifikation mit sieben Nächste-Nachbarn (7NN) und verschiedene Klassifikationen mit Hintergrundwissen und unterschiedlicher Anzahl Nächste-Nachbarn (xNN_HW).

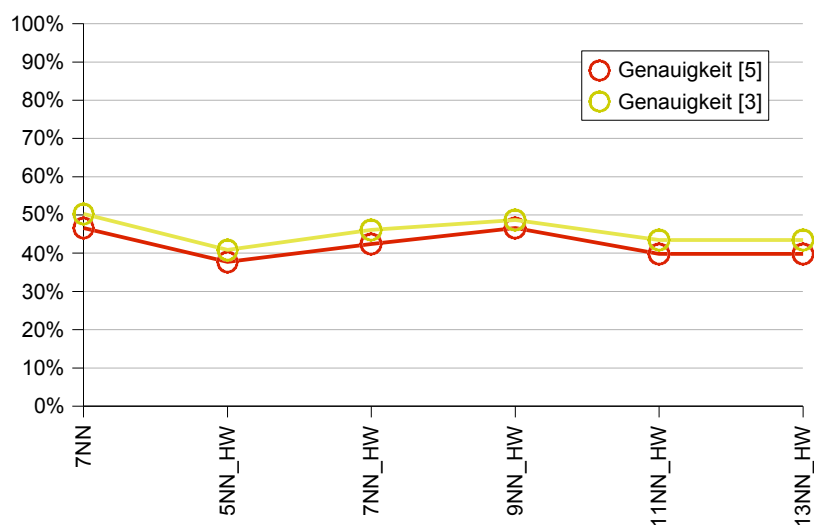


Abbildung 4.11: Genauigkeit 5-Klassen-Problem und Genauigkeit 3-Klassen-Problem bei verschiedenen KNN mit Hintergrundwissen

Das Hintergrundwissen verbessert die Klassifikation nicht. Mit einer Genauigkeit von 46.6% beim 5-Klassen-Problem fällt die Klassifikation mit neun Nächste-Nachbarn und Hintergrundwissen (9NN_HW) exakt gleich gut aus wie die einfache Klassifikation mit sieben Nächste-Nachbarn (7NN). Bei Betrachtung des 3-Klassen-Problems liegt sie sogar einen Prozentpunkt tiefer. Interessant ist die Verschiebung der besten Methode von sieben zu neun Nächste-Nachbarn.

4.7.4 Klassifikation mit Rückfragen

Wie bereits im Kapitel 4.6.4 beschrieben, ist es sinnvoll, die Ermittlung der Alternativklasse und die Suche nach dem besten Fragewort entkoppelt zu betrachten. Dementsprechend wird erst am Schluss dieses Kapitels die Klassifikation mit Rückfragen gesamthaft evaluiert.

Es wird versucht, bei der Ermittlung der Alternativklasse und des Fragewortes den jeweils besten Ansatz zu finden. Erst danach werden die beiden Methoden kombiniert und mit der besten der einfachen Klassifikation (7NN) verglichen.

Alternativklasse

Die im Kapitel 4.6.4 beschriebenen Ansätze zur Bestimmung der alternativen Klasse werden wie die einfache Klassifikation auf ihre Genauigkeit beim 5- und beim 3-Klassen-Problem untersucht. Die Berechnung der Werte (z.B. Precision) für die verschiedenen Ansätze werden anhand der Confusion Matrix der 7NN-Klassifikation durchgeführt¹.

Beim 5-Klassen-Problem schneiden zwei Methoden genau gleich gut ab. Dies kann mit den Eigenschaften des Experimentdatensatzes und des daraus gebildeten Klassifikationsmodells erklärt werden². Bei einem anderen Datensatz können die Resultate verschieden sein. Als Vergleich dazu sind auch die Werte der 7NN-Klassifikation in der Abbildung 4.12 dargestellt.

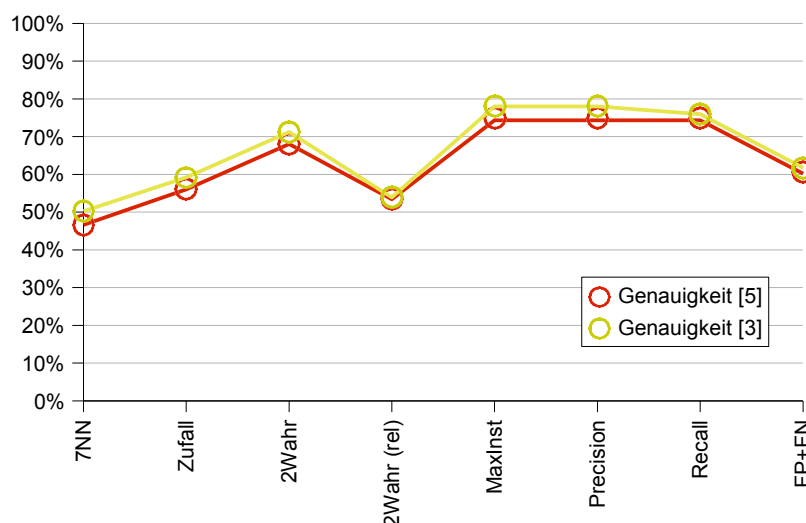


Abbildung 4.12: Genauigkeit 5-Klassen-Problem und Genauigkeit 3-Klassen-Problem bei verschiedenen Methoden zur Bestimmung der Alternativklasse

¹ Tabelle 4.3

² ungleichmässige Verteilung der Klassen

Mit der Wahl der beiden besten Methoden, “Precision” oder “Klasse mit grösster Anzahl Instanzen” (MaxInst), kann bei einer Klassifikation mit Rückfragen maximal 74.4% Genauigkeit erreicht werden. Dies entspricht einer Verbesserung von mehr als der Hälfte gegenüber der einfachen Klassifikation (7NN). Im Vergleich zum Grundwert von 37.7% verdoppelt sich die Genauigkeit fast. Voraussetzung für das Erreichen dieses Wertes ist die perfekte Methode zur Ermittlung des Fragewortes und die wahrheitsgemässe Beantwortung davon durch den Anrufer.

Fragewort

Anders als bei der Evaluation der Alternativklasse läuft die Auswertung bei der Ermittlung eines Fragewortes nicht automatisch ab. Jede vom Machine-Learning-System erzeugte Frage muss manuell beurteilt und beantwortet werden. Aus diesem Grund werden nicht alle 191 Instanzen verwendet, sondern aus jeder Klasse vier möglichst unterschiedliche Anrufe ausgewählt. Dies ergibt gesamthaft 20 Anrufe oder gut 10% des Datensatzes. Bei der Evaluation wird für alle ausgewählten Anrufe ein Fragewort für jede fremde Klasse erzeugt. So entstehen pro Anruf maximal vier Fragen oder 80 bei jeder Evaluation einer Methode.

