Konfidenzmaße im Dialogmanagement

Studienarbeit am

Institut für Logik, Komplexität und Deduktionssysteme

Prof. Dr. A. Waibel

von

Anja Hauth
anja.hauth@gmx.de

Betreuer:
Prof. Dr. A. Waibel, Christian Fügen, Petra Gieselmann

Abgabe: 29. November 2004

Übersicht Ein Dialogsystem kann nur so gut sein, wie die vorhandenen Informationen, auf denen das Dialogsystem aufbaut. Ziel dieser Studienarbeit ist es dem Dialogsystem mit Hilfe von bekannten Konfidenzmaßen und maschinellen Lernmethoden ein Maß für die Qualität der Erkennung an die Hand zu geben. Durch die Bestimmung von falsch erkannten Wörtern soll der Diskurs erweitert werden.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1

Einführung

1.1 Zielsetzung

Ziel meiner Studienarbeit ist es, das Verständnis des Dialogsystems Tapas zur Steuerung eines Haushaltsroboters mit dem Namen Robbi zu verbessern. Robbi unterstützt bei der Arbeit im Haushalt. Die Grundvoraussetzung dafür, daß der Roboter die Anweisungen, welche an ihn gestellt werden, versteht, ist eine gute Spracherkennnung und ein gutes Sprachverstehen. Diese Studienarbeit beschäftigt sich mit dem Sprachverstehen.

1.1.1 Das Dialogsystem Tapas

Der Aufbau der Mensch- Maschine- Schnittstelle ist in Abbildung 1.1 dargestellt. Sie besteht aus einem Spracherkenner, welcher die gesprochenen Laute vom Menschen aufnimmt und einem Gestenerkenner, der die Bewegungen und Deutungen wahrnimmt. Über die Sprachhypothese läuft ein grammatischer Parser, welcher aus der wortbasierten Repräsentation eine semantische Darstellung erzeugt.

Aus der semantischen Repräsentation des Gesprochenen, den Gesteninformationen und seinem Umweltmodell, gehalten in einer Datenbank, bestimmt der Roboter Robbi seine Reaktion. Das Umweltmodell enthält Informationen über das Arbeitsumfeld des Roboters, die Küche und den Verlauf des Gespräches.

Die Reaktion des Roboters wird durch Synthese der Antwort und durch Taten bestimmt. Durch die Sprachsynthese kann der Roboter Nachfragen stellen und seine verstandene Hypothese bestätigen.

1.2 Herausforderungen

Diese Studienarbeit beschäftigt sich mit der Frage, wie aus dem Worthypothesengraphen aussagekräftige Informationen für das Dialogsystem extrahiert werden können. Die Erkennung der Hypothese durch den Spracherkenner erfolgt mittels einem Hidden Markov Modells. Mit Hilfe des Hidden Markov Models wird ein Worthypothesengraphen aufbaut. Unter einem Worthypothesengraphen versteht man einen Ergebnisgraphen, welcher alle wahrscheinlichen Sätze abdeckt.

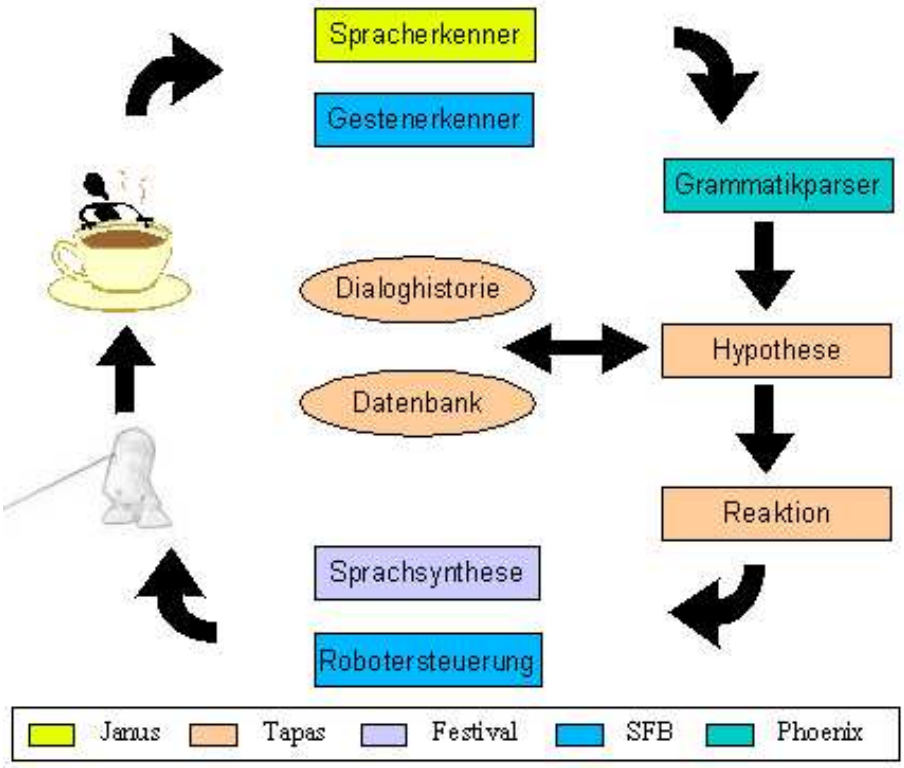


Abbildung 1.1: Hier sieht man den Aufbau und die Integration des Dialogsystems Tapas. Das komplette System stellt die Schnittstelle zwischen Mensch und Roboter dar. Das Herzstück des Dialogsystems besteht in der Bestimmung der korrekten Hypothese. Diese soll die Intention des Menschen in einer semantischen, maschinenverständlichen Repräsentation korrekt erfassen. Aus der semantischen Repräsentation wird die Reaktion abgeleitet, die vom Menschen erwartet wurde.

Aus diesem Ergebnisgraphen wird zur Zeit die wahrscheinlichste Hypothese extrahiert und als Grundlage für die Erkennung der semantischen Repräsentation verwendet. Ist aber die Hypothese des Spracherkenners nicht korrekt erkannt, so wird es für den semantischen Parser schwierig bis unmöglich die richtige semantische Repräsentation der Anfrage durch den Benutzer zu finden.

In dieser Studienarbeit soll analysiert werden, in welcher Form dem semantischen Parser zusätzliche Informationen zur Verfügung gestellt werden kann, um zu entscheiden, welches Vertrauen in die Hypothese gesetzt werden kann. Auf dieser Information kann folgende Verarbeitung aufsetzen:

1. Die beste Hypothese des Spracherkenners kann anhand eines weiteren Kriteriums ausgewählt werden. Die beste Hypothese bestimmt sich nicht mehr nur aus der höchsten Wahrscheinlichkeit, welche durch den Spracherkenner bestimmt wurde, sondern auch aus dem Vertrauen, welches das Dialogsystem in die Hypothese hat.

2. Konnte auf der Grundlage des Kontextes kein sinnvoller Parse aus der Hypothese bestimmt werden, so kann analysiert werden, welcher Satzteil falsch erkannt wurde. Das Dialogsystem kann dann gezielt nach der sich dahinter versteckenden Information beim Anwender nachfragen.

Wort	Vertrauen	a posteriori Score
switch	0.999990	5.34
on	0.894773	4.97
that	0.395221	2.45
and	0.393012	9.59
robbi	0.999961	8.48
Gesamt	0.736591	30.83

Tabelle 1.1: Vertrauen in die Wörter, welche vom Spracherkenner unterstützt werden

Wort	Vertrauen	a posteriori Score
switch	0.999990	5.34
on	0.894773	4.97
the	0.316853	5.31
lamp	0.506931	9.89
robbi	0.999961	8.48
Gesamt	0.743702	33.99

Tabelle 1.2: Vertrauen in die Wörter, welche vom Vertrauen unterstützt werden

Auf diese zwei Aspekte soll an dieser Stelle noch ein wenig näher anhand eines Beispieles eingegangen werden.

1.2.1 Extraktion der Sprachhypothese

Das Vertrauen in eine Hypothese ist ein zusätzliches Kriterium, auf Grund welchem das Dialogsystem seine Wahl der Hypothese des Spracherkenners aus den Hypothesen des Worthypothesengraphen treffen kann. Wie dies von Nutzen sein kann, soll das nachfolgende Beispiel veranschaulichen.

Der Spracherkenner liefert den in der Abbildung 1.2 dargestellten Worthypothesengraphen. Aus dem Worthypothesengraphen sollen zwei Hypothesen näher betrachtet werden:

Hypothese basierend auf dem Score des Spracherkenners:
Switch on that and Robbi.
(*Schalte dies an und Robbi.*)

Hypothese basierend auf dem Vertrauen in die Hypothese:
Switch on the lamp Robbi.
(*Schalte das Licht an Robbi.*)

Ein weit verbreitetes Verfahren bei der Wahl der besten Hypothese beruht auf dem a posteriori Score, welche mit Hilfe eines Hidden Markov Modells bestimmt wird. Die Werte für das Vertrauen und dem a posteriori Score sind in den Tabellen 1.1 und 1.2 dargestellt.

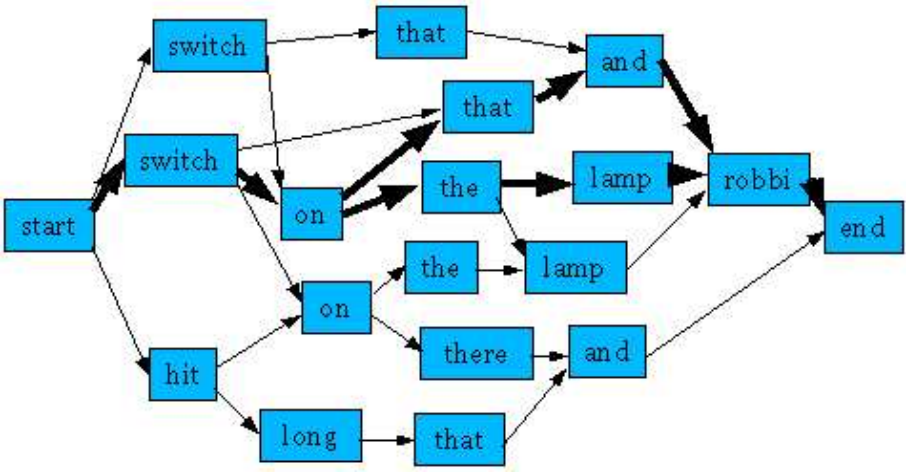


Abbildung 1.2: Die Abbildung zeigt einen Worthypothesengraphen des Spracherkenners. Die Knoten enthalten die möglichen Wörter. Die Kanten enthalten die Übergangswahrscheinlichkeitenen zwischen den Wörtern. Die zwei näher betrachteten Hypothesen sind im Worthypothesengraphen hervorgehoben.

Beim Lesen der Tabelle ist zu beachten, daß die Konfidenz zwischen 0 und 1 liegt. Je größer die Konfidenz des erkannten Wortes ist, desto mehr Vertrauen wird in dieses Wort gesetzt. Beim a posteriori Score handelt es sich um eine Wertangabe, welche aus der a posteriori Wahrscheinlichkeit bestimmt wird durch:

$$\text{Score} = -log(\text{Wahrscheinlichkeit})$$

Damit schwankt der Betrag des Wertes zwischen 0 und unendlich. Ein hoher a posteriori Score spricht für ein a posteriori unwahrscheinlicheres Wort gegeben der akustischen Beobachtung.

Betrachtet man ausschließlich den a posteriori Score, wie es zur Zeit üblich ist, so erreicht die erste Hypothese eine geringere a posteriori Bewertung als die zweite. Die Wahl der zu verarbeitenden Hypothese würde auf die vom Spracherkenner empfohlene Hypothese fallen. Betrachtet man aber das Vertrauen in die erkannten Wörter, dann fällt die Entscheidung auf die zweite Hypothese, weil für diese Hypothese ein höheres Vertrauen vorhanden ist.

Wählt man im einfachsten Fall eine passende Gewichtung zwischen a posteriori Score und Vertrauen, so könnte man die Wahl der zweiten Hypothese bewirken, welche für das Dialogsystem bessere Informationen enthält.

1.2.2 Klärungsdialog

Gibt es im Dialogsystem Zweifel über eine bestimmte semantische Hypothese, dann ist es die Aufgabe des Klärungsdialogs, gezielte Fragen an den Benutzer zu stellen, um die Zweifel zu beseitigen. Das Vertrauen könnte ein Kriterium sein, anhand dessen der Dialogmanager Zweifel bekunden könnte.

An einem Beispiel läßt sich das am Besten veranschaulichen. Der Spracherkenner stellt folgende Hypothese zur Verfügung:

Hypothese:
Switch on that and Robbi.
(*Schalte dies an und Robbi.*)

Aus dieser Hypothese erkennt das Dialogsystem, daß es sich um das Dialogziel 'act_switch' handelt, welches keine Informationen darüber spezifiziert hat, was angeschaltet werden soll.

```
act_switch
        robbi:ONOFF [ robbi:prp_onoff
                robbi:BOOL [ true ]
        ]
```

Also wird es die Gesteninformationen des Gestenerkenners prüfen. Diese enthalten auch keine relevanten Informationen. Um die Anfrage des Benutzers trotzdem bearbeiten zu können, geht das Dialogsystem zur Zeit davon aus, daß das Objekt, welches angeschaltet werden soll, falsch verstanden oder nicht spezifiziert wurde und fragt danach nach. Aber genauso gut hätte das Dialogziel nicht korrekt erkannt worden sein können.

Was falsch erkannt wurde, kann jetzt anhand der Vertraunenswerte der einzelnen Wörter bestimmt werden. Es ergeben sich die Werte, welche in der Tabelle 1.1 abgebildet sind.

Es ist deutlich zu erkennen, daß die Wörter 'that' und 'and' die geringsten Vertrauenswerte haben. Dies bestätigt die Annahme, daß das Referenzobjekt falsch erkannt wurde. Anhand dieser Information kann der Worthypothesengraph des Spracherkenners untersucht werden. Das System wird gezielt nach alternativen Wörtern suchen und 'the lamp' (*die Lampe*) finden.

Sollte der Worthypothesengraph keine Anhaltspunkte enthalten, dann kann mit Hilfe des Klärungsdialogs die Aufgabe verfeinert werden. Eine mögliche Nachfrage könnte in der Art erfolgen:

Antwort:
What may I switch on for you?
(*Was kann ich für Sie anschalten?*)

Antwortet der Benutzer mit 'the lamp' (*die Lampe*), so weiß das Dialogsystem, daß es die Lampe anschalten soll. Der Spracherkenner wird die Hypothese richtig erkennen, weil die Anwort kurz und prezise formuliert werden kann. Die korrekte semantische Hypothese ist:

```
act_switch
        robbi:ONOFF [ robbi:prp_onoff
                robbi:BOOL [ true ]
        ]
        robbi:OBJ [ robbi:obj_lamp
                robbi:SG [ false ]
        ]
```

1.3 Herangehensweise

1.3.1 Forschung

Schon vor mir haben sich einige Forschungsgruppen mit dem Thema der Vertrauensbestimmung beschäftigt. In [Kemp97] wird ein Vertrauensmaß vorgestellt, welches dazu verwendet wird die Erkennungsleistung des Spracherkenners zu verbessern.