Insgesamt werden die drei im Kapitel 4.6.4 vorgestellten Methoden mit verschiedenen Einstellungen zur Ermittlung eines Fragewortes getestet. Dies sind: Die “Anzahl Nächste-

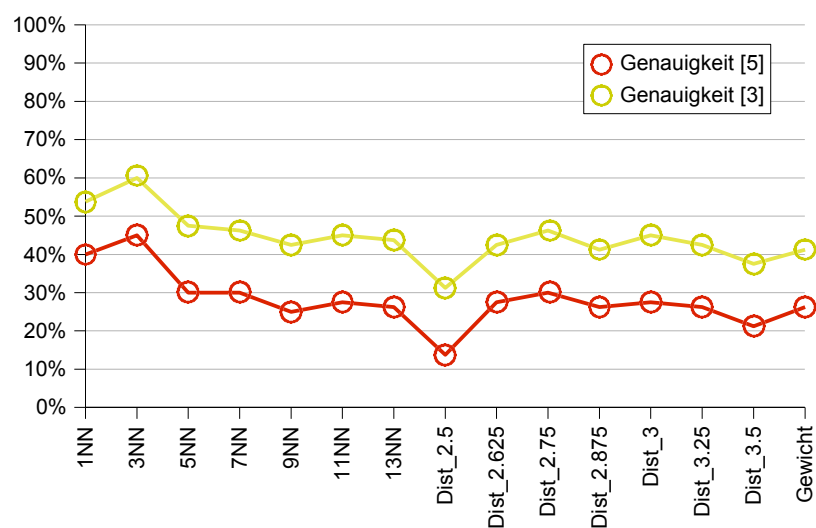


Abbildung 4.13: Genauigkeit 5-Klassen-Problem und Genauigkeit 3-Klassen-Problem bei verschiedenen Methoden zur Bestimmung des Fragewortes

Nachbarn" (xNN), der "maximale Abstand der Nächste-Nachbarn" (Dist_x) und die "Gewichtung aller Instanzen" (Gewicht).

Mit 45% Genauigkeit schneidet dabei die Generierung des Fragewortes mit drei Nächste-Nachbarn (3NN) klar am besten ab. Weder die Berücksichtigung der Nächste-Nachbarn innerhalb einer bestimmten Distanz (Dist_x) noch die Gewichtung aller Instanzen mit ihrem Abstand (Gewicht) überzeugen.

Gründe dafür sind in der Abbildung 4.14 ersichtlich. In ihr ist die Genauigkeit beim 5-Klassen-Problem, die Anzahl der erzeugten Fragen und die Anzahl aller positiven Antworten dargestellt. Dabei werden für die Bestimmung der Genauigkeit nur jene Klassifikationen berücksichtigt, bei denen überhaupt eine Frage erstellt wird.

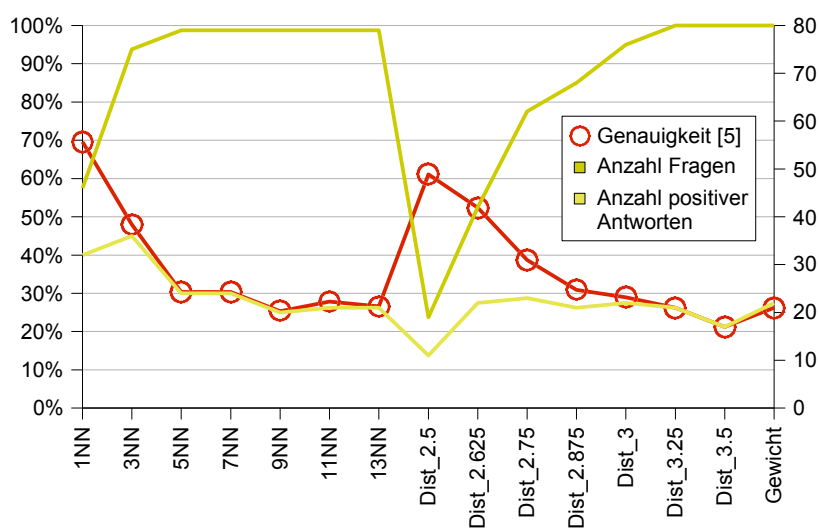


Abbildung 4.14: Genauigkeit 5-Klassen-Problem bei vorhandener Frage, Anzahl Fragen und Anzahl positiver Antworten bei verschiedenen Methoden zur Bestimmung des Fragewortes

Mit steigender Anzahl Fragen nimmt ihre Qualität markant ab. Methoden, die eine hohe Genauigkeit aufweisen (1NN und Dist_2.5), tragen jedoch wegen ihrer geringen Anzahl Fragen nur wenig zu der in Abbildung 4.13 dargestellten Gesamtgenauigkeit bei.

Kombination Alternativklasse und Fragewort

Die jeweils besten Methoden zur Ermittlung der Alternativklasse (Precision)¹ und der Bestimmung des Fragewortes (3NN)² werden in die beste Methode der einfachen Klassifikation (7NN)³ integriert.

Zum einen werden die im Unterkapitel Fragewort ausgewählten Instanzen evaluiert (7NN mit Rückfragen, 20 Instanzen), zum anderen wird die Methode auf alle Instanzen des Datensatzes angewendet (7NN mit Rückfragen, 191 Instanzen). Die verwendeten Masse sind im Kapitel 4.7.2 beschrieben und als Vergleich ist auch die einfache 7NN-Klassifikation dargestellt.