Auch im Dialogsystem wurden bereits Erfahrungen mit Vertrauensmaßen gesammelt. Die Publikationen von [SanSegundo01], [Bohus02] und [Hatzen00] zeigen einen Ausschnitt aus den aktuell erreichten Ergebnissen.

[SanSegundo01] verwendet für seine Untersuchungen ausschließlich neuronale Netze. Diese werden benutztF, um falsch erkannte Wörter, Äußerungen oder Konzepte zu erkennen. Auf Wortebene wurde bestimmt, wie gut ein Wort erkannt wurde. Auf dieser Grundlage wurde aus dem Worthypothesengraphen eine passende Hypothese extrahiert. Damit wurde eine Verbesserung von $6,2\%$ erreicht. Auf der Ebene der einzelnen Äußerungen interessiert man sich für die korrekte Bestimmung des Bereiches der Aussage. Damit wurde eine Verbesserung von $4,2\%$ durch Erkennung von falsch erkannten Themenbereichen erreicht. Die Konzeptebene beschäftigt sich mit der korrekten Erkennung von sinntragenden Wörtern. Durch das Ignorieren von Füllworten konnte eine Verbesserung von $13,5\%$ erreicht werden.

[Hatzen00] entwickelte ein Bewertungssystem, welches die Erkennungsleistung des Dialogsystems stark verbesserte. Sein Ziel war es den Spracherkenner JUPITER dahingehend zu erweitern, daß die Qualität des Dialogsystems gesteigert werden konnte. Dies wurde erreicht durch die Zurückweisung von falsch erkannt Hypothesen und die Verwendung der Konfidenz zur Initialisierung des Klärungsdialoges. So konnten 35% der falsch erkannten Hypothesen erkannt werden und es wurden 5% der Klärungsdialoge sinnvoller eingesetzt, d.h. kein Klärungsdialog initiiert, wenn die Hypothese bereits korrekt verstanden wurde oder ein Klärungsdialog gestartet bei einer falsch verstandenen Hypothese.

Die Publikation von [Bohus02] erreicht mit Hilfe von maschinellen ein robusteres Dialogsystem. Die Robustheit des Dialogsystems wird durch Zurückweisen von falsch erkannten Hypothesen erreicht. Die verwendeten Lernverfahren waren Bayes Netzwerke, Boosting, Entscheidungsbäume, neuronale Netze, Support Vektor Lernen und naive Bayes Klassifikatoren. Die Lernverfahren erreichen vergleichbare Ergebnisse und verbessern das bestehende System um bis zu $49,5\%$.

1.3.2 Idee

In dieser Studienarbeit soll im ersten Teil eine kurze Einführung in die für das Verständnis der Studienarbeit wichtigen Grundlagen gegeben werden. Hier werden kurz verschiedene Lernverfahren vorstellt, welche später für die Klassifikation verwendet werden und auf die wichtigen Kriterien und Kenngrößen eingegangen

Der zweite Teil zeigt die Herangehensweise auf, welche verwendet wurde, um ein gutes Vertrauensmaß zu finden. Es werden bekannte Verfahren zur Bestimmung des Vertrauens in erkannte Wörter vorgestellt und die Umsetzung verschiedener Lernverfahren um aussagekräftige Aussagen treffen zu können. Gleichzeitig werden die erzielten Ergebnisse für die einzelnen Verfahren der Ver-

trauensbestimmung beschrieben und der Erfolg aufgezeigt, welcher durch den Einsatz des Vertrauens in das bestehende Dialogsystem Tapas erreicht werden kann.

Teil I

Grundlagen

Kapitel 2

Entscheidungsbäume

2.1 Aufbau

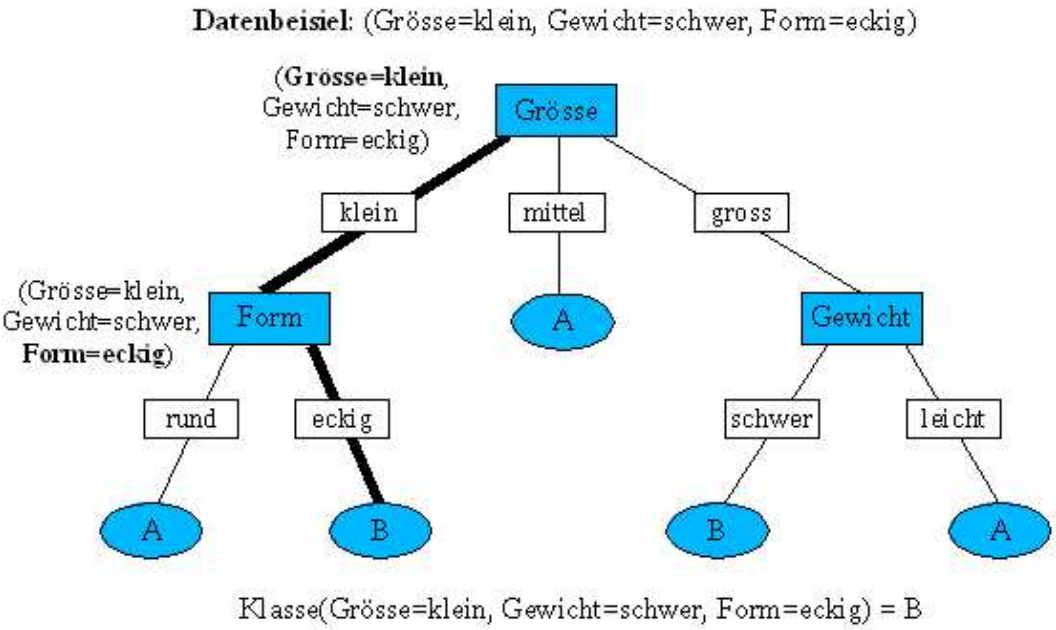


Abbildung 2.1: Die Knoten des Entscheidungsbaums sind als blaue Rechtecke dargestellt und enthalten die Kriterien nach denen die Daten getrennt werden. Die weißen Rechtecke stellen die verschiedenen Werte des Kriteriums dar. Die Blätter sind als blaue Ovale dargestellt und geben die Klassenzugehörigkeit der Daten an.
Möchte man ein Datenbeispiel mit den Eigenschaften (Größe = klein, Gewicht=schwer, Form=eckig) auswerten, so prüft man am Wurzelknoten das Kriterium 'Größe'. Da 'Größe' den Wert 'klein' hat, verzweigt man in den linken Zweig. Dort wird das Kriterium 'Form' als 'eckig' ausgewertet. Damit kommt man in den rechten Zweig, welcher auf ein Blatt zeigt. Aus diesem Blatt kann die Klasse B abgelesen werden.

Bei den Entscheidungsbäumen [Zöllner03], [Mitchell97] werden Daten in einer Baumstruktur dargestellt. In den Knoten werden die Daten anhand eines Kriteriums unterteilt. In den Blättern befindet sich die den Kriterien zugeord-

nete Klasse. Damit ergibt sich die in Abbildung 2.1 dargestellte Struktur.

2.2 Auswertung

Die Auswertung eines neuen Datenbeispieles erfolgt durch Traversierung des Entscheidungsbaums. An jedem Knoten wird das angegebene Kriterium für das Datenbeispiel untersucht. Die Auswertung bestimmt die Verzweigung in den entsprechenden Pfad. Dieses Vorgehen erfolgt bis man sich in einem Blatt befindet, von dem man die Klasse, welche dem Datenbeispiel zugeordnet wird, ablesen kann.

Wie man an dem in Abbildung 2.1 gezeigten Beispiel sehen kann, müssen nicht alle Eigenschaften eines Datenbeispiels für die Zuordnung zu einer Klasse genutzt werden. Trotzdem ist eine notwendige Voraussetzung für die Verwendung eines Entscheidungsbaumes, das für alle zu klassifizierenden Daten die verwendeten Kriterien im Entscheidungsbaum bekannt sind.

Mit dem beschriebenen Vorgehen kann allen Datenbeispielen eine Klasse zugeordnet werden. Damit diese Klasse auch für unbekannte Datenbeispielen möglichst korrekt und einfach zu bestimmten ist, sollte der Baum möglichst optimal aufgebaut werden. Ein Entscheidungsbaum ist optimal, wenn die Generalisierungsfähigkeit gegeben ist und bestimmte Gefahren wie Overfitting berücksichtigt wurden.

Kapitel 3

Neuronale Netze

Neuronale Netze wurden vom Aufbau des menschlichen Gehirns inspiriert. Ziel war es ein Netz aus Neuronen aufzubauen, welche über Synapsen so verbunden werden, so daß bei einem Eingangsreiz der entsprechende Ausgangsreiz hervorgerufen wird.

Die neuronalen Netze, welche sich auf Grund dieser Überlegung entwickelt haben, haben nicht mehr viel mit dem menschlichen Gehirn gemeinsam: das menschliche Gehirn hat eine viel komplexere Verschaltung und leitet Reize viel langsamer weiter. Aber mit den künstlichen neuronalen Netzen ist eine weit verbreitete Lernmethode entstanden, mit der viele Probleme aus unterschiedlichen Bereichen, wie Kursvorhersage für die Börse oder Steuerung eines Roboterarmes, gelöst worden sind.

3.1 Aufbau

Ein neuronales Netz [Zöllner03], [Mitchell97], [Callan03], [Waibel04] besteht aus vielen Neuronen. Ein Neuron hat mehrere Eingänge, welche gewichtet werden. Eine Propagierungsfunktion, welche im Kapitel Auswertung näher besprochen wird, entscheidet darüber, ob ein Neuron ein Signal weiterleitet oder nicht. Den Aufbau eines Neurons kann man in Abbildung 3.1 sehen.

Aus diesen Neuronen können beliebige Netzstrukturen aufgebaut werden. Die elementarste Netzstruktur besteht aus einer Eingabeschicht, mehreren versteckten Schichten und einer Ausgabeschicht, wie in Abbildung 3.2 dargestellt. Jedes Neuron in der Vorgängerschicht ist mit allen Neuronen in der nachfolgenden Schicht verbunden. Damit ist nur eine vorwärtsgerichtete Beeinflussung von Neuronen möglich. Darum heißen diese Art von neuronalen Netzen auch vorwärtsgerichtete neuronale Netze.

Aber auch andere Strukturen werden in der Praxis eingesetzt, wie zum Beispiel rückwärtsgerichtete neuronale Netze, bei denen auch die Möglichkeit vorgesehen ist, dass die Verschaltung in vorher durchlaufende Schichten möglich ist. Auf dies Möglichkeit soll an dieser Stelle nicht eingegangen werden, da nur ein Überblick über die verschiedenen Lernmethoden geben werden soll.

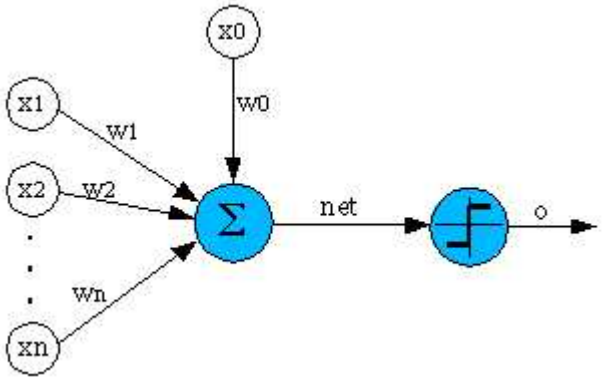


Abbildung 3.1: Ein Neuron bekommt verschiedene Eingaben x_i von anderen Neuronen oder der Eingabe des neuronalen Netzes. Die Eingaben werden mit den Gewichten w_i gewichtet. Die Eingabe x_0 bezeichnet man als Threshold. Der Threshold bestimmt die Schranke, ab welcher das Neuron schalten soll. Das zugehörige Gewicht w_0 ist immer 1.
Die gewichtete Eingabe bezeichnet man als 'net'. Aus dem 'net' wird mittels der Propagierungsfunktion die Ausgabe o bestimmt. In der Abbildung sieht man eine Stufenfunktion. Die Ausgabe des Neurons kann als Eingabe für weitere Neuronen verwendet oder als Ergebnis des neuronalen Netzes interpretiert werden.

3.2 Auswertung

Die Auswertung eines neuronalen Netzes erfolgt mittels Propagierung. Die Propagierung besteht aus fünf Schritten:

1. Die Eingaben eines neuronalen Netzes müssen in die richtige Form für das neuronale Netz gebracht werden. Eingaben eines Netzes müssen Dezimalzahlen sein, am besten zwischen 0 und 1. Durch die Wahl eines ähnlichen Intervalls für alle Eingaben, wird das Training der neuronalen Netze vereinfacht. Für die Umsetzung der Eingaben in Dezimalwerte wurden verschiedene Methoden entwickelt. Auf diese soll an dieser Stelle nicht weiter eingegangen werden, da sie im Rahmen der Bestimmung eines Vertrauensmaßes der Sprachhypothese nicht nötig sind.

2. Jedes Neuron hat mehrere Eingaben. x_{ji} beschreibt den i-ten Eingabewert für das Neuron j und das Gewicht w_{ji} beschreibt das zugehörige Gewicht zum Eingang x_{ji}. Daraus ergibt sich die gewichtete Summe für das j-te Neuron, welche mit 'net' bezeichnet wird, mit $net_j = \sum_i w_{ji}x_{ji}$.

3. Aus der gewichteten Summe wird mittels Propagierungsfunktion die Ausgabe bestimmt. Häufige Propagierungsfunktionen werden in der Abbildung 3.3 gezeigt. Die am weitesten verbreitete Propagierungsfunktion ist die Sigmoidfunktion $\sigma(net) = \frac{1}{1+e^{-net}}$, welche eine Ausgabe zwischen 0 und 1 erzeugt.

4. Die Schritte zwei und drei werden durch alle Schichten hindurch propagiert. Die Propagierung erfolgt Schicht für Schicht. Die Ausgabe der vorhergehenden Schicht bildet die Eingabe für die nächste. Die Ausgabeschicht ist die letzte Schicht, welche die entgültige Ausgabe des neuronalen Netzes erzeugt.