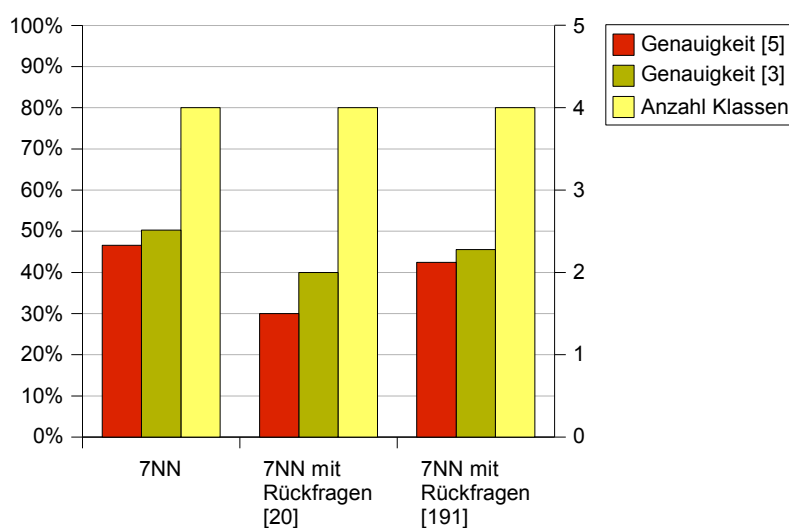


Abbildung 4.15: Genauigkeit 5-Klassen-Problem, Genauigkeit 3-Klassen-Problem und Anzahl vorausgesagter Klassen bei verschiedenen Methoden

Ersichtlich ist, dass die beiden Klassifikationen mit Rückfragen mit einer Genauigkeit beim 5-Klassen-Problem von 42.4% bei 191 Instanzen, respektive 30.0% bei 20 Instanzen, klar schlechter abschneiden als die einfache Klassifikation mit 46.6%. Die Differenz beträgt zwischen den beiden alle Instanzen berücksichtigenden Klassifikationen 9%.

Der Hauptgrund für das schlechte Abschneiden der Klassifikation mit Rückfragen ist die geringe Anzahl an positiven Antworten. Von 189 erzeugten Fragen konnten nur 48 (25%) als zutreffend beantwortet werden. Bei den restlichen stimmte das ermittelte Fragewort nicht mit dem Inhalt des Anrufes überein und führte deshalb zu einer Ablehnung der Frage.

¹ Kapitel 4.7.4 Alternativklasse

² Kapitel 4.7.4 Fragewort

³ Kapitel 4.7.2

In der Tabelle 4.5 ist die Confusion Matrix der Klassifikation mit Rückfragen über allen 191 Instanzen dargestellt. Im Vergleich zur Confusion Matrix der einfachen Klassifikation in Tabelle 4.6 fällt die grössere Anzahl an Voraussagen der Klassen 3 und 4 auf. Dies hängt direkt mit der Wahl der Alternativklasse zusammen. Es werden durch den Datensatz dieses Experiments und der verwendeten Methode “Precision” ausschliesslich Fragen zu diesen beiden Klassen erzeugt. Bei positiver Beantwortung einer Frage wird die betreffende Instanz deshalb einer dieser Klassen zugeteilt.

		Voraussage					
		Klasse	1	2	3	4	5
Realität	1	7	0	5	1	0	
	2	2	2	21	4	0	
	3	2	7	48	15	0	
	4	3	0	38	24	0	
	5	0	1	7	4	0	

Tabelle 4.5: Confusion Matrix 7NN mit Rückfragen, 191 Instanzen

		Voraussage					
		Klasse	1	2	3	4	5
Realität	1	7	1	4	1	0	
	2	4	1	19	5	0	
	3	4	3	52	13	0	
	4	2	1	33	29	0	
	5	1	1	8	2	0	

Tabelle 4.6: Confusion Matrix 7NN

4.8 Einschränkungen

Die Einschränkungen, denen dieses Experiment unterliegt, sind vielfältig. Grösstenteils hängt dies mit der Simulation der Anrufe zusammen. Eine Simulation kann definitionsgemäss die Realität nie perfekt abbilden.

Bereits beim ersten Schritt, der Annahme der Anrufverteilung und ihrer Thematiken, sind unweigerlich Ungenauigkeiten vorhanden. Ausserdem werden saisonale Schwankungen und Verschiebungen über die Zeit nicht berücksichtigt.

Die Durchführung der Anrufe birgt weitere Fehlerquellen. Zum einen besitzen die ausgewählten Personen nicht das Wissen der wirklichen Anrufer, zum anderen können sie die Anweisungen falsch interpretieren. Auch kann sich durch das wiederholte Anrufen, innerhalb der kurzen Zeitspanne ihrer Anrufe, eine gewisse Monotonie einschleichen.

Die bis jetzt erwähnten Fehler werden teilweise ausgeglichen, indem die Testperson die Anrufe erst nach ihrer Durchführung klassiert. Diese Bewertung selbst ist jedoch auch nicht konsistent. Von drei doppelt im Datensatz enthaltenen Anrufen wurden zwei unterschiedlich bewertet. Die Abweichungen betragen jeweils eine Klasse.

Ein weiterer Punkt ist die relativ geringe Zahl an Anrufen (191), die für die Klassifikation zur Verfügung stehen. Das gleiche Problem tritt zumindest in einer Anfangsphase auch in der Realität auf, in der die Grösse des Datensatzes bei null beginnt und teilweise nur langsam anwächst.

Bei der Klassifikation mit Hintergrundwissen muss der kleine Datensatz des Hintergrundwissens erwähnt werden. Er entspricht nur gut der Hälfte der Anrufe. Bei einer viel grösseren Datenmenge könnten die Auswirkungen des Hintergrundwissens stärker zum Tragen kommen.

Da dieses Experiment nur den Telefonanschluss einer Person einbezieht, können gewonnene Resultate nicht verallgemeinert werden, beziehungsweise auf eine andere Person übertragen werden.

5 Schluss

5.1 Diskussion

Die in dieser Diplomarbeit entwickelte Methode zur Verbesserung einer einfachen Klassifikation von Telefonanrufen durch Rückfragen erzielt schlechtere Resultate als die einfache Klassifikation ohne Rückfragen. Auch die einfache Klassifikation befindet sich nicht auf einem für den praktischen Einsatz tauglichen Niveau.

Grundsätzlich wäre eine Steigerung der Qualität der einfachen Klassifikation mit Rückfragen möglich, wie Abbildung 4.12 zeigt. Die Methode zur Auswahl der Alternativklasse ist genügend gut. Die Gründe für das schlechte Abschneiden der Klassifikation mit Rückfragen liegen hauptsächlich in der unzureichenden Ermittlung des Fragewortes.

Eine kleine Analyse des Datensatzes zeigt, dass ein Themengebiet nicht einer Klasse und eine Klasse nicht einem Themengebiet zugeordnet werden kann. Das Wort *Student* taucht beispielsweise in allen Klassen mit Ausnahme der Klasse *sehr wichtig* auf. Trotz Einbezug der lokalen Umgebung kann die wirkliche Bedeutung eines Anrufes nicht ausreichend gut erfasst werden.

Ein Grund dafür kann auch in der Art der Fragestellung mit nur einem inhaltstragenden Wort liegen. Aus einer positiven Antwort auf die Frage “Geht es bei ihrem Anliegen um einen *Studenten*?” kann nicht abgeleitet werden, ob es im Anruf um eine kurzfristige Terminabsage oder eine Anfrage bezüglich einer Vorlesung geht. Die beiden Thematiken werden jedoch von der Testperson in ihrer Wichtigkeit komplett anders wahrgenommen.

Dieses Problem kann sehr wahrscheinlich auch bei einer viel grösseren Anzahl an Dokumenten im Trainingsdatensatz mit den vorgestellten Methoden nicht gelöst werden. Als Alternative bietet sich der Einbezug der semantischen Bedeutung der Anrufe in die Rückfragen an. Dieser Ansatz wird im folgenden Kapitel näher vorgestellt.

5.2 Ausblick

In diesem Kapitel wird ein Blick auf mögliche Weiterentwicklungen dieser Diplomarbeit geworfen. Ob und wie diese Ideen in spätere Arbeiten einfließen, ist noch ungewiss.

Erweitertes Klassifikationsmodell

Bewusst wurde in dieser Diplomarbeit die Klassifikation eines Telefonanrufs nur anhand des Inhalts vorgenommen. Um die Resultate zu verbessern, könnten weitere Informationen wie beispielsweise die Rufnummer oder die Gesprächsdauer in die Bildung des Klassifikationsmodells einfließen.

Durch die Erfassung des aktuellen Arbeitskontexts der Testperson kann eine weitere Komponente der Klassifikation eines Anrufs hinzugefügt werden. So können nach [Mark et al. 2005] fördernde Anrufe priorisiert werden. Ansätze dazu sind in den Arbeiten [Marti 2005] [Horvitz et al. 2002] zu finden.

Bereits publizierte Arbeiten zeigen, dass die Kombination verschiedener Klassifikationsmodelle um einiges robuster sind, als die Systeme für sich alleine betrachtet. [Ting und Low 1997].

Liveexperiment

Aufschlussreich wäre sicherlich die Durchführung eines Liveexperiments. Nur so kann das Funktionieren des Systems in der Realität überprüft werden und viele der im Kapitel 4.8 erwähnten Einschränkungen aufgehoben werden. Grundvoraussetzung dazu ist eine Spracherkennung, die genügend schnell und zuverlässig die gesprochene Sprache in für den Computer verarbeitbaren Text umwandelt. Auch müssten vorgängig rechtliche Aspekte wie der Schutz der gewonnenen Daten geklärt werden.

Bei einem ausführlicheren Experiment wäre auch eine Untersuchung der Übertragbarkeit auf weitere Testpersonen mit anderen Anrufergruppen interessant. Auch das Zeitverhalten der Voraussagequalität kann durch die längere Dauer ermittelt werden.

Semantik

Ein wichtiges Mittel zur Steigerung der Qualität der Rückfragen wird der Einbezug der semantischen Zusammenhänge sein. Dazu muss der Inhalt eines Anrufes semantisch erfasst und in der selben Ontologie wie die bereits bestehende Wissensbasis beschrieben

werden können. Mit Hilfe von Ähnlichkeitsmessungen können die Inhalte der Rückfragen ermittelt werden [Ziegler et al. 2006]. Dadurch werden die Rückfragen spezifischer auf den Inhalt des Anrufes ausgerichtet.

Entdeckung von Manipulationen

Bei einem automatisierten System muss die Gefahr absichtlicher Manipulationen durch die Benutzer berücksichtigt werden. Ein bedeutendes Kriterium des Menschen für die Abschätzung der Absichten seines Gegenübers sind die Emotionen. Mit ihnen werden sehr viele Informationen übertragen, die nur anhand des Inhalts eines Gesprächs nicht ersichtlich sind. Einem Computer fällt die unter Umständen vorhandene Nervosität eines Schwindlers am anderen Ende der Leitung nicht auf, der sich als persönliche Sekretärin ausgibt. Erste Forschungsansätze sehen vielversprechend aus [Athanaselisa 2005] [Fragopanagos 2005] [Toivanen et al. 2006].

Literaturverzeichnis

- [Athanaselisa 2005] Athanaselisa, t.; Bakamidisa, S.; Dologloua, I.; Cowieb, R.; Douglas-Cowieb, E.; Cox, C.: ASR for emotional speech: Clarifying the issues and enhancing performance, *Neural Networks* 18 (2005):437-444, 2005
- [Bayes 1763] Bayes, T.: An Essay towards solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London* 53:370–418, 1763
- [Boulevard und Dupont 1996] Boulevard, H.; Dupont, S.: A new ASR approach based on independent processing and recombination of partial frequency bands, in *ICSLP 1*: 426-429, Philadelphia, Pennsylvania, 1996.
- [Clausius 1865] Clausius, R.: *The Mechanical Theory of Heat – with its Applications to the Steam Engine and to Physical Properties of Bodies*, London: John van Voorst, 1 Paternoster Row, 1865
- [Chu-Carroll und Carpenter 1999] Chu-Carroll, J.; Carpenter, B.: *Vector-based Natural Language Call Routing*, Association for Computational Linguistics, 1999
- [Cole et al. 1997] Cole, R. A.; Mariani, J.; Uszkoreit, H.; Zaenen, A.; Zue, V.: *Survey of the State of the Art in Human Language Technology*, Center for Spoken Language Understanding CSLU, Carnegie Mellon University, Pittsburgh, 1997
- [Cover und Hart 1967] Cover, T. M.; Hart, P. E.: Nearest-neighbor pattern classification, *IEEE, Transactions on Information Theory* IT-13:21–27, 1967
- [Donner 2005] Donner, M.: *Heterogenität von Benutzergruppen in mobilen, kontextbewussten Anwendungen*, Diplomarbeit, Institut für Informatik der Universität Zürich, 2005
- [Fragopanagos 2005] Fragopanagos, N.; Taylor, J.G.: Emotion recognition in human-computer interaction, *Neural Networks* 18 (2005):389-405, 2005
- [Graham-Rowe 2006] Graham-Rowe, D.: A Sentinel to Screen Phone Calls, <http://www.technologyreview.com>, 14.8.2006
- [Haapalainen und Majorin 1994] Haapalainen, M.; Majorin, A.: *GERTWOL: Ein System zur automatischen Wortformenerkennung deutscher Wörter*, 1994
- [Horvitz et al. 2002] Horvitz, E.; Koch, P.; Kadie, C.; Jacobs, A.: Coordinate: Probabilistic Forecasting of Presence and Availability, *Conference on Uncertainty and Artificial Intelligence* 18:224-233, 2002
- [Joachims 2002] Joachims, T.: *Learning to Classify Text using Support Vector Machines*,