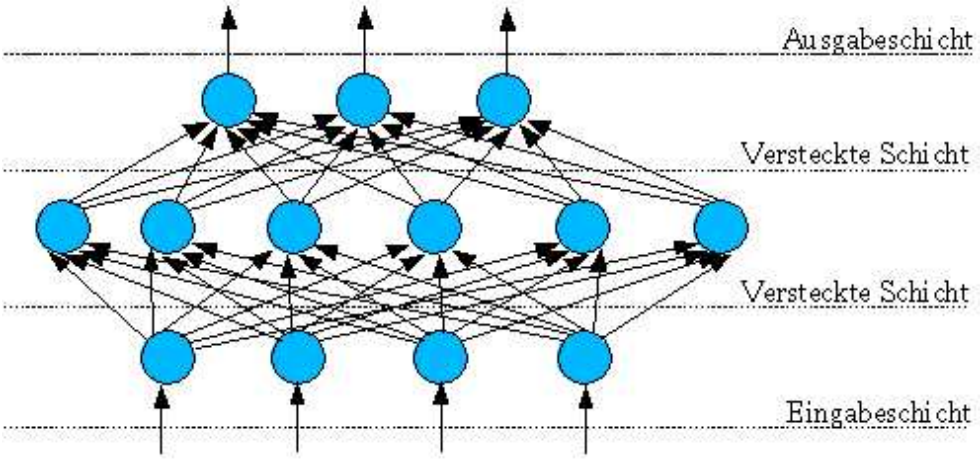


Abbildung 3.2: In der Abbildung ist ein vorwärtsgerichtetes neuronales Netz mit zwei versteckten Schichten dargestellt. Es gibt vier Eingaben, welche in drei Ausgaben vom neuronalen Netz umgesetzt werden. Wie man sieht werden die Ausgaben der vorhergehenden Schicht als Eingabe dieser Schicht verwendet.

5. Der letzte Schritt besteht in der Umsetzung der Ausgabewerte des neuronalen Netzes in das interpretierte Ergebnis. So können mehrere Neuronen verwendet werden, wobei jedes Neuron für eine andere Klasse steht. Dann entspricht die Klasse der Eingabedaten der Klasse des aktivsten Neurons. Eine andere Möglichkeit liegt in der Festlegung bestimmter Ausgabebereiche eines Neurons für die verschiedenen Klassen.

Ein Beispiel für die Struktur eines neuronalen Netzes sieht man in Abbildung 3.2.

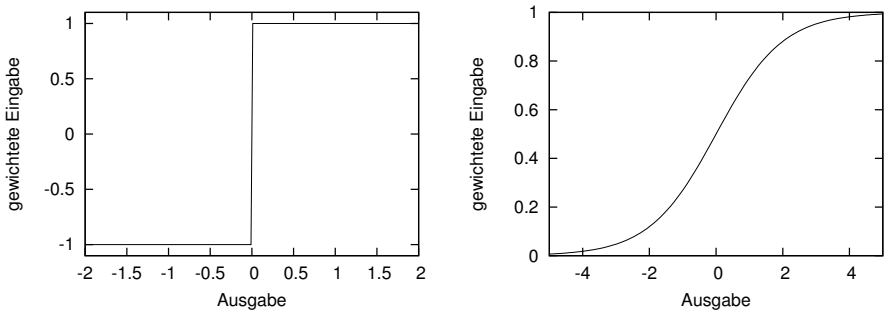


Abbildung 3.3: In dem linkem Bild sieht man ein Stufenfunktion. Eine Stufenfunktion gibt für alle Werte kleiner als 0 den Wert -1 aus und für alle Werte größer als 0 den Wert 1. Da die Stufenfunktion nicht stetig ist, ist das Integral nicht definiert und damit ein Netz mit Stufenfunktion schwer zu trainieren.
Im rechten Bild sieht man eine Sigmoidfunktion. Diese konvergiert für $-\infty$ gegen 0 und für ∞ gegen 1. Da die Sigmoidfunktion stetig und leicht zu integrieren ist, hat sich die Sigmoidfunktion durchgesetzt und wird hauptsächlich für neuronale Netze verwendet.

Kapitel 4

Support Vector Machines

4.1 Aufbau

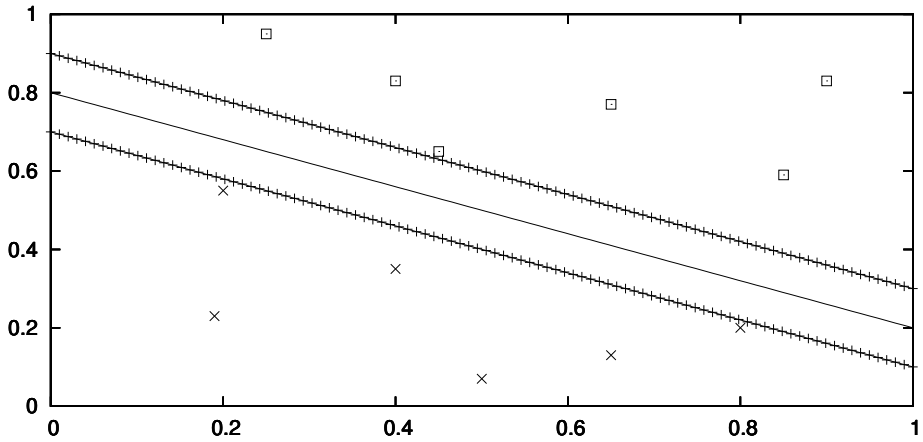


Abbildung 4.1: In der Abbildung ist die Lösung für ein Klassifikationsbeispiel dargestellt. Es gibt zwei Klassen, welche durch eine Gerade linear getrennt werden. Dabei wurde die Gerade so gewählt, daß die Support Vektoren alle den gleichen Abstand von der Gerade haben. Diesen Abstand nennt man Margin.
Die Support Vektoren befinden sich am Rand der Margin. Da diese Beispiele allein den Verlauf der Trenngeraden bestimmen, werden die restlichen Beipieldaten nicht für die weitere Klassifikation verwendet.

Support Vector Machines [Zöllner03], [Vapnik00], [Schölkopf00], [Cristianini00] stellen eine relativ neue Klassifikationsmethode dar, welche von Vapnik entwickelt wurde. Die Lernmethode hat ihren Namen von den sogenannten Support Vektoren, welche die Grenzfälle für die Klassen darstellen.

Im Gegensatz zu neuronalen Netzen wird bei den Support Vector Machines nicht eine mögliche Trennung gesucht, sondern die Hyperebene, welche die Klassen trennt und einen möglichst großen Abstand von allen Trainingsdaten hat. Die Vektoren, welche die optimale Hyperebene bestimmen, werden als Support Vektoren bezeichnet.

Die ursprünglichen Support Vector Machines konnten Daten nur linear tren-

nen. Durch Verfahren, welche in diesem Kapitel kurz vorstellt werden, konnte aber auch eine nichtlineare Trennung der Trainingsdaten erreicht werden.

4.2 Auswertung

Die Trenngerade ist gegeben durch folgende Gleichung:

$$f(x, \alpha, b) = \sum_{i=1}^{l} y_i \alpha_i < x_i, x > +b \tag{4.1}$$

$$= \sum_{i \in SV} y_i \alpha_i < x_i, x > +b \tag{4.2}$$

Dabei ist x die neu zu klassifizierende Beobachtung. α_i nimmt nur einen von 0 verschiedenen Wert an, wenn x_i ein Support Vektor ist. Damit können bei der Bestimmung von $f(x, \alpha, b)$ nur die Support Vektoren berücksichtigt werden, was den Rechenaufwand und den Speicheraufwand bei der Klassifikation erheblich verbessert.

Die Trenngerade verwendet einen Kernel $< \cdot, \cdot >$. Die Kernelfunktion wird einmalig gewählt und beeinflußt die Eigenschaften des Klassifikators erheblich. Im einfachsten Fall, für linear trennbare Daten, kann das Standardskalarprodukt verwendet werden.

Die Bestimmung von α_i ist Aufgabe des Trainings. Der y_i Wert gibt die zugehörige Klasse zum Support Vektor an. Aus den gegebenem α_i Werten kann b bestimmt werden als:

$$b = -\frac{max_{y_i=-1}(< w, x_i >) + min_{y_i=1}(< w, x_i >)}{2} \tag{4.3}$$

$$w = \sum_{i=1}^{l} y_i \alpha_i x_i \tag{4.4}$$

Um die Klasse für eine Beobachtung zu bestimmen, setzt man die Werte der Beobachtung in die Gleichung (4.2) ein. Ist das Ergebniss von $f(x, \alpha, b)$ positiv, so gehört die Beobachtung zu der Klasse oberhalb der Trenngraden, sonst zu der anderen Klasse.

Kapitel 5

Bayes Klassifikation

Die Grundlage des Bayes Klassifikators [Zöllner03], [Mitchell97] bildet die Statistik, da der Bayes Klassifikator Grundlage der Trainingsdaten ein statistisches Modell für die Umwelt aufbaut.

5.1 Aufbau

Der Bayes Klassifikator setzt auf der Bayes Regel auf.

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)} \tag{5.1}$$

Die Bayes Regel bestimmt die Wahrscheinlichkeit für die Klasse y_i, wenn der Datensatz x beobachtet wird. Diese Wahrscheinlichkeit bezeichnet man als die a posteriori Wahrscheinlichkeit $P(y_i|x)$ von y_i. Sie ist abhängig von $P(x|y_i)$, der klassenbedingten Wahrscheinlichkeit, von $P(y_i)$ der a priori Wahrscheinlichkeit der Klasse y_i und von $P(x)$ der Wahrscheinlichkeit für die Beobachtung x.

Die Wahrscheinlichkeitsverteilung $P(x|y_i)$ legt fest, wie wahrscheinlich eine Beobachtung x in der Klasse y_i ist. Es ist klar, das die Wahrscheinlichkeit für diese Klasse größer wird, wenn die Beobachtung x für die Klasse y_i häufiger auftritt.

Die a priori Wahrscheinlichkeit $P(y_i)$ gibt an, wie häufig eine Klasse auftritt. Wenn eine Klasse y_i öfter auftritt, dann wird auch die Klasse für die gegebene Beobachtung x wahrscheinlicher.

Die Wahrscheinlichkeit $P(x)$ spielt beim Bayes Klassifikator keine Rolle, denn die Wahrscheinlichkeit ist unabhängig von der Klasse y_i und Ziel ist es ja die wahrscheinlichste Klasse für die Beobachtung x zu bestimmen.

5.2 Auswertung

Ziel einer jeden Klassifikation ist es eine Klasse für einen Datensatz zu bestimmen, welche möglichst die korrekte Klasse für den Datensatz ist. Darum ordnet der Bayes Klassifikator dem Datensatz die wahrscheinlichste Klasse gegeben seinem Modell zu.

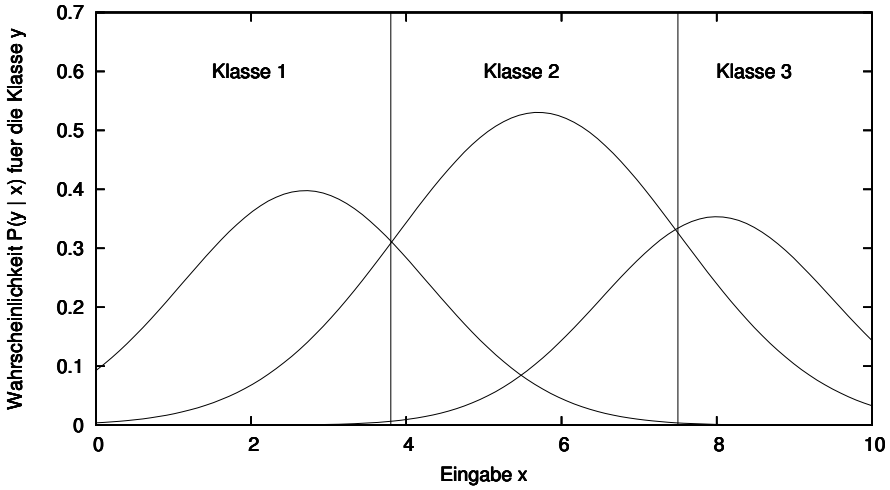


Abbildung 5.1: In der Abbildung sieht man die Verteilung der Wahrscheinlichkeiten der Klassen für verschiedene Beobachtungswerte x. Welche Klasse würde man also der Beobachtung 6 zuordnen? Die Wahrscheinlichkeit für die Klasse 1 liegt bei 0.05, die Wahrscheinlichkeit für die Klasse 2 bei 0.52 und für die Klasse 3 bei 0.14. Damit ist die wahrscheinlichste Klasse die Klasse 2, welche wir also der Beobachtung 6 zuordnen würden.
Diese Art der Zuordnung von Beobachtungen zu der wahrscheinlichsten Klasse bezeichnet man als das Maximum Likelyhood Verfahren. In unserem Beispiel würden allen Beobachtungen unter 3.8 der ersten Klasse, alle Beobachtungen zwischen 3.8 und 7.5 der Klasse 2 und alle darüber der Klasse 3 zuordnet. Die Grenzen sind in der Abbildung eingezeichnet.

Nach der Bayes Regel in Gleichung (5.1) ist die wahrscheinlichste Klasse gegeben, indem man das Maximum aller Klassen wählt. Damit ergibt sich für einen Datensatz x die Klasse y_i, für welche gilt:

$$y = argmax_{y_i \in Y} P(y_i|x) \tag{5.2}$$
$$= argmax_{y_i \in Y} \frac{P(x|y_i)P(y_i)}{P(x)} \tag{5.3}$$
$$= argmax_{y_i \in Y} P(x|y_i)P(y_i) \tag{5.4}$$
$$= argmax_{y_i \in Y} ln(P(x|y_i)) + ln(P(y_i)) \tag{5.5}$$

Wie bereits angekündigt, kann $P(x)$ ignoriert werden, weil die Wahrscheinlichkeit für die Beobachtung x für alle Klassen y_i gleich ist. Durch diese Umformung erhält man keine Wahrscheinlichkeit mehr, aber die wertmäßige Reihenfolge wird nicht geändert, welches das interessante Kriterium für das Maximum ist.

Die Gleichung wird durch den Logarithmus in einen anderen Raum transformiert, um einen Überlauf beim tatsächlichen Umgang mit den Wahrscheinlichkeiten zu vermeiden. Denn bei vielen möglichen Klassen und Beobachtungen, wie z.B. bei der Spracherkennung, welche miteinander multipliziert werden, kann man sich gut vorstellen, das es zu einem Unterlauf der Darstellung der Werte im Rechner kommen kann.

Für die Auswertung eines Datensatzes bleibt also die Frage zu beantworten, wie sich $P(x|y_i)$ und $P(y_i)$ bestimmen lassen. Die Bestimmung von $P(y_i)$ erfolgt durch folgende Formel:

$$P(y_i) = \frac{\#\text{ Beobachtungen für die Klasse } y_i}{\#\text{ Beobachtungen für alle Klassen } y \in Y} \tag{5.6}$$

Die Wahrscheinlichkeit $P(x|y_i)$ für die Beobachtung x in der Klasse y_i ist definiert als:

$$P(x|y_i) = P(x_1, x_2, \ldots, x_n|y_i) \tag{5.7}$$

$$= \frac{\#\text{ Beobachtungen } (x_1, x_2, \ldots, x_n) \text{ in der Klasse } y_i}{\#\text{ aller Beobachtungen } x \in X \text{ in der Klassen } y_i} \tag{5.8}$$

Man betrachtet also alle Trainingsdaten dieser Klasse, sucht nach der Beobachtung $x = (x_1, x_2, \ldots, x_n)$ und bestimmt die Häufigkeit, mit welcher diese Beobachtung in der Klasse aufgetreten ist. Man kann sich gut vorstellen, daß es bei komplexen Beobachtungen schwer ist ausreichend viele Trainingsdaten zu bekommen, damit eine aussagekräftige Aussage über das Auftreten der Beobachtungen gemacht werden kann.