- Vol. 668, Kluwer, 2002
- [Kohavi 1995] Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, 14th International Joint Conference on Artificial Intelligence, 1995
- [Lee et al. 1998] Lee, C.-H.; Carpenter, R.; Chu-Carroll, J.; Reichl, W.; Saad, A.; Zhou, Q.: A Study on Natural Language Call Routing, IEEE, 1998
- [Lovins 1968] Lovins, J. B.: Development of a Stemming Algorithm, Mechanical Translation and Computational Linguistics, 1968
- [Mark et al. 2005] Mark, G.; Gonzalez, V. M.; Harris, J.: No Task Left Behind? Examining the Nature of Fragmented Work, CHI, 2005
- [Marti 2005] Marti, S.: Autonomous Interactive Intermediaries: Social Intelligence for Mobile Communication Agents, Ph.D. Thesis in Media Arts and Sciences, MIT Media Lab, 2005
- [Marti und Schmandt 2005] Marti, S.; Schmandt, C.: Physical Embodiments for Mobile Communication Agents. Proceedings of UIST, 2005
- [McCallum und Nigam 1998] McCallum, A.; Nigam, K.: A Comparison of Event Models for Naive Bayes Text Classification, AAI/ICML-98 Workshop on Learning for Text Categorization, AAAI Press, 1998
- [Nadas 1984] Nadas, A.: Estimation of probabilities in the language model of the IBM speech recognition system, in IEEE Transaction on Acoustics, Signal and Speech Processing, 32(4): 859 - 861, 1984.
- [Salton et al. 1975] Salton, G.; Wong, A.; Yang, C. S.: A Vector Space Model for Automatic Indexing, Commun. ACM, 18, p. 613-620, 1975.
- [Schiller et al. 1999] Schiller, A.; Teufel, S.; Stöckert, C.: Guidelines für das Tagging deutscher Textcorpora mit STTS, 1999
- [Schmid 1995] Schmid, H.: Improvements In Part-of-Speech Tagging With an Application To German, Proceedings of the ACL SIGDAT-Workshop, 1995
- [Snowball 2006] <http://snowball.tartarus.org>, 14.9.2006
- [Suhm et al. 2002] Suhm, B.; Bers, J.; McCarthy, D.; Freeman, B.; Getty, D.; Godfrey, K.; Peterson, P. : A comparative study of speech in the call center: natural language call routing vs. touch-tone menus, SIGCHI, 2002
- [Ting und Low 1997] Ting, K. M.; Low, B. T.: Model Combination in the Multiple-Data-Batches Scenario, European Conference on Machine Learning, 1997
- [Toivanen et al. 2006] Toivanen, J.; Waaramaa, T.; Alku, P.; Laukkanen, A.; Seppanen, T.;

- Vayrynen, E.; Airas, M.: Emotions in [a]: A perceptual and acoustic study, *Logopedics Phoniatrics Vocology* 31:43-48, 2006
- [Vorbürger et al. 2005] Vorbürger, P. ; Bernstein, A.; Zurfluh, A.: Interruptability Prediction Using Motion Detection, *First International Workshop on Managing Context Information in Mobile and Pervasive Environments, MCMP-05*, 2005.
- [Waibel et al. 1989] Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K.; Lang, K.J.: Phoneme recognition using time-delay neural networks, *Acoustics, Speech, and Signal Processing, IEEE Transactions*, 37(3): 328-339, 1989
- [Willée et al. 2002] Willée, G.; Schröder, B.; Schmitz, H.: *Computerlinguistik - Was geht, was kommt?*, Festschrift für Winfried Lenders, Gardez! Verlag, Sankt Augustin, 2002
- [Witten 2004] Witten, I. H.: Text Mining, in M. P. Singh, editor, *Practical Handbook of Internet Computing*, Boca Raton, FL: CRC Press, 2004
- [Witten und Frank 2005] Witten, I. H.; Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, 2005
- [Yang und Wilbur 1996] Yang, Y.; Wilbur, J.: Using Corpus Statistics to Remove Redundant Words in Text Categorization, *JASIS* 47 (5): 357-369, 1996
- [Zelikovitz und Hirsh 2000] Zelikovitz, S.; Hirsh, H.: Improving Short-Text Classification Using Unlabeled Background Knowledge to Assess Document Similarity, *ICML*, 2000
- [Ziegler et al. 2006] Ziegler, P.; Kiefer, C.; Sturm, C.; Dittrich, K.; Bernstein, A.: Generic Similarity Detection in Ontologies with the SOQA-SimPack Toolkit, 2006 *ACM SIGMOD International Conference on Management of Data*, 2006

Abbildungsverzeichnis

Abbildung 1.1: Konzept des Klassifikationssystems.....	7
Abbildung 3.1: Spracherkennung I.....	14
Abbildung 3.2: Spracherkennung II.....	15
Abbildung 3.3: Phasenmodell einer Textklassifikation.....	16
Abbildung 3.4: Vektorraum.....	19
Abbildung 3.5: Support-Vektor-Maschinen.....	21
Abbildung 4.1: Problemstellung.....	23
Abbildung 4.2: Experiment.....	24
Abbildung 4.3: Themenkarte.....	26
Abbildung 4.4: Klassendiagramm des Machine-Learning-Framework.....	28
Abbildung 4.5: Methoden der Klassifikation.....	30
Abbildung 4.6: Erstellung einer Rückfrage.....	32
Abbildung 4.7: Anzahl Anrufe pro Gruppe.....	36
Abbildung 4.8: Relative Verteilung der Wichtigkeitsklassen.....	36
Abbildung 4.9: Relative Verteilung der Wichtigkeitsklassen pro Gruppe.....	37
Abbildung 4.10: Genauigkeit 5-Klassen-Problem, Genauigkeit 3-Klassen-Problem und Anzahl vorausgesagter Klassen bei verschiedenen KNN.....	38
Abbildung 4.11: Genauigkeit 5-Klassen-Problem und Genauigkeit 3-Klassen-Problem bei verschiedenen KNN mit Hintergrundwissen.....	40
Abbildung 4.12: Genauigkeit 5-Klassen-Problem und Genauigkeit 3-Klassen-Problem bei verschiedenen Methoden zur Bestimmung der Alternativklasse.....	41
Abbildung 4.13: Genauigkeit 5-Klassen-Problem und Genauigkeit 3-Klassen-Problem bei verschiedenen Methoden zur Bestimmung des Fragewortes.....	42
Abbildung 4.14: Genauigkeit 5-Klassen-Problem bei vorhandener Frage, Anzahl Fragen und Anzahl positiver Antworten bei verschiedenen Methoden zur Bestimmung des Fragewortes.....	43
Abbildung 4.15: Genauigkeit 5-Klassen-Problem, Genauigkeit 3-Klassen-Problem und Anzahl vorausgesagter Klassen bei verschiedenen Methoden.....	44

Tabellenverzeichnis

Tabelle 3.1: Confusion Matrix.....	22
Tabelle 4.1: Anrufergruppen.....	25
Tabelle 4.2: Klassen.....	26
Tabelle 4.3: Confusion Matrix 7NN.....	39
Tabelle 4.4: Vergleich der Klassifikationsmethoden.....	39
Tabelle 4.5: Confusion Matrix 7NN mit Rückfragen, 191 Instanzen.....	45
Tabelle 4.6: Confusion Matrix 7NN.....	45

A Vorverarbeitung dreier Mustersätze

Folgend werden drei Mustersätze analog zur Vorverarbeitung im Machine-Learning-Framework verarbeitet.