Für die Bestimmung von $P(x|y_i)$ gibt es mehrere Ansätze, von denen im Folgenden zwei Ansätze vorgestellt werden sollen. Eine sehr vereinfachte, aber relativ effektive und erstaunlich gute Bestimmung von $P(x|y_i)$ ist der naive Bayes.

Ein parametrischer Ansatz, welcher entsprechend weniger Trainingsdaten benötigt, ist die Normalverteilung. Der parametrische Ansatz kann mit dem naiven Bayes kombiniert werden, um mit kontinuierlichen Parametern umgehen zu können.

5.2.1 Naiver Bayes

Der naive Bayes 5.1 trifft die Annahme, daß die Wahrscheinlichkeitsverteilung für die Werte im Beobachtungstuppel $(x_1, x_2, \ldots, x_n)$ voneinander unabhängig sind. Damit kann folgende Aussage getroffen werden:

$$P(x|y_i) = P(x_1, x_2, \ldots, x_n|y_i) \tag{5.9}$$

$$= \prod_{j=1}^{n} P(x_j|y_i) \tag{5.10}$$

Wie man schön sieht, bestimmt sich jetzt die Wahrscheinlichkeit der Beobachtung $(x_1, x_2, \ldots, x_n)$ für die Klasse y_i aus dem Produkt der Wahrscheinlichkeiten der Einzelbeobachtungen x_j für die Klasse y_i. Es ist klar, daß man mehr Datensätze für eine Einzelbeobachtung findet, als für die Gesamtbeobachtung.

5.2.2 Normalverteilung

Im Gegensatz zu dem vorher vorgestellten Verfahren zur Bestimmung der klassenbedingten Wahrscheinlichkeit, muß man bei der Normalverteilung 5.2 nicht die Annahme treffen, daß die Werte im Beobachtungstuppel $(x_1, x_2, \ldots, x_n)$ voneinander unabhängig sind. Die Annahme würde aber die Bestimmung die Kovarianzmatrix erleichtern, da nur die Werte für eine Diagonalmatrix bestimmt werden müssen.

Klasse NaiverBayes(Datensatz, Klassen, statistisches Modell)
maximale-Wahrscheinlichkeit $\Leftarrow 0$
Für alle Klassen $< y_i >$
$P(y_i) \Leftarrow \frac{statistischesModell \Rightarrow Beobachtungen(y_i)}{statistischesModell \Rightarrow Beobachtungen(alle)}$
Für alle Teilbeobachtungen x_j des Datensatzes $(x_1, x_2, \ldots, x_n)$
$P(x_j|y_i) \Leftarrow \frac{statistischesModell \Rightarrow Beobachtungen(x_j, y_i)}{statistischesModell \Rightarrow Beobachtungen(y_i)}$
$P(x|y_i) \Leftarrow P(x|y_i) * P(x_j|y_i)$
$P(y_i|x) \Leftarrow ln(P(y_i) + ln(P(x|y_i))$
Wenn maximale-Wahrscheinlichkeit $< P(y_i|x)$
maximale-Wahrscheinlichkeit $\Leftarrow P(y_i|x)$
Klasse $\Leftarrow y_i$
return Klasse

Tabelle 5.1: Um einen Datensatz mittels des naiven Bayes zu klassifizieren, benötigt man ein statistisches Modell. Man geht alle Klassen durch und berechnet die Wahrscheinlichkeit für jede Klasse. Die Klasse mit der größten Wahrscheinlichkeit merkt man sich. Sie ist abhängig von der a priori Wahrscheinlichkeit $P(y_i)$, welche angibt wie häufig die Klasse auftritt und der a posteriori Wahrscheinlichkeit $P(x|y_i)$, welche die Wahrscheinlichkeit für x innerhalb der Klasse y_i angibt.
Der angegebene Algorithmus bestimmt die Wahrscheinlichkeit für alle Klassen y_i nach den für den naiven Bayes beschrieben Formeln. Häufig findet eine weitere Vereinfachung des Algorithmuses statt, in dem man eine weitere Annahme trifft. Die Annahme geht davon aus, daß alle Klassen gleich wahrscheinlich sind. Damit fällt die Bestimmung der a priori Wahrscheinlichkeit weg.

Klasse Normalverteilung(Datensatz, Klassen, statistisches Modell)
maximale-Wahrscheinlichkeit $= 0$
Für alle Klassen $< y_i >$
$P(y_i) \Leftarrow \frac{statistischesModell \Rightarrow Beobachtungen(y_i)}{statistischesModell \Rightarrow Beobachtungen(alle)}$
$P(x|y_i) \Leftarrow$ bestimme Normalverteilung(
statistisches Modell $\Rightarrow$ Mittelwert(y_i),
statistisches Modell $\Rightarrow$ Kovarianzmatrix(y_i),
x_j)
$P(y_i|x) \Leftarrow ln(P(y_i) + ln(P(x_j|y_i))$
Wenn maximale-Wahrscheinlichkeit $< P(y_i|x)$
maximale-Wahrscheinlichkeit $\Leftarrow P(y_i|x)$
Klasse $\Leftarrow y_i$
return Klasse

Tabelle 5.2: Das statistische Modell für die Normalverteilung ist durch die Mittelwerte und die Kovarianzmatrizen für die verschiedenen Klassen y_i bestimmt. Um die wahrscheinlichste Klasse zu bestimmen, berechnet man die a priori Wahrscheinlichkeit $P(y_i)$ und die klassenbedingte Wahrscheinlichkeit $P(x|y_i)$ für alle Klassen.
Die klassenbedingte Wahrscheinlichkeit kann einfach durch einsetzen des Mittelwertvektors und der Kovarianzmatrix in die Gleichung für die Normalverteilung bestimmt werden.

Allgemein bestimmt man den Mittelwert und die Kovarianzmatrix nach folgenden Formeln:

$$\bar{\mu} = \frac{1}{N} \sum_{k=1}^{N} \bar{x_k} \tag{5.11}$$

$$\Sigma = \frac{1}{N} \sum_{k=1}^{N} (\bar{x_k} - \bar{\mu})(\bar{x_k} - \bar{\mu})^T \tag{5.12}$$

Die klassenbedingte Wahrscheinlichkeit kann dann nach der folgenden Formel berechnet werden.

$$p(x_j|y_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{d}{2}}} e^{\left(-\frac{1}{2}(\bar{x}-\bar{\mu})^t \Sigma^{-1} \bar{x}-\bar{\mu})^t\right)} \tag{5.13}$$

Die Verwendung der Normalverteilung ermöglicht mit kontinuierlichen Werten umzugehen und die Anzahl der benötigten Trainingsbeispiele zu verringern. Diese Vorteile werden durch die Annahme der Normalverteilung der Daten erreicht. Durch den parametrischen Ansatz sollte die unterstellte Verteilung auch auf die Daten passen, damit korrekte Ergebnisse erzielt werden können.

Kapitel 6

Hidden Markov Modelle

6.1 Aufbau

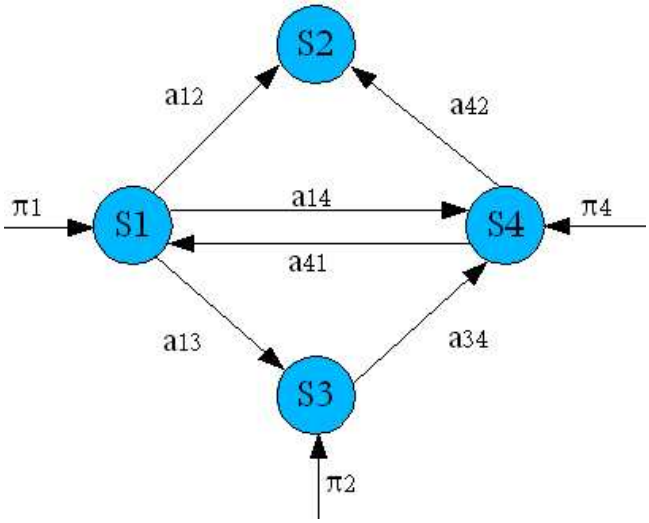


Abbildung 6.1: In der Abbildung sieht man ein mögliches Hidden Markov Modell. Zustandsübergänge, welche den Wert $a_{ij} = 0$ haben, sind der Übersichtlichkeit halber nicht mit eingezeichnet.
Die Struktur eines Hidden Markov Modells kann beliebig gewählt werden. Sie weist aber meist eine Struktur auf wie vorwärtsgerichtete Netze, welche keinen Rücksprung zu einem einmal verlassenen Zustand zulassen. Dabei sind Zustandsübergänge auf sich selber a_{ii} erlaubt.

Ein Hidden Markov Modell [Zöllner03], [Mitchell97], [Fink03] ist eine statistische Lernmethode, welche ein Netz aufbaut, das überwacht trainiert wird. Ein Netz besteht wie in der Abbildung 6.1 dargestellt aus:

- **Zuständen** $\{S_1, \ldots, S_n\}$, welche die möglichen Zustände abbilden, die eine Beobachtung annehmen kann
- **Zustandsübergänge** a_{ij}, welche die möglichen Wechsel zwischen den Zuständen beschreiben. Der Wert von a_{ij} gibt die Wahrscheinlichkeit an, mit welcher eine Beobachtung vom Zustand S_i in den Zustand S_j wechselt.
- **Initiale Zustandswahrscheinlichkeiten** π_i geben die Wahrscheinlichkeit an, mit welcher der Zustand S_i als Anfangszustand beobachtet wurde.
- **Ausgabezeichen** $\{O_1, \ldots, O_n\}$ sind die möglichen Ausgaben, welche beim Durchlaufen eines Hidden Markov Modells ausgegeben werden können.

- **Emissionswahrscheinlichkeiten** $b(i)(O_k)$, welche die Wahrscheinlichkeiten angeben, das im Zustand S_i die Ausgabe O_k ausgegeben wird.

6.2 Auswertung

Die Auswertung eines Hidden Markov Modells kann von zwei Perspektiven betrachtet werden. Man spricht von der Evaluation, wenn man sich dafür interessiert, wie wahrscheinlich eine Beobachtung ist. Das Dekoding beschäftigt sich mit der Suche der wahrscheinlichsten Zustandsfolge für eine Beobachtungssequenz. Beide Gesichtspunkte möchte ich im Folgenden kurz vorstellen.

6.2.1 Evaluation

Die Evaluation beschäftigt sich mit der Bestimmung der Wahrscheinlichkeit einer Beobachtungsfolge. Diese Information ist wichtig, wenn man eine Beobachtung voraussagen möchte. Der dafür verwendete Algorithmus ist der Forward Algorithmus. Dieser ist von der dynamischen Programmierung beeinflusst und bestimmt die Wahrscheinlichkeiten der Zustandsübergänge iterativ.

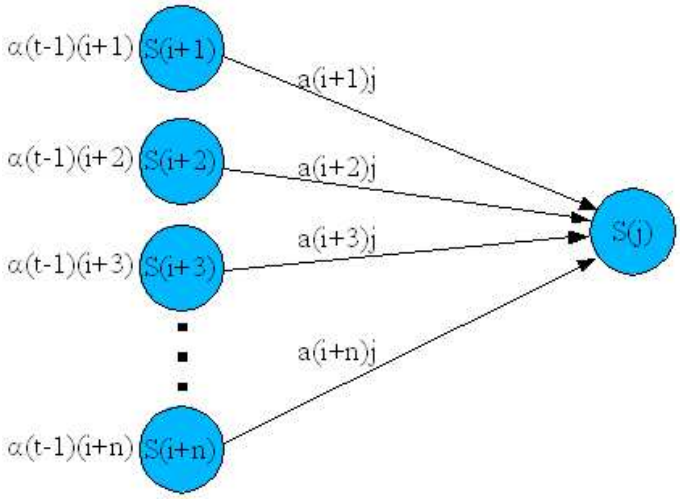


Abbildung 6.2: In dem Schaubild ist die rekursive Berechnung der Wahrscheinlichkeit der Zustandsfolge veranschaulicht. Man kennt die Wahrscheinlichkeit der Vorgängerzustände $\alpha_{t-1}(i+1)$ bis $\alpha_{t-1}(i+n)$ und die Übergangswahrscheinlichkeit vom Zustand S_{i+1} in den Zustand S_j etc. gegeben als $a_{(i+1)j}$. Damit hat man alles, was man für die Berechnung der Wahrscheinlichkeit des aktuellen Zustandes benötigt.
Darüber, wie man die Wahrscheinlichkeiten für die Vorgängerzustände berechnet, braucht man sich an dieser Stelle nicht zu überlegen. Dies geschieht auf die gleiche Art und Weise, wie die Berechnung dieser Wahrscheinlichkeit. Die Ausnahme bildet die Startwahrscheinlichkeit $\alpha_0(i)$. Dieser wird die Emissionswahrscheinlichkeit π_i zugeordnet.

Kennt man die Wahrscheinlichkeiten aller möglichen Vorgängerzustände $\alpha_{t-1}(i)$, dann kann für ein gegebenes Modell die Wahrscheinlichkeit dieses Zustandes wie folgt bestimmt werden:

Beobachtungswahrscheinlichkeit Forward(Modell, Beobachtungsfolge)
    Initialisiere Wahrscheinlichkeiten $\alpha_0(i)$ mit π_i
    Für alle Beobachtungen O_t in der Beobachtungsfolge $(O_1, \ldots, O_m)$
        Für alle Zustände S_j aus $(S_1, \ldots, S_n)$
            Für alle Zustände S_i aus $(S_1, \ldots, S_n)$
                $\alpha_t(j) \Leftarrow \alpha_t(j) + \alpha_{t-1}(i)a_{ij}$
            $\alpha_t(j) \Leftarrow \alpha_t(j) * b_j(O_t)$
    Für alle Zustände S_i aus $(S_1, \ldots, S_n)$
        $p \Leftarrow p + \alpha_n(i)$
    return p

Tabelle 6.1: Aus den gegebenen Übergangswahrscheinlichkeiten a_{ij} und den Emissionswahrscheinlichkeiten $b_i(O_t)$ wird die Wahrscheinlichkeit bestimmt mit welcher sich die einzelnen Zustände $(S_1 \ldots S_n)$ zum Zeitpunkt m beobachten lassen und die Ausgabe $(O_1, \ldots, O_m)$ erzeugt wurde.
Die Bestimmung erfolgt dynamisch. Dies bedeutet, das im ersten Schritt die Bestimmung der Wahrscheinlichkeit von (O_1), dann der Wahrscheinlichkeit (O_1, O_2) etc. durchgeführt wird, bis die Wahrscheinlichkeit der Beobachtung $(O_1, \ldots, O_m)$ für alle Zustände bekannt ist.
Durch die Aufsummierung der einzelnen Wahrscheinlichkeiten der Zustände erhält man die Wahrscheinlichkeit, mit welcher die Beobachtung $(O_1, \ldots, O_m)$ gemacht wird.

$$\alpha_t(j) = [\sum_{i=0}^{N} \alpha_{t-1}(i)a_{ij}]b_j(t)(O_t) \tag{6.1}$$

Das heißt, daß einfach die Wahrscheinlichkeiten der Vorgängerzustände $\alpha_{t-1}(i)$ entsprechend der Übergangswahrscheinlichkeiten a_{ij} gewichtet werden und aufsummiert. Denn mit einer Wahrscheinlichkeit von a_{ij} wird er in den Zustand S_j vom Zustand S_i wechseln, in dem er sich zu $\alpha_{t-1}(i)$ befindet. Damit ist die Wahrscheinlichkeit für einen Wechsel von S_i nach S_j gegeben durch $a_{ij}\alpha_{t-1}(i)$. Da es dem Zustand egal ist, von welchem Zustand in ihn gewechselt wird, werden die Wahrscheinlichkeiten für die verschiedenen Zustandsübergänge einfach aufsummiert.

Es bleibt die Frage, wie $\alpha_0(i)$ initialisiert werden soll. $\alpha_0(i)$ wird die initiale Zustandswahrscheinlichkeit π_i zugewiesen. Der hintere Term $b(j)(O_t)$ gibt die Ausgabewahrscheinlichkeit der gemachten Teilbeobachtung O_t im neuem Zustand S_j an.

Wichtig ist zu sehen, daß die Wahrscheinlichkeit von der Beobachtung steigt, wenn die Wahrscheinlichkeit der Vorgängerzustände groß ist, der Zustandswechsel wahrscheinlich ist und die Teilbeobachtung häufig in dem Zustand gemacht wird.

6.2.2 Dekodierung

Die Dekodierung beschäftigt sich mit der Frage, was die wahrscheinlichste Zustandsübergangsfolge in einem gegebenen Markov Modell ist. Um diese Frage zu beantworten, wird der Forward Algorithmus so angepaßt, daß nicht mehr alle Vorgängerzustände berücksichtigt werden, sondern nur der wahrscheinlichste. Damit wird aus dem Forward Algorithmus der Viterbi Algorithmus 6.2.

ZustandsfolgeViterbi(Modell, Beobachtungsfolge)
Initialisiere Wahrscheinlichkeiten $\alpha_0(i)$ mit π_i
sonst mit 0
Für alle Beobachtungen O_t in der Beobachtungsfolge $(O_1, \ldots, O_m)$
Für alle Zustände S_i aus $(S_1, \ldots, S_n)$
Für alle Zustände S_j aus $(S_1, \ldots, S_n)$
Wenn $(\alpha_{t-1}(i)a_{ij}b_j(O_t) > \alpha_t(j))$
$\alpha_t(j) \Leftarrow \alpha_{t-1}(i)a_{ij}b_j(O_t)$
Vorgänger $\Leftarrow S_i$
Zustandsgraph $\Leftarrow$ Hinzufügen(Vorgänger, S_j, t)
Für alle Zustände S_j aus $(S_1, \ldots, S_n)$
Wenn $(p > \alpha_n(j))$
p $\Leftarrow \alpha_n(j)$
Vorgänger $\Leftarrow S_i$
Zustandsfolge $\Leftarrow$ VerfolgeZustandsfolge(Vorgänger, Zustandsgraph)
return Zustandsfolge

Tabelle 6.2: Der Viterbi Algorithmus bestimmt die wahrscheinlichste Zustandsfolge, indem er sich für jeden Zustand $(S_1, \ldots, S_n)$ und Zeitpunkt t, bestimmt durch die Beobachtung O_t, den wahrscheinlichsten Vorgängerzustand S_i merkt.
Um die letztendlich am wahrscheinlichste Zustandsfolge zu bestimmen, werden alle Endzustände nach dem wahrscheinlichsten Zustand durchsucht. Für den wahrscheinlichsten Zustand wird dann die Zustandsfolge, welche zu der Wahrscheinlichkeit geführt hat, zurückverfolgt.

Da der Viterbi Algorithmus immer nur die wahrscheinlichste Zustandsfolge weiter berücksichtigt, erhält man am Ende auch die Wahrscheinlichkeit für die wahrscheinlichste Zustandsfolge. Hat man sich bei jedem rekursivem Schritt der Bestimmung der Zustandsfolge den wahrscheinlichsten Vorgängerzustand gemerkt, dann kann daraus die wahrscheinlichste Zustandsübergangsfolge rekonstruiert werden.

Kapitel 7

Konfidenzmaße

Das Wort Konfidenz bedeutet Zuversicht und Vertrauen. Damit ist ein Konfidenzmaß in der Spracherkennung eine Größe, welche angibt, wie stark das Vertrauen in das erkannte Wort bzw. den erkannten Satz ist. Mit der Aufgabe, wie man die Konfidenz für ein Wort bestimmen kann, haben sich schon verschiedene Arbeiten beschäftigt [Metze04], [Mangu99], [Kemp97]. Zwei Konfidenzen, welche im Weiteren in der Studienarbeit verwendet werden, werden in diesem Abschnitt vorgestellt. Dabei handelt es sich um die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit und die Konfidenz auf Basis des Consensus.

7.1 Konfidenz auf Basis der a posteriori Wahrscheinlichkeit

Grundlage für die Bestimmung der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit [Metze04] ist, wie der Name schon sagt, die a posteriori Wahrscheinlichkeit. Die a posteriori Wahrscheinlichkeit ist die Ausgabe eines Hidden Markov Modells, welches häufig in der Spracherkennung eingesetzt wird, um die wahrscheinlichste gesprochene Hypothese für ein gegebenes akustisches Signal zu bestimmen. Die Hidden Markov Modelle wurden im Abschnitt der Lernverfahren vorgestellt.

Die a posteriori Wahrscheinlichkeit für die erkannte Hypothese ist durch einen Pfad durch die Zustände des Hidden Markov Modells bestimmt. Mittels Dekodierung kann dieser Pfad bestimmt werden. Aus dem Pfad kann die Zustandsfolge für die einzelnen Wörtern bestimmt werden.

Das akustische Signal kann also in Teilabschnitte zerlegt werden, welche den einzelnen Wörtern in der erkannten Hypothese entsprechen. Bestimmt man für diesen Ausschnitt aus dem akustischen Signal mit Hilfe des Hidden Markov Modells die a posteriori Wahrscheinlichkeit $p^I(w)$, so erhält man die wahrscheinlichste Wortsequenz für diese akustische Beobachtung.

Die a posteriori Wahrscheinlichkeit des gefundenen Wortes kann mit der a posteriori Wahrscheinlichkeit des Wortes der kompletten Wortsequenz $p(w)$ verglichen werden. Die a posteriori Wahrscheinlichkeit für das im Zusammenhang erkannte Wort bestimmt sich aus der Differenz der Wahrscheinlichkeit bei Eintritt in das Wort p_E und bei Austritt aus dem Wort p_A, wenn wir uns in der

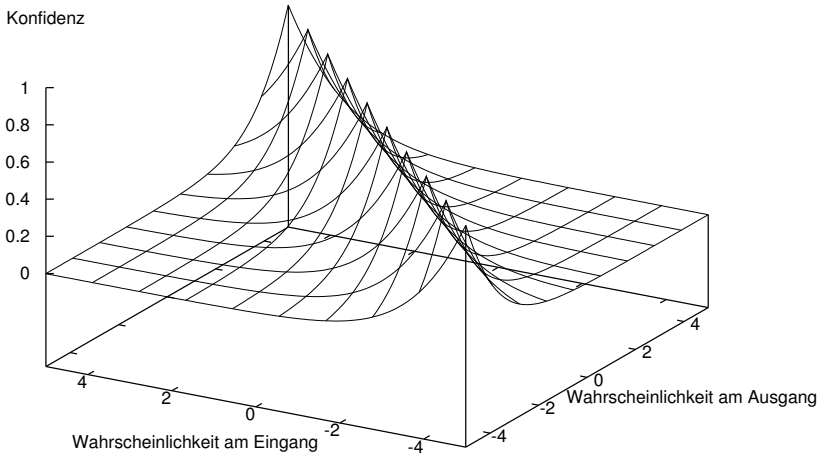


Abbildung 7.1: In der Abbildung sieht man den Verlauf des Konfidenzmaßes auf Basis der a posteriori Wahrscheinlichkeit abhängig von der lokalen a posteriori Wahrscheinlichkeit $P^I(w)$ und der globalen a posteriori Wahrscheinlichkeit $P(w)$. In der Diagonalen, in welcher beide Wahrscheinlichkeitswerte den selben Wert annehmen, hat das Konfidenzmaß den Wert 1. Je stärker sich die beiden Wahrscheinlichkeitswerte unterscheiden, desto weiter sinkt das Konfidenzmaß ab und konvergiert gegen den Wert 0.

logarithmischen Transformation befinden. p_E und p_A können mittels Forward- und Backward-Algorithmus bestimmt werden.

Die a posteriori Wahrscheinlichkeit für den Ausschnitt aus dem Signal $p^I(w)$ und die a posteriori Wahrscheinlichkeit für das erkannt Wort $p(w)$ bestimmen die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit wie folgt:

$$Conf_{Post}(w) = log(e^{|p(w)-p^I(w)|} + 1)$$

Das Konfidenzmaß $Conf_{Post}(w)$ ist hoch, wenn der Wahrscheinlichkeitsunterschied gering ist und damit die global gefundene Wortsequenz sehr nah an der lokal gefundenen Wortsequenz liegt. Weicht das lokal gefundene Ergebnis stark von dem global gefundenen Ergebnis ab, dann ist der Wahrscheinlichkeitsunterschied groß und damit liegt das Konfidenzmaß nahe 0. Den Zusammenhang der beiden Werte sieht man schön in Abbildung 7.1.

7.2 Konfidenz auf Basis des Consensus

Die Konfidenz auf Basis des Consensus wird in [Mangu99] vorgestellt. Die Grundlage für die Bestimmung der Consensus basierten Konfidenz ist ein Konfusionsnetzwerk. Ein Konfusionsnetzwerk wird aus dem Worterkennungsgraphen in zwei Schritten erstellt:

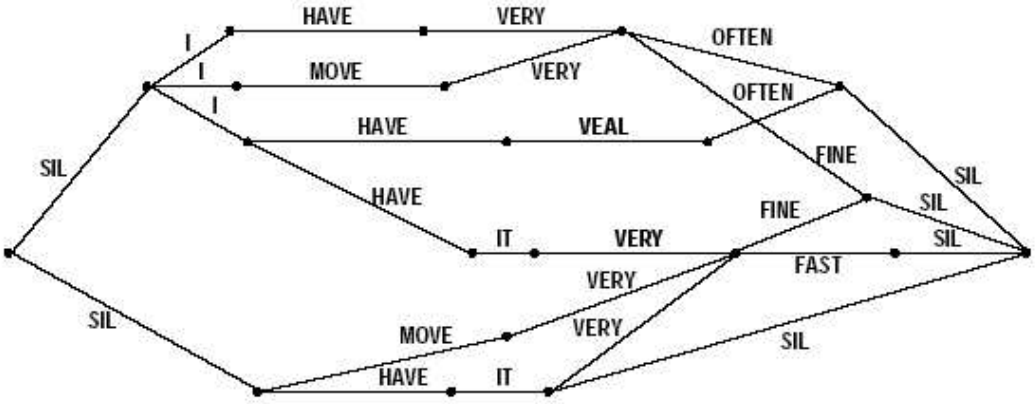


Abbildung 7.2: Der Erkennungsgraph ist eine kompakte Darstellung der erkannten Hypothesen für die akustische Beobachtung. Die Kanten stellen die erkannten Wörter dar. Mit Knoten werden die möglichen Übergänge zwischen den Wörtern veranschaulicht. Dabei ist eine n - m Zuordnung zwischen den Wörtern möglich. Das Beispiel ist aus dem Paper [Mangu99] entnommen.

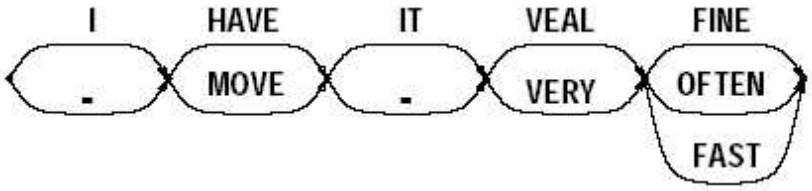


Abbildung 7.3: Das Konfusionsnetzwerk gibt den erkannten Wörtern eine Ordnung. In der spezielleren Form des Konfusionsnetzwerkes sind leichte Unterschiede im Startpunkt der möglichen gesprochenen Wörter beseitigt und Wörter können nur parallel oder hintereinander angeordnet sein. Dies reduziert die Möglichkeit der erkannten Sätze durch Abstraktion und schafft eine Vergleichsmöglichkeit der erkannten Wörter.

1. **Intra Wort Clusterung:** Im ersten Schritt findet eine Intra Wort Clusterung statt. Diese Intra Wort Clusterung faßt Kanten zusammen, welche dem gleichem Wort zugeordnet sind und sich überschneiden. Die Ähnlichkeit zweier Cluster bestimmt sich wie folgt:

$$
\begin{aligned}
Sim(C_1, C_2) &= max_{c_1 \in C_1, c_2 \in C_2} \text{Überlagerung}(c_1, c_2) p(c_1) p(c_2) \\
&\text{mit} \\
p(c_i) &= \text{a posteriori Wahrscheinlichkeit für Cluster i} \\
\text{Überlagerung}(c_1, c_2) &= \text{normierte akustische Wortüberschneidung} \\
&= \frac{\text{gemeinsame Zeit}}{Zeit(c_1) + Zeit(c_2)}
\end{aligned}
$$

   Die Intra Wort Clusterung bewirkt, daß gleiche Wörter als eine Einheit betrachtet werden.

2. **Inter Wort Clusterung:** Im zweiten Schritt findet die Inter Wort Clusterung statt. Die Inter Wort Clusterung erzeugt aus den verschiedenen Clustern eine Ordnung. Cluster, welche zueinander noch nicht geordnet

sind, d.h. zeitlich nicht hintereinander oder gleichzeitig liegen, werden zu einem Cluster zusammengefaßt.

Die Ähnlichkeit zweier zusammengefaßter Cluster bestimmt sich nach folgender Formel:

$$
\begin{aligned}
Sim(C_1, C_2) &= avg_{w_1 \in Wort(C_1), w_2 \in Wort(C_2)} sim(w_1, w_2) p_{C_1}(w_1), p_{C_1}(w_2) \\
&\text{mit} \\
p_{C_j}(w_i) &= \text{a posteriori Wahrscheinlichkeit für das Wort i im Cluster j} \\
sim(w_1, w_2) &= \text{normierte Editierdistanz} \\
&= 1 - \frac{phonetischeEditierdistanz(w_1, w_2)}{Zeit(w_1) + Zeit(w_2)}
\end{aligned}
$$

Am Ende der Inter Wort Clusterung bilden die Cluster ein Konfusionsnetzwerk.