Mustersätze

- a) Hans hat eine Katze.
- b) Hans füttert seine Katze.
- c) Das ist die Katze.

Morphologie

- a) Hans haben ein Katze.
- b) Hans füttern er Katze.
- c) Das sein die Katze.

Filterung

- a) Hans haben Katze.
- b) Hans füttern Katze.
- c) Sein Katze.

Vektordarstellung

	Hans	haben	Katze	füttern	sein
a	1	1	1	0	0
b	1	0	1	1	0
c	0	0	1	0	1

Tabelle A.1: Vorkommen der Wörter

	Hans	haben	Katze	füttern	sein
a	0.33	0.33	0.33	0	0
b	0.33	0	0.33	0.33	0
c	0	0	0.5	0	0.5

Tabelle A.2.: Termfrequenzen

	Hans	haben	Katze	füttern	sein
IDF	1.5	3	1	3	3

Tabelle A.3: Inverse Dokumentfrequenzen

	Hans	haben	Katze	füttern	sein
a	0.059	0.159	0	0	0
b	0.059	0	0	0.159	0
c	0	0	0	0	0.239

Tabelle A.4: TF-IDF-Werte

B Aufstellung der Anrufergruppen

Gruppe	Untergruppe	Thema
Institut für Informatik	50% Sekretariat	65% Terminvereinbarung Anwesenheitsabklärung Absagemitteilung Reisebuchungen Kolloquien Gästebetreuung (Zimmer, Hotel, Telefon) Spesenabrechnung (Belege) Unterschrift für Personalmutation Bestätigung Zwischen-/Schlussberichte Allgemeine Korrespondenzen Konferenz
	Assistenten	15% Forschungsprojekt
	Professoren	10% Organisation Ifl Traktanden Institutssitzung Synchronisation Übung/Vorlesung Absprache Vorlesungsinhalt Plagiat
	Studenten	5% Termin für TLE Diplomarbeit
	Techniker	5% Neuanschaffung Reparatur
Erweitertes Arbeitsumfeld	45% Uni Zürich	25% Prorektorate Unitetra Verwaltung
	Konferenz	20% Organisator Anmeldeprozess Proceedings
	Firmen	15% Sponsoring Konferenz Vortrag
	Verband	15% SI Organisation SI Berufsmarketing SI ECDL SwissICT Kbildung
	Professoren Nationalfond	25% Organisation NF-Projekte Gutachten Berufungskommission Forschungsdiskussion Koordination Forschungsprojekt
Spam	5%	100% Fehlwahl, Werbung, Umfrage

C Anweisung für Simulationsanruf

Bei diesem Experiment geht es um ein System, das ankommender Anrufe nach deren Wichtigkeit einstuft. Dabei wird ausschliesslich auf inhaltliche Aspekte geachtet. Das System braucht dafür von jedem Anrufer eine kurze Beschreibung seines Anliegen, ähnlich wie bei einem Telefonbeantworter.

Um an Testdaten zu gelangen, sind wir auf deine Mithilfe angewiesen. Bitte versetze dich möglichst gut in die auf den beigelegten Karten beschriebenen Situationen. Sie sind bewusst knapp gehalten und du bist eingeladen, wo es dir sinnvoll erscheint, sie mit z.B. fiktiven Namen, Orten, Zeiten und Gegenständen auszuschnücken. Sprich dabei klar und deutlich in Hochdeutsch.

Das System ist noch nicht lauffähig und deine Anrufe werden deshalb von einem normalen Telefonbeantworter aufgenommen. Falls du Fragen hast, bin ich am besten per Mail (s0090832@access.unizh.ch) erreichbar. Vielen Dank für deine Teilnahme am Experiment!

Simon Bleher

Beispiel:

Experiment Telefonassistent

Bitte beachte beim Anrufen folgende Punkte:

- Telefonnummer: 044 63 57 140
- Sprache: Hochdeutsch
- Code am Ende des Anrufes nennen.

- Rolle: Student
- Ansprechperson: Prof. Bernstein
- Thema: Anfrage zu Diplomarbeiten
- Code: 24-2-1 X

„Guten Tag Herr Bernstein hier spricht Peter Müller. Ich besuche bei ihnen die Vorlesung Business Intelligence. Gerne würde ich wissen welche aktuellen Diplomarbeits-themen momentan frei sind. Darf ich heute Nachmittag kurz vorbeikommen? Vierundzwanzig – Zwei – Eins – X (24-2-1 X)“

D Korrekturen der Anrufe

Datei	Original	Änderung
A008	Müller	Markwalder
A009	Müller	Unterwasser
C005	Meier	Gubler
D008	Mundart	Übersetzung in Hochdeutsch
F001	Felix Muster vom Sekretariat	Eveline
F002	Felix Muster vom Sekretariat	Lotti
F003	Felix Muster vom Sekretariat	Kilit
F003	ihnen	dir
F003	sie	dich
F004	Felix Muster vom Sekretariat	Eveline
F004	schreiben sie	schreib
F005	Felix Muster vom Sekretariat	Eveline
F005	ihnen	dir
F010	Muster	Bannwart
F011	Muster	Schneebeli
F012	Muster	Lehmann
F013	Muster	Walser
F014	Muster	Müller
F015	Muster	Helfenstein
G001	Herr Bernstein	Avi
G001	Frau Reinecke	Katharina
G001	sie aufsetzen wollten	du aufsetzen wolltest
G002	Herr Bernstein	Avi
G002	Frau Reinecke	Katharina
G003	Herr Bernstein	Avi
G003	Frau Schreiberling	Katharina
G003	können sie sich	könntest du
G005	Reinecke	Halder
G006	Reinecke	Jucker
G009	Reinecke	Hermann
G010	Lüthi	Hächler
G011	Lüthi	Eigenmann
I001	Stefanie Hauske	Maria Heiniger
I002	Stefanie Hauske	Isolde Brennwald
I003	Stefanie Hauske	Luzia Schmid
I004	Stefanie Hauske	Luzia Schmid
I006	Stefanie Hauske	Anna Metzger
I007	Stefanie	Lotti
I008	Stefanie	Zehra
I009	Stefanie	Eveline
I012	Stefanie Hauske	Isolde Brennwald
K001	das Sekretariat	Lotti
K003	Michel Hagenauer	Alfred Kunz
K009	Meier	Vogt
K010	das Sekretariat	Eveline
K010	Könnten sie	Könntest du
L001	Manuel Ziltener	Bünter
L002	Manuel Ziltener	Gasser
L003	Manuel vom Sekretariat	Zehra
L004	Manuel vom Sekretariat	Eveline