Das Ergebnis der Intra und der Inter Wort Clusterung ist in Abbildung 7.2 anhand eines Beispiels dargestellt.

In einem dritten Schritt kann ein Pruning des Konfusionsnetzwerkes vorgenommen werden. Beim Pruning werden unbedeutende Cluster, welche von anderen Clustern stark dominiert werden, entfernt. Damit werden Cluster entfernt, welche nur von unwahrscheinlichen Hypothesen gestützt werden.

Aus dem Konfusionsnetzwerk kann für jedes Cluster die Konfidenz auf Basis des Consensus abgelesen werden durch:

$$
\begin{aligned}
Conf_{Cons}(x) &= \sum_{x \in C_i, Word(x) = w_i} p(x) && \text{wenn } w_i \neq' -' \\
&= \sum_{x \in C_i} p(x) && \text{wenn } w_i =' -'
\end{aligned}
$$

Die Konfidenz auf Basis des Consensus ist hoch, wenn sich das wirklich erkannte Wort in einem Cluster mit hoher Wahrscheinlichkeit befindet. Die Konfidenz auf Basis des Consensus schwankt zwischen 0 und 1. Ein Wert nahe 1 gibt dem erkannten Wort eine hohe Konfidenz.

Kapitel 8

Merkmale

Die Wahl der Merkmale ist eine wichtige Entscheidung, denn die Merkmale bilden die Grundlage, auf denen die Lernmethoden aufsetzten. Es gilt aussagekräftige Kriterien zu finden, anhand welcher die Beobachtungen gut in die beiden Klassen 'korrekt erkannt' und 'inkorrekt erkannt' eingeordnet werden können.

Dies ist in diesem Fall besonders schwierig, denn die im ersten Schritt vorgenommene Spracherkennung ist bereits sehr weit entwickelt und greift schon auf viele aussagekräftige Kriterien zurück.

8.1 Wortlänge

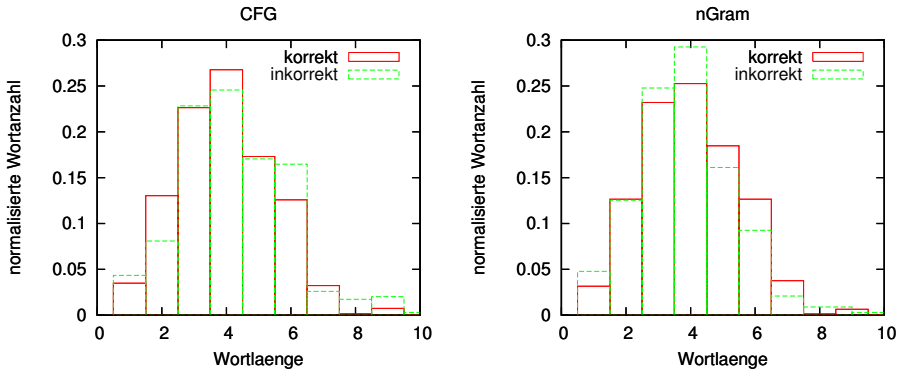


Abbildung 8.1: In der Abbildung sieht man wie die Wortlänge für die korrekt erkannten (rot) und die nicht korrekt erkannten Wörter (grün) verteilt ist. Links sieht man die Verteilung für das CFG- basierte Sprachmodell und rechts für das nGram- basierte Sprachmodell. Die Anzahl der Wörter wurde auf alle untersuchten Wörter normiert, damit die beiden Reihen besser verglichen werden können, obwohl die ausgewertete Anzahl an Beispielen für falsch und korrekt erkannt Wörter voneinander abweicht.

Wie man in der Abbildung 8.1 sieht, werden beim CFG- basierten Spracherkenner längere Wörter eher falsch verstanden, während beim nGram- basierten Spracherkenner kürzere Wörter falsch erkannt werden. Der CFG- basierte Spracherkenner hat nur ein sehr begrenztes Vokabular, welches mit der Grammatik

festgelegt ist. Besonders lange Wörter sind dem Vokabular der Grammatik nicht bekannt, so daß lange Wörter nicht erkannt werden.

Der nGram- basierte Spracherkenner muß dagegen mit dem Problem kämpfen, daß kurze Wörter akustisch sehr ähnlich sind. Damit ist es für das nGram-Sprachmodell schwierig, kurze Wörter nicht zu vertauschen. Diese Information könnte den Lernmethoden helfen, falsch erkannte Wörter zu bestimmen.

8.2 a posteriori Score

Der a posteriori Score ist ein statistisches Maß, welches angibt, wie wahrscheinlich das erkannte Wort für die beobachtete akustische Hypothese ist. Die Wahrscheinlichkeit ist nach Bayes durch das akustische Modell (= klassenbedingte Wahrscheinlichkeit) und das Sprachmodell (= a priori Wahrscheinlichkeit) bestimmt. Das akustische Modell gibt an, wie üblich das Phonem für die akustische Beobachtung ist. Das Sprachmodell gibt an, wie wahrscheinlich ein Wort auf ein anderes folgt und formt so aus der Akustik Sätze, welche üblicherweise gesprochen werden.

Die Idee ist, daß Wörter, welche einen niedrigen a posteriori Score aufweisen, korrekt erkannt werden, während falsch erkannte Wörter einen höheren a posteriori Score haben.

Der a posteriori Score selbst ist aber nicht aussagekräftig, weil die Wortlänge einen starken Einfluß auf den a posteriori Score hat. Denn mit steigender Länge der akustischen Beobachtung nimmt der a posteriori Score für das Wort ständig zu.

Darum wird der normierte a posteriori Score verwendet, wenn vom a posteriori Score die Rede ist. Beim normierten a posteriori Score findet eine Normierung über die im Wort enthaltenen Phoneme statt. Die Annahme dahinter ist, daß lange Wörter aus einer längeren akustischen Sequenz entstanden sind.

Der normierte a posteriori Score p_n läßt sich aus dem a posteriori Score p und der Anzahl der Phoneme n wie folgt bestimmen:

$$p_n = ln(n) + p$$

In Abbildung 8.2 sieht man, daß es sowohl Wörter mit niedrigerem a posteriori Score gibt als auch mit hohem a posteriori Score gibt. Die beiden Kurven sind aber sehr ähnlich, so das beim normiertem a posteriori Score kein signifikanter Unterschied zwischen falsch und korrekt erkannten Wörtern sichtbar ist.

8.3 Konfidenzmaße

Eine weitere Überlegung war, daß ein gutes Merkmal zur Bestimmung des Vertrauens in ein Wort ein Merkmal sein sollte, welches dafür entwickelt wurde, die zu bestimmende Charakteristik widerzuspiegeln. Darum wird auf die vorgestellten Konfidenzen als Merkmal für die verschiedenen Klassifikatoren zurückgegriffen.

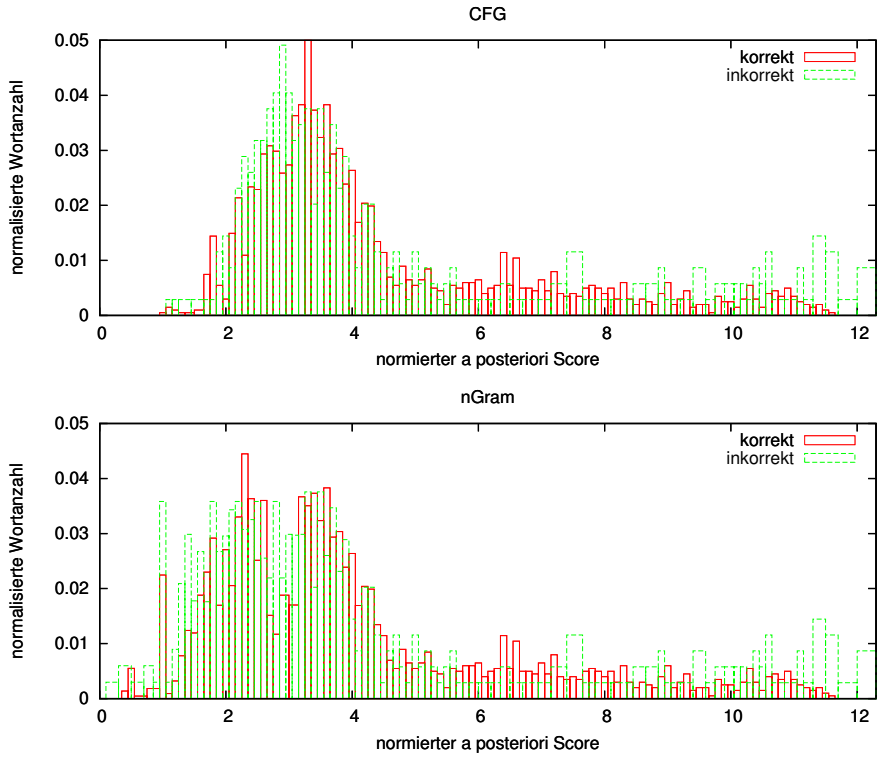


Abbildung 8.2: In der Abbildung ist der a posteriori Score dargestellt, welcher auf die Phonemanzahl des Wortes normiert wurde. Die Wahl der Phonemanzahl ist nicht optimal, da eine Normierung auf die Zustandsfolgenlänge vorgenommen werden müßte. Aber da die Phonemanzahl die Zustandsfolgenlänge stark beeinflußt, wurde an dieser Stelle die Heuristik der Phonemanzahl verwendet.
Die richtig erkannten Wörter wurden mit der Farbe rot dargestellt, während für die falsch klassifizierten Wörter die Farbe grün gewählt wurde. Die obere Abbildung zeigt die normierten a posteriori Scores für den CFG- basierte Spracherkenner. Die Scores für den nGram- basierten Spracherkenner sind in der unteren Abbildung zu sehen.

8.3.1 Konfidenz auf Basis der a posteriori Wahrscheinlichkeit

Die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit wurde bereits im Abschnitt über die Grundlagen ausführlich vorgestellt. An dieser Stelle soll darauf eingegangen werden, wie gut die falsch und die korrekt erkannten Wörter bestimmt werden und eine Separierung der beiden Klassen mit Hilfe der a posteriori Wahrscheinlichkeit basierten Konfidenz möglich ist.

Wie die Abbildung 8.3 zeigt, erreichen viele der falsch erkannten Wörter eine geringere Konfidenz auf Basis der a posteriori Wahrscheinlichkeit, während richtig erkannte Wörter eine höhere Konfidenz erzielen. Eine Ausnahme bildet der Grenzwerte 1.

Dieser wird sowohl von korrekt als auch inkorrekt erkannten Wörtern häufig angenommen. Eine Konfidenz von 1 wird erreicht, wenn kein anderes Wort für diese akustische Beobachtung denkbar gewesen wäre.

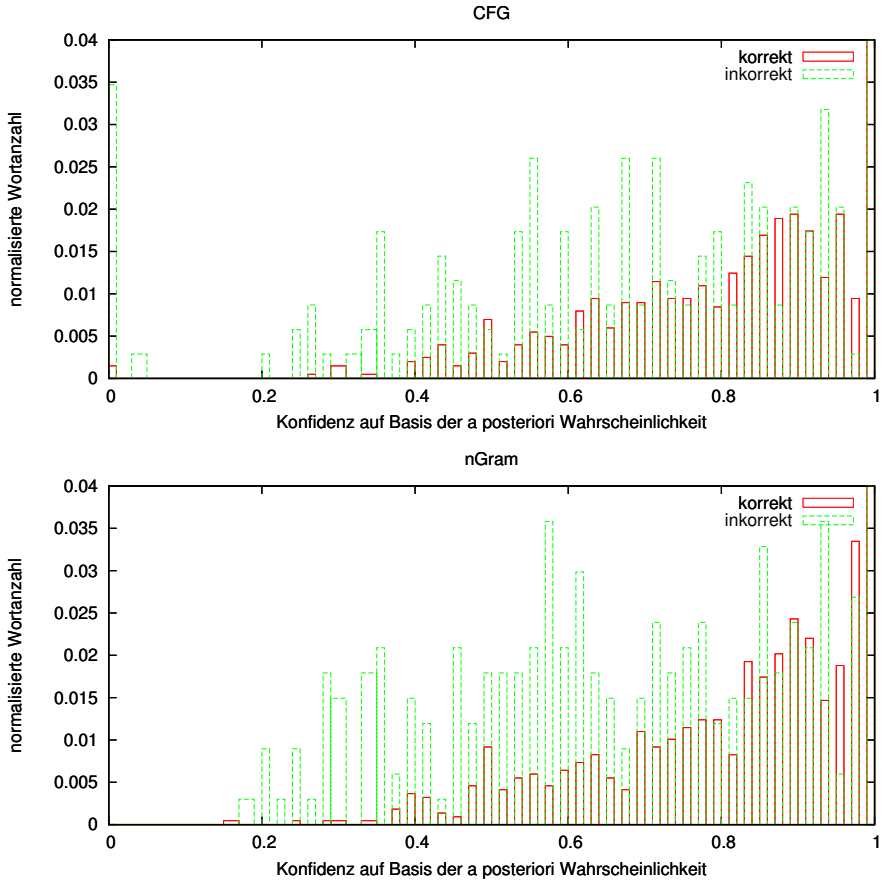


Abbildung 8.3: In der Abbildung sind zwei Kurven zu sehen. Die rote Kurve stellt die korrekt erkannten Wörter in Abhängigkeit der dafür bestimmten Konfidenz auf Basis der a posteriori Wahrscheinlichkeit da. Die grüne Kurve betrachtet die Wörter, welche vom Spracherkenner falsch erkannt wurden.
Die in der oberen Abbildung dargestellt Konfidenzen sind für einen CFG- basierten Spracherkenner bestimmt worden. Die Konfidenzen für den nGram- basierten Spracherkenner findet man in der unteren Abbildung.

8.3.2 Konfidenz auf Basis des Consensus

Die Konfidenz auf Basis des Consensus verhält sich ähnlich wie die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit. Es ist eine schöne Tendenz zu erkennen, daß richtig erkannte Wörter eine hohe Konfidenz zugeordnet bekommen und falsch erkannte Wörter eine niedrige Konfidenz. Für richtig erkannte Wörter wird eine sehr hohe Konfidenz auf Basis des Consensus bestimmt.

Wieder bildet die Konfidenz von 1 eine Ausnahme. Eine Konfidenz auf Basis des Consensus von 1 wird immer erreicht, wenn keine weitere Alternative zum vom Spracherkenner bestimmten Wort im Konfusionsnetzwerk enthalten ist. Dies trifft häufig sowohl für korrekt als auch inkorrekt erkannte Wörter auf.

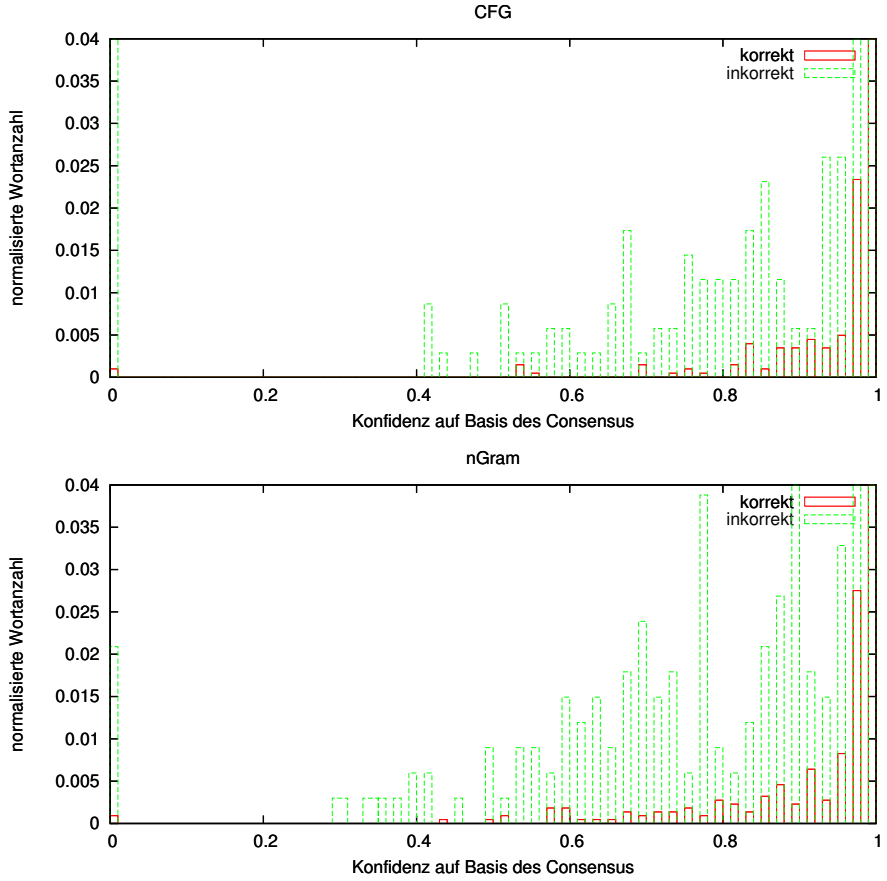


Abbildung 8.4: In den Abbildungen sind mit rot die Konfidenzen für korrekt erkannte Wörter dargestellt. Die grüne Kurve zeigt die Konfidenz auf Basis des Consensus für die vom Spracherkenner inkorrekt erkannten Wörter. In der oberen Abbildung wurde ein CFG Sprachmodell verwendet, in der unteren Abbildung ein nGram- Sprachmodell.

Teil II

Umsetzung Dialogmanagement

Kapitel 9

Spracherkenner und Dialogsystem

In diesem Abschnitt soll darauf eingegangen werden, in wie weit der Einsatz der Verfahren zur Bestimmung von Konfidenzen, das Dialogmangement verbessern kann. Aber bevor darauf näher eingegangen wird, soll das Umfeld beschrieben werden, unter welchem die Untersuchungen stattgefunden haben.

Der Spracherkenner Der eingesetzte Spracherkenner ist Janus [Rogina95]. Die Daten, auf welchen der Spracherkenner trainiert wurde, kamen aus dem Haushaltsbereich. Es wurde ein Nahbesprechungsmikrophon für die Aufnahmen von 8 verschiedenen Sprechern verwendet.

Für den CFG- basierten Spracherkenner wurde eine kontextfreie Grammatik entwickelt, welche aus 142 Regeln aufgebaut ist. Weiterreichende Informationen zur Konfiguation des Spracherkenners findet man in [Fuegen04]. Der nGram-basierte Spracherkenner wurde auf der ausgerollten Grammatik trainiert. Es wurde ein trigram Sprachmodell aufgebaut.

Das Dialogsystem Allen Trainingsdaten wurde die entsprechende semantische Referenzhypothese zugeordnet. Daraufhin konnte einfach untersucht werden, in wie weit das derzeitige Dialogsystem bereits die Anforderungen erfüllt und welches Potenzial noch genutzt werden kann.

Besonders interessant sind die Trainingsdaten, für welche keine korrekte semantische Repräsentation erstellt werden konnte, aber im Worthypothesengraphen des Spracherkenners die semantische Hypothese enthalten ist.

Die oben beschriebenen Kriterien erfüllen nur eine geringe Anzahl an Beispielen. Damit sind auch nur geringe Verbesserungen zu erwarten. Alle anderen Beispiele sind weniger interessant, weil sie entweder schon die korrekte semantische Hypothese erzeugen, oder gar nicht erzeugen können, weil die notwendige Information nicht im Worthypothesengraphen enthalten ist. Die Tabelle 9.1 zeigt die Verbesserungen, welche im optimistischsten Fall erreicht werden könnten.

Die 'Ist'-Werte geben Auskunft darüber, wie gut die semantische Repräsentationsfindung ohne Nutzung der Konfidenzangabe funktioniert. Die 'Kann'-Werte sind eine optimistische Angabe, welche das Potenzial angibt, wie viele

Merkmal	CFG Ist	CFG Kann	nGram Ist	nGram Kann
Hypothesen	646	646	646	646
korrekte Aktionen	517	527	459	486
korrekte Parameter	463	470	436	462

Tabelle 9.1: erreichbare Verbesserungen im Dialogsystem

semantische Hypothesen aus dem Worthypothesengraphen überhaupt erkannt werden könnten.

In der Zeile 'Hypothesen' kann man ablesen, wie viele Anfragen an das Dialogsystem untersucht wurden. In der Zeile 'korrekte Aktionen' sieht man die Anzahl der Anfragen, denen die richtige Aktion zugeordnet wurde. Die Aktion gibt an, was gemacht werden soll. Im unten angegebenen Beispiel ist die Aktion 'act_switch', welche angibt, daß etwas geschaltet werden soll.

```
act_switch
        robbi:ONOFF [ robbi:prp_onoff
                robbi:BOOL [ true ]
        ]
```

Wurde die Aktion richtig erkannt, dann wird ein Blick auf die Parameter geworfen, um zu schauen, ob auch diese Angabe richtig erkannt wurde. Die Parameter spezifizieren die Aktion näher. Ein Beispiel für einen Parameter ist 'robbi:ONOFF', welcher angibt, ob etwas an- oder ausgeschaltet werden soll: In diesem Fall an. Die Zahl der Hypothesen, für welche die komplette Hypothese korrekt erkannt wurde, befindet sich in der Zeile 'korrekte Parameter'.

Man sieht, daß auf der Grundlage des CFG- basierten Spracherkenners bessere semantische Ergebnisse erzielt werden, obwohl die Erkenungsleistung des Spracherkenners selber schlechter ist (76.48% statt 83.46% für den nGram- basierten Spracherkenner). Dafür ist das Verbesserungspotential geringer. Lediglich 17 semantische Hypothesen können verbessert werden. Beim nGram- basierten Spracherkenner können 30 Hypothesen verbessert werden. Die optimalen semantischen Ergebnisse bleiben trotzdem unter den Ergebnissen, die mit Hilfe des CFG- basierten Spracherkenners erreicht werden können.

Das Verbesserungspotential könnte erhöht werden, indem ein breiterer Hypothesengraph durch den Spracherkenner erzeugt werden würde. Ein höherer Beam als die derzeitigen 200 des Graphen würde aber einen höheren Aufwand für die Bearbeitung der Anfragen eines Benutzers bedeuten, da ein größerer Suchraum durchsucht werden müßte.

Kapitel 10

Wahl der Hypothese des Spracherkenners

Das Problem, welches näher betrachten werden soll, ist die Wahl der Hypothese aus dem Worthypothesengraphen des Spracherkenners. Im bestehenden Dialogsystem 'Tapas' wird die wahrscheinlichste Hypothese des Spracherkenners für die semantische Verarbeitung gewählt. Unter Zuhilfenahme der Konfidenz könnte eine bessere Wahl getroffen werden. Zu den untersuchten Konfidenzmaßen gehören die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit, auf Basis des Consensus und auf Basis des Bayes Klassifikators.

10.1 Konfidenz auf Basis der a posteriori Wahrscheinlichkeit

Das Schaubild 10.1 zeigt die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit für die wahrscheinlichste Hypothese des Spracherkenners, welche zur Zeit im Dialogsystem verarbeitet wird und die semantisch bessere Hypothese, welche im Worthypothesengraphen des Spracherkenners gefunden werden kann. Ein brauchbares Ergebnis würde den semantisch korrekteren Hypothesen eine höhere Konfidenz auf Basis der a posteriori Wahrscheinlichkeit zuordnen. Damit sollte die rote Kurve über der grünen Kurve liegen.

Dies ist aber wie man in den beiden Abbildungen sieht nicht der Fall. Es läßt sich sogar der umgekehrte Trend erkennen. Beim CFG- basierten Spracherkenner haben nur 6 der 17 Hypothesen im Durchschnitt über alle Wörter des Satzes eine höhere Konfidenz auf Basis der a posteriori Wahrscheinlichkeit. Dies entspricht $35,3\%$ der untersuchten Hypothesen. Noch gravierender sieht es beim nGram- basierten Spracherkenner aus. Nur 5 der 30 Hypothesen weisen eine höhere Konfidenz für die semantisch bessere Hypothese auf. Dies entspricht $13,3\%$ der untersuchten Aussagen.

10.2 Konfidenz auf Basis des Consensus

Die erzielten Konfidenzen auf Basis des Consensus für das CFG- Sprachmodell sind im oberen Teil der Abbildung 10.2 dargestellt. 7 von 17 der Konfidenzen auf

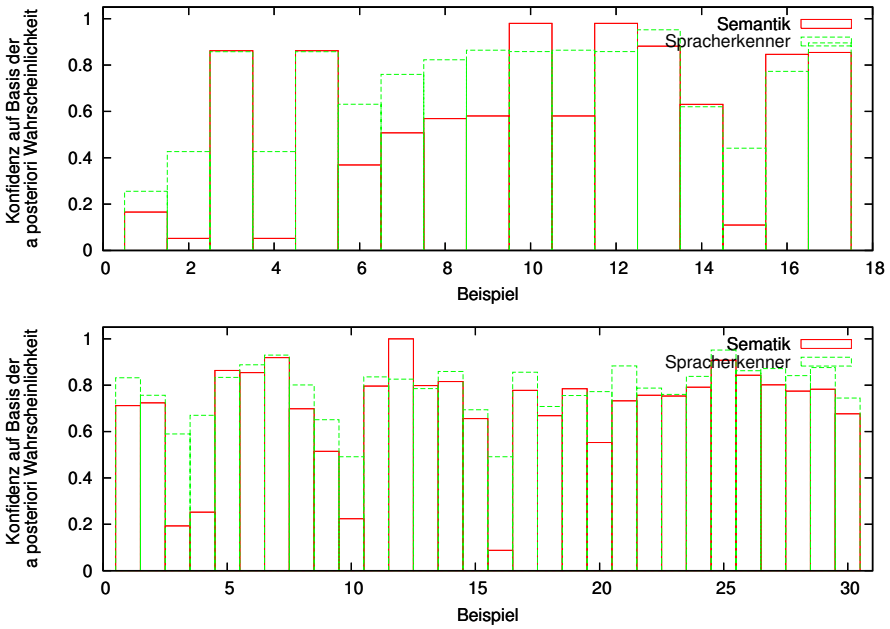


Abbildung 10.1: Beide Abbildungen zeigt zwei Kurven, welche für die untersuchten Beispiele, die sematisch besser erkannt werden könnten, die durchschnittlichen Konfidenzen auf Basis der a posteriori Wahrscheinlichkeit angeben. Die rote Kurve berücksichtigt die sematisch korrektere Hypothese. Die Kurve mit der grünen Linie zeigt die Konfidenzen für die wahrscheinlichste Hypothese des Spracherkenners. In der oberen Abbildung wurde ein CFG- Sprachmodell verwendet. Die Ergebnisse der unteren Abbildung basieren auf einem nGram- basierten Spracherkenner.

Basis des Consensus sind für die semantisch bessere Hypothese höher als für die vom Spracherkenner wahrscheinlichste Hypothese. Dies entspricht $41,2\%$ der untersuchten Datenbeispiele und ist damit leicht höher als für die Konfidenzen auf Basis der a posteriori Wahrscheinlichkeit.

Beim nGram- Sprachmodell werden 9 von 30 Datenbeispielen eine höherere Konfidenz auf Basis des Consensus für die semantisch aussagekräftigere Hypothese zugeordnet. Die $30,0\%$ sind wesentlich besser als die $13,3\%$ für die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit. Man sieht deutlich, daß für das CFG- Sprachmodell die Konfidenzmaße aussagekräftiger sind.

10.3 Bayes Klassifikator

Als letztes soll vorgestellt werden, welche Ergebnisse mit Hilfe des naiven Bayes bei der Findung der korrekten semantischen Hypothese erreicht werden können. Die Ergebnisse, welche in der Abbildung 10.3 dargestellt sind, beziehen sich auf die untersuchten Hypothesen. Von den 17 Hypothesen bestimmt mit Hilfe des CFG- Sprachmodells wurde 6 Hypothesen der semantisch besseren Hypothese eine niedere Konfidenz auf Basis des Bayes Klassifikator zugewiesen. Dies entspricht $35,3\%$ der Wörter. Für den nGram- basierten Spracherkenner ist für $40,0\%$ der Hypothesen die Konfidenzaussage richtig. 12 Hypothesen wird ein niedrigerer a posteriori Score als Konfidenz auf Basis des Bayes Klassifikators

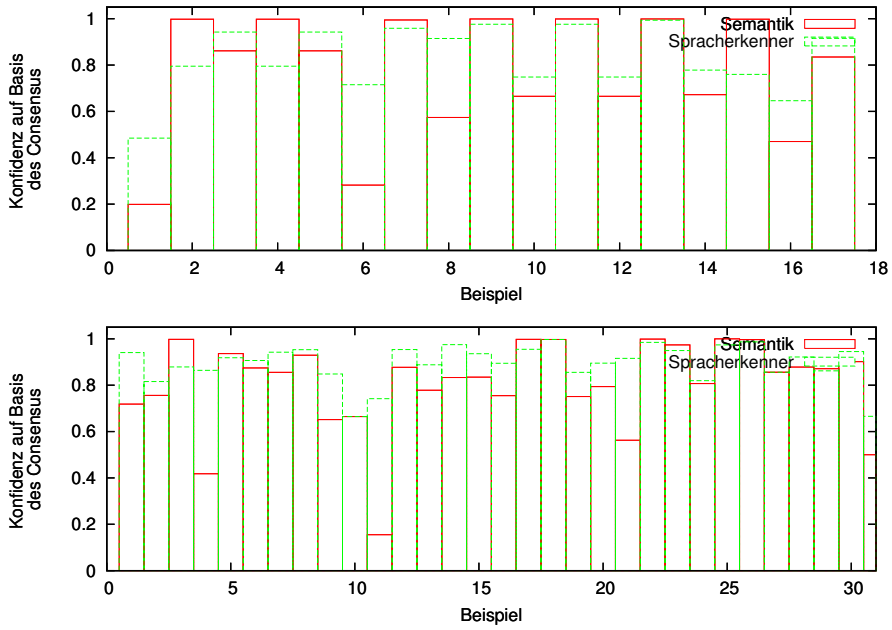


Abbildung 10.2: Die rote Kurve zeigt die Konfidenzen auf Basis des Consensus für die semantisch aussagekräftigen Hypothesen. Die vom Spracherkenner ausgewählten wahrscheinlichsten Hypothesen und die dazu bestimmten Konfidenzen sind in der grünen Kurve veranschaulicht. Die obere Abbildung basiert auf einem Spracherkenner mit CFG- Sprachmodell und die untere Abbildung basiert auf einem nGram- basierten Spracherkenner.

zugeordnet und 18 Hypothesen ein höherer Score für die semantisch korrektere Hypothese.