L007	Manuel Ziltener	Beat Rageth
L008	Manuel Ziltener	Beat Rageth
M001	KKK	KCK
M002	Hannelore Hanselmann	Monika Estermann
M004	Hannelore Hanselmann	Anne Poire
M005	Hannelore Hanselmann vom Sekretariat	Lotti
M006	die Hannelore Hanselmann vom Sekretariat	Eveline
M007	die Hannelore Hanselmann vom Sekretariat	Eveline
M009	die Hannelore Hanselmann	Ramona Schläpfer
M011	die Hannelore Hanselmann	Gertrud Melcher

E Einfache Klassifikation

1	c1	c2	c3	c4	c5		
c1	6	2	4	1	0	13	
c2	3	6	14	5	1	29	
c3	9	9	37	12	5	72	
c4	13	5	28	19	0	65	
c5	0	1	6	5	0	12	
	31	23	89	42	6	191	

3	c1	c2	c3	c4	c5		
c1	9	1	2	1	0	13	
c2	12	2	14	1	0	29	
c3	27	5	34	5	1	72	
c4	35	2	21	7	0	65	
c5	3		7	2	0	12	
	86	10	78	16	1	191	

5	c1	c2	c3	c4	c5		
c1	8	1	3	1	0	13	
c2	3	2	21	3	0	29	
c3	6	2	50	13	1	72	
c4	5		43	17	0	65	
c5	1		10	1	0	12	
	23	5	127	35	1	191	

7	c1	c2	c3	c4	c5		
c1	7	1	4	1	0	13	
c2	4	1	19	5	0	29	
c3	4	3	52	13	0	72	
c4	2	1	33	29	0	65	
c5	1	1	8	2	0	12	
	18	7	116	50	0	191	

9	c1	c2	c3	c4	c5		
c1	1	0	10	2	0	13	
c2	1	1	22	5	0	29	
c3	1	2	51	18	0	72	
c4	2	0	39	24	0	65	
c5	0	0	8	4	0	12	
	5	3	130	53	0	191	

11	c1	c2	c3	c4	c5	
c1	1	0	9	3	0	13
c2	1	1	22	5	0	29
c3	1	3	54	14	0	72
c4	2	0	42	21	0	65
c5	0	0	8	4	0	12
	5	4	135	47	0	191

13	c1	c2	c3	c4	c5	
c1	0	0	11	2	0	13
c2	0	1	23	5	0	29
c3	2	1	54	15	0	72
c4	1	0	41	23	0	65
c5	0	0	6	6	0	12
	3	2	135	51	0	191

15	c1	c2	c3	c4	c5	
c1	0	0	10	3	0	13
c2	0	1	24	4	0	29
c3	1	1	55	15	0	72
c4	0	0	44	21	0	65
c5	0	0	6	6	0	12
	1	2	139	49	0	191

17	c1	c2	c3	c4	c5	
c1	0	0	11	2	0	13
c2	0	1	24	4	0	29
c3	1	0	58	13	0	72
c4	0	0	48	17	0	65
c5	0	0	6	6	0	12
	1	1	147	42	0	191

19	c1	c2	c3	c4	c5	
c1	0	0	10	3	0	13
c2	0	1	24	4	0	29
c3	1	0	57	14	0	72
c4	0	0	44	21	0	65
c5	0	0	7	5	0	12
	1	1	142	47	0	191

21	c1	c2	c3	c4	c5	
c1	0	0	11	2	0	13
c2	0	1	24	4	0	29
c3	1	0	58	13	0	72
c4	0	0	46	19	0	65
c5	0	0	6	6	0	12
	1	1	145	44	0	191

23	c1	c2	c3	c4	c5	
c1	0	0	12	1	0	13
c2	0	0	25	4	0	29
c3	1	0	57	14	0	72
c4	0	0	43	22	0	65
c5	0	0	6	6	0	12
	1	0	143	47	0	191

SVM	c1	c2	c3	c4	c5	
c1	5	0	4	4	0	13
c2	0	6	15	6	2	29
c3	2	11	39	17	3	72
c4	2	9	16	37	1	65
c5	0	1	4	7	0	12
	9	0	78	71	6	191

NB	c1	c2	c3	c4	c5	
c1	0	0	5	8	0	13
c2	0	0	23	6	0	29
c3	0	0	53	19	0	72
c4	0	0	33	32	0	65
c5	0	0	6	6	0	12
	0	0	120	71	0	191

F Klassifikation mit Hintergrundwissen

HW5	c1	c2	c3	c4	c5	
c1	8	0	4	1	0	13
c2	5	2	20	2	0	29
c3	5	3	49	14	1	72
c4	4	1	47	13	0	65
c5	3	0	8	1	0	12
	25	6	128	31	1	191

HW7	c1	c2	c3	c4	c5	
c1	7	0	5	1	0	13
c2	2	2	21	4	0	29
c3	2	7	48	15	0	72
c4	3	0	37	24	1	65
c5	0	1	7	4	0	12
	14	10	118	48	1	191

HW9	c1	c2	c3	c4	c5	
c1	6	0	6	1	0	13
c2	1	2	22	4	0	29
c3	2	3	52	15	0	72
c4	4	0	32	29	0	65
c5	0	0	9	3	0	12
	13	5	121	52	0	191

HW11	c1	c2	c3	c4	c5	
c1	0	0	11	2	0	13
c2	1	1	20	7	0	29
c3	2	2	49	19	0	72
c4	1	0	38	26	0	65
c5	0	0	6	6	0	12
	4	3	124	60	0	191

HW13	c1	c2	c3	c4	c5	
c1	0	0	11	2	0	13
c2	1	1	20	7	0	29
c3	2	2	49	19	0	72
c4	1	0	38	26	0	65
c5	0	0	6	6	0	12
	4	3	124	60	0	191

G Klassifikation mit Rückfragen

Alternativklasse

Zufall	c1	c2	c3	c4	c5	
c1	8	1	3	1	0	13
c2	3	6	16	4	0	29
c3	3	2	56	11	0	72
c4	1	1	28	35	0	65
c5	1	1	6	2	2	12
	16	11	109	53	2	191

2Wahr	c1	c2	c3	c4	c5	
c1	9	1	3	0	0	13
c2	3	4	18	4	0	29
c3	3	1	64	4	0	72
c4	2	1	9	53	0	65
c5	1	1	8	2	0	12
	18	8	102	63	0	191

Wahr (rec1	c1	c2	c3	c4	c5	
c1	12	0	1	0	0	13
c2	0	5	19	5	0	29
c3	4	3	53	12	0	72
c4	2	1	33	29	0	65
c5	1	1	6	1	3	12
	19	10	112	47	3	191

MaxInst	c1	c2	c3	c4	c5	
c1	7	1	4	1	0	13
c2	4	1	19	5	0	29
c3	0	0	72	0	0	72
c4	2	1	0	62	0	65
c5	1	1	8	2	0	12
	14	4	103	70	0	191

Precisio	c1	c2	c3	c4	c5	
c1	7	1	4	1	0	13
c2	4	1	19	5	0	29
c3	0	0	72	0	0	72
c4	2	1	0	62	0	65
c5	1	1	8	2	0	12
	14	4	103	70	0	191

Recall	c1	c2	c3	c4	c5		
c1	7	1	4	1	0	13	
c2	0	5	19	5	0	29	
c3	4	0	68	0	0	72	
c4	2	1	0	62	0	65	
c5	1	1	8	2	0	12	
	14	8	99	70	0	191	

FP+FN	c1	c2	c3	c4	c5		
c1	7	1	4	1	0	13	
c2	0	5	19	5	0	29	
c3	4	3	65	0	0	72	
c4	2	1	33	29	0	65	
c5	1	0	0	2	9	12	
	14	10	121	37	9	191	

Fragewort

1	c1	c2	c3	c4	c5		
c1	5	4	2	3	2	16	
c2	2	7	2	3	2	16	
c3	2	3	7	2	2	16	
c4	2	2	3	7	2	16	
c5	2	3	2	3	6	16	
	13	19	16	18	14	80	

3	c1	c2	c3	c4	c5		
c1	8	3	2	2	1	16	
c2	3	6	1	3	3	16	
c3	2	3	7	2	2	16	
c4	2	2	2	8	2	16	
c5	1	3	1	4	7	16	
	16	17	13	19	15	80	

5	c1	c2	c3	c4	c5		
c1	7	3	2	2	2	16	
c2	3	6	1	3	3	16	
c3	2	3	5	4	2	16	
c4	3	3	3	3	4	16	
c5	2	4	3	4	3	16	
	17	19	14	16	14	80	

7	c1	c2	c3	c4	c5		
c1	7	3	2	2	2	2	16
c2	3	6	1	3	3	3	16
c3	2	3	6	3	2	2	16
c4	3	4	4	2	3	3	16
c5	2	4	3	4	3	3	16
	17	20	16	14	13		80

9	c1	c2	c3	c4	c5		
c1	7	3	2	2	2	2	16
c2	3	6	1	3	3	3	16
c3	3	3	4	3	3	3	16
c4	4	4	4	0	4	4	16
c5	2	4	3	4	3	3	16
	19	20	14	12	15		80

11	c1	c2	c3	c4	c5		
c1	7	3	2	2	2	2	16
c2	3	6	1	3	3	3	16
c3	3	2	5	3	3	3	16
c4	4	4	4	0	4	4	16
c5	1	4	3	4	4	4	16
	18	19	15	12	16		80

13	c1	c2	c3	c4	c5		
c1	7	3	2	2	2	2	16
c2	3	6	1	3	3	3	16
c3	3	2	5	3	3	3	16
c4	4	4	4	0	4	4	16
c5	2	4	3	4	3	3	16
	19	19	15	12	15		80

2.5	c1	c2	c3	c4	c5		
c1	3	4	3	3	3	3	16
c2	4	0	4	4	4	4	16
c3	4	4	1	3	4	4	16
c4	2	2	3	7	2	2	16
c5	4	4	4	4	0	3	16
	17	14	15	21	13		80

2.63	c1	c2	c3	c4	c5		
c1	8	3	2	2	1	1	16
c2	4	0	4	4	4	4	16
c3	3	3	4	3	3	3	16
c4	2	2	3	7	2	2	16
c5	3	4	3	3	3	3	16
	20	12	16	19	13		80

2.75	c1	c2	c3	c4	c5		
c1	8	3	2	2	1	16	
c2	3	3	3	4	3	16	
c3	2	3	6	3	2	16	
c4	3	3	4	3	3	16	
c5	2	4	2	4	4	16	
	18	16	17	16	13	80	

2.87	c1	c2	c3	c4	c5		
c1	8	3	2	2	1	16	
c2	2	4	3	4	3	16	
c3	3	3	3	4	3	16	
c4	3	3	4	3	3	16	
c5	2	4	3	4	3	16	
	18	17	15	17	13	80	

3	c1	c2	c3	c4	c5		
c1	7	3	2	2	2	16	
c2	3	7	1	2	3	16	
c3	3	3	4	3	3	16	
c4	4	4	4	0	4	16	
c5	2	4	2	4	4	16	
	19	21	13	11	16	80	

3.25	c1	c2	c3	c4	c5		
c1	6	2	2	3	3	16	
c2	3	4	3	3	3	16	
c3	3	1	6	3	3	16	
c4	4	4	3	1	4	16	
c5	1	4	3	4	4	16	
	17	15	17	14	17	80	

3.5	c1	c2	c3	c4	c5		
c1	6	2	2	3	3	16	
c2	3	3	3	4	3	16	
c3	3	3	4	3	3	16	
c4	4	4	3	1	4	16	
c5	1	4	4	4	3	16	
	17	16	16	15	16	80	

Kombination Alternativklasse und Fragewort

[20]	c1	c2	c3	c4	c5	
c1	4	0	0	0	0	4
c2	0	0	4	0	0	4
c3	0	1	2	1	0	4
c4	1	0	3	0	0	4
c5	0	1	1	2	0	4
	5	2	10	3	0	20

[191]	c1	c2	c3	c4	c5	
c1	7	0	5	1	0	13
c2	2	2	21	4	0	29
c3	2	7	48	15	0	72
c4	3	0	38	24	0	65
c5	0	1	7	4	0	12
	14	10	119	48	0	191