10.4 Einsatz

Die zu erwartenden Verbesserungen durch den Einsatz von Konfidenzmaßen im Dialogmanagement sind gering. Nicht nur, daß nur für eine geringe Anzahl an Hypothesen semantische Verbesserungen zu erwarten sind, es wird auch nur optimistisch betrachtet ca. 40% der Hypothesen ein höherer Konfidenzwert zugewiesen.

Aber warum fällt das Ergebnis so schlecht aus? Betrachtet man die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit, so ist diese durch die Distanz zwischen globaler a posteriori Wahrscheinlichkeit und lokaler a posteriori Wahrscheinlichkeit des Wortes bestimmt. Wurde in einem Satz nur ein Wort falsch erkannt, so gilt folgender Zusammenhang:

$$\begin{aligned} P_{Sprach} > P_{Sem} &\Rightarrow P_{local|Sprach} > P_{local|Sem} \text{ und } P_{global|Sprach} = P_{global|Sem} \\ &\Rightarrow P_{local|Sprach} - P_{lokal} > P_{local|Sem} - P_{lokal} \\ &\Rightarrow Conf_{aposteriori}(Sprach) > Conf_{aposteriori}(Sem) \end{aligned}$$

Ist die globale a posteriori Wortwahrscheinlichkeit für die wahrscheinlichste Hypothese P_{Sprach} größer als für die semantisch erwünschte Hypothese P_{Sem}

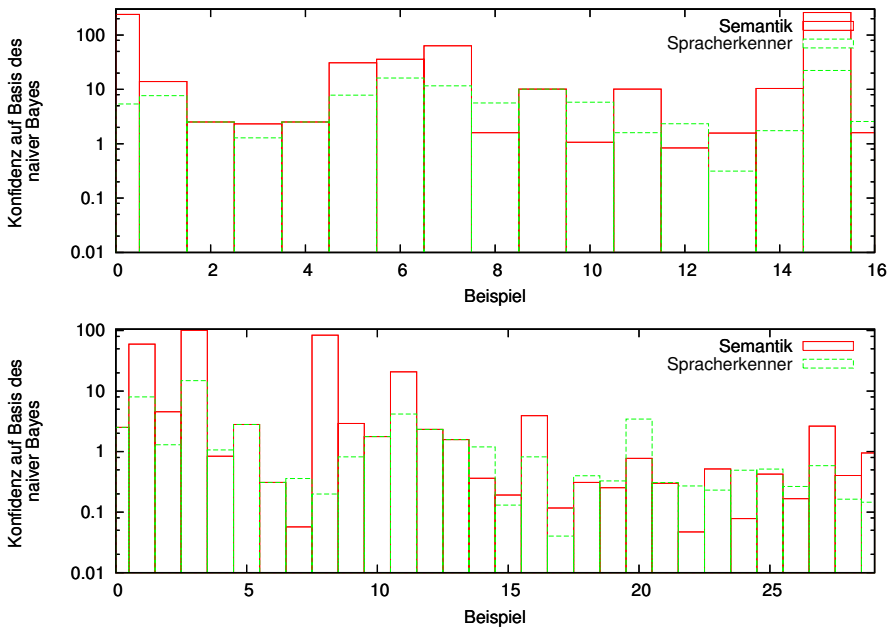


Abbildung 10.3: Die Abbildung zeigt zwei Kurven, welche für alle Hypothesen, welche untersucht wurden, die bestimmten a posteriori Scores für den Bayes Klassifikator anzeigen. Die roten Balken stellen die Konfidenzen, welche bei der Untersuchung der semantisch korrekteren Hypothesen bestimmt wurden, auf Basis des Bayes Klassifikators da. Mit den grünen Balken werden die vom Spracherkenner bestimmten wahrscheinlichsten Hypothesen veranschaulicht. Für die obere Abbildung wurden die Konfidenzen auf Basis des Bayes Klassifikators auf der Grundlage des CFG-basierten Spracherkenners bestimmt. Die untere Abbildung basiert auf einem nGram- basierten Spracherkenner.

und die lokale a posteriori Wortwahrscheinlichkeit P_{lokal} gleich, dann ergibt sich für die wahrscheinlichste Hypothese auch der höhere Konfidenzwert auf Basis der a posteriori Wahrscheinlichkeit. Also kann die Konfidenz nur für eine semantische Hypothese größer sein, wenn mehr als ein Wort falsch erkannt wurde.

Gleiches läßt sich für die Konfidenz auf Basis des Consensus sagen. Dieser ordnet alle möglichen Wörter und macht sie austauschbar. Nehmen wir an, daß nur ein Wort im Satz falsch erkannt wurde. So wird das erkannte Wort mit dem semantischen Wort im Konfusionsnetzwerk ausgetauscht. Es ergibt sich also:

$$\begin{aligned} P_{Sprach} > P_{Sem} &\Rightarrow P_{Cluster|Sprach} > P_{Custer|Sem} \\ &\Rightarrow Conf_{Consensus}(Sprach) > Conf_{Consensus}(Sem) \end{aligned}$$

Der Spracherkenner wird die Hypothese als am wahrscheinlichsten betrachten, wenn das Cluster stärker von dem falsch erkannten Wort als von dem korrekten Wort gestützt wird. Dies ist gleichbedeutend mit einer höheren Konfidenz auf Basis des Consensus.

Kapitel 11

Bestimmung semantisch falsch verstandener Sätze

Durch die Bestimmung von falsch erkannten Sätzen kann das Dialogsystem gezielt eine Fehlerbehandlung anstossen und so die Qualität des Dialogsystems verbessern. Der Benutzer würde sich nicht falsch verstanden fühlen, sondern das Dialogsystem könnte durch gezielte Nachfragen das Gespräch in die richtige Richtung führen. Mit Hilfe der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit und auf Basis des Consensus wurde untersucht, ob sich falsch erkannte Hypothesen von richtig erkannten unterscheiden lassen.

11.1 Konfidenz auf Basis der a posteriori Wahrscheinlichkeit

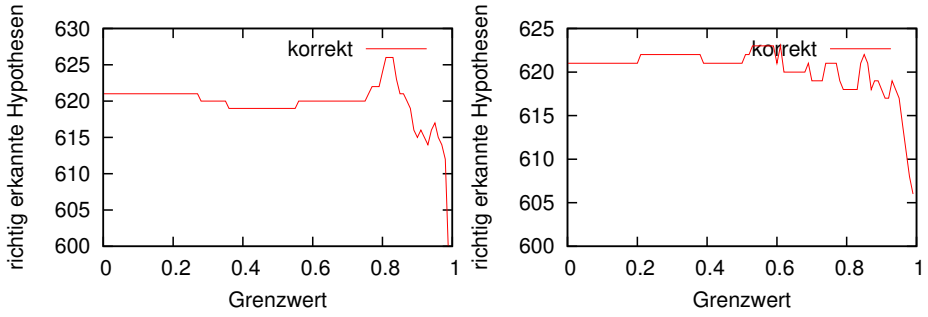


Abbildung 11.1: In der Abbildung sind für die verschiedenen Grenzwerte die Erkennungsleistungen dargestellt, welche unter Zuhilfenahme der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit erreicht werden. Alle Konfidenzen auf Basis der a posteriori Wahrscheinlichkeit unter dem Grenzwert werden als inkorrekte Hypothesen angenommen, während Hypothesen mit einer größeren Konfidenz als korrekt erkannt angenommen werden. Für die Ergebnisse in der linken Abbildung wurde ein CFG- Sprachmodell verwendet. Auf Grundlage eines nGram- Sprachmodells wurden die in der rechten Abbildung dargestellten Werte erreicht.

Die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit wurde sowohl für den CFG- basierten als auch für den nGram- basierten Spracherkenner un-

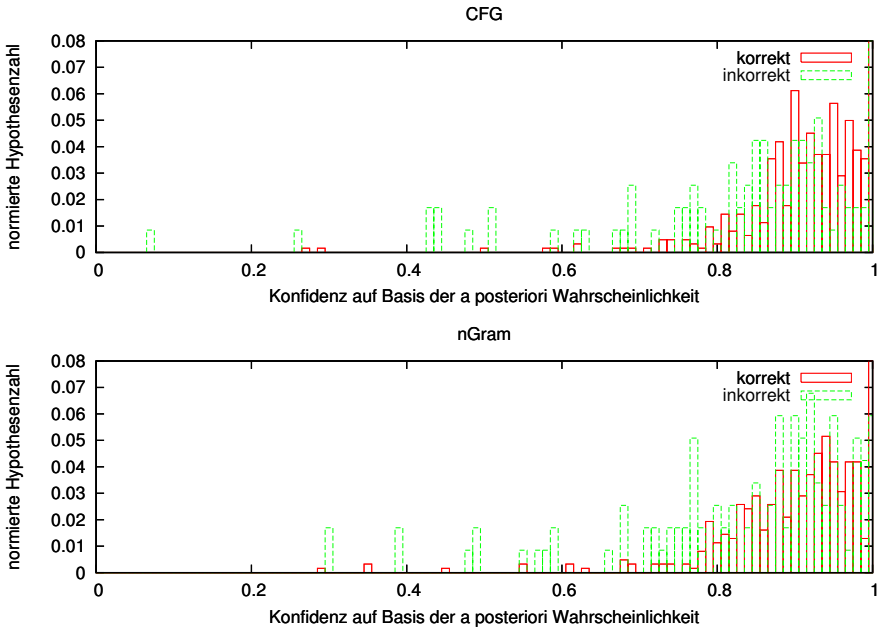


Abbildung 11.2: In der Abbildung sind sowohl für korrekt erkannte semantische Hypothesen (rot) als auch für inkorrekt erkannte Hypothesen (grün) die durchschnittlichen Konfidenzen auf Basis der a posteriori Wahrscheinlichkeiten eingezeichnet. In der oberen Abbildung wurde ein Spracherkenner verwendet, welcher auf einer kontextfreien Grammatik basiert. Die untere Abbildung wurde auf Basis eines nGram- Sprachmodells erstellt.

tersucht (Abbildung 11.2). Die Ergebnisse liegen im gleichen Bereich. Der CFG-basierte Spracherkenner erreicht bei einem Trennwert von 0.81 ein optimales Trennergebniss von 113 falsch zugeordneten Hypothesen. Dies entspricht $15,3\%$ der untersuchten Hypothesen.

Das nGram- Sprachmodell ordnet 116 Hypothesen die falsche Erkennungsleistung zu. Die Fehlerrate von $15,7\%$ wird bei einem Grenzwert von 0.53 erreicht. Die Erkennungsleistungen für die unterschiedlichen Grenzwerte sind in der Abbildung 11.1 für nGram- und CFG- Sprachmodell dargestellt.

11.2 Konfidenz auf Basis des Consensus

Die Konfidenz auf Basis des Consensus erreicht für beide untersuchte Sprachmodelle ein besseres Trennergebnisse als die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit (siehe Abbildung 11.4). Im Unterschied zur Konfidenz auf Basis der a posteriori Wahrscheinlichkeit wird aber beim nGram- basierten Sprachmodell ein besseres Ergebnis erreicht als beim CFG- basiertem Sprachmodell. Das nGram Sprachmodell bestimmt die Korrektheit der erkannten Hypothese nur für 110 Hypothesen falsch. Diese Fehlerrate von $14,9\%$ wird bei einem Trennwert von 0.78 erreicht.

Eine Hypothese ordnete die Konfidenz auf Basis des Consensus auf Grundlage des CFG- Sprachmodells der falschen Erkennungsleistung mehr zu. Somit werden 111 Hypothesen und damit $15,0\%$ bei einem Trennwert von 0.70 falsch

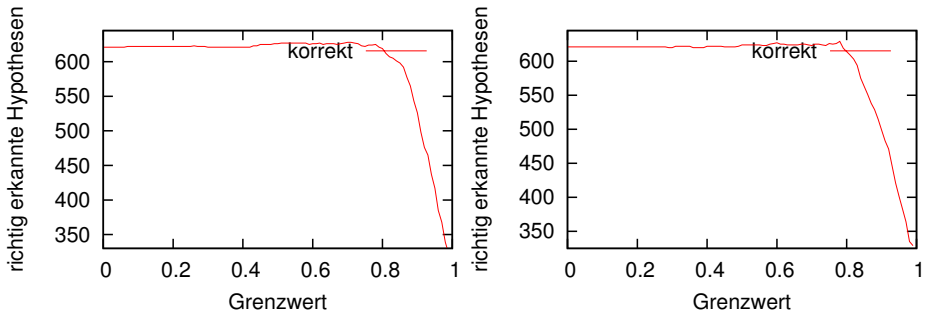


Abbildung 11.3: Die Abbildung zeigt links die korrekt eingeordneten Hypothesen für die verschiedenen Grenzwerte des Grammatik basierten Spracherkenners. Eine Hypothese wird als semantisch korrekt erkannt eingeordnet, wenn die Konfidenz auf Basis des Consensus über dem Grenzwert liegt und als falsch, wenn der Consensus unter dem Grenzwert liegt. Das rechte Schaubild zeigt die korrekt eingeordneten Hypothesen für einen nGram- basierten Spracherkenner.

zugeordnet. Die Schwankungen in den Trennleistungen sind in der Abbildung 11.3. abgebildet.

Nachdem das Dialogsystem jetzt die falsch erkannten Hypothesen automatisch bestimmen kann, ist die nächste Aufgabe, falsch erkannte Ausschnitte aus der Hypothese des Spracherkenners zu bestimmen. Durch die Bestimmung des falsch erkannten Satzteiles kann nach einer alternativen Hypothese im Worthypothesengraphen geschaut werden oder ein gezielter Klärungsdialog eingeleitet werden. Dies soll im nächsten Kapitel untersucht werden.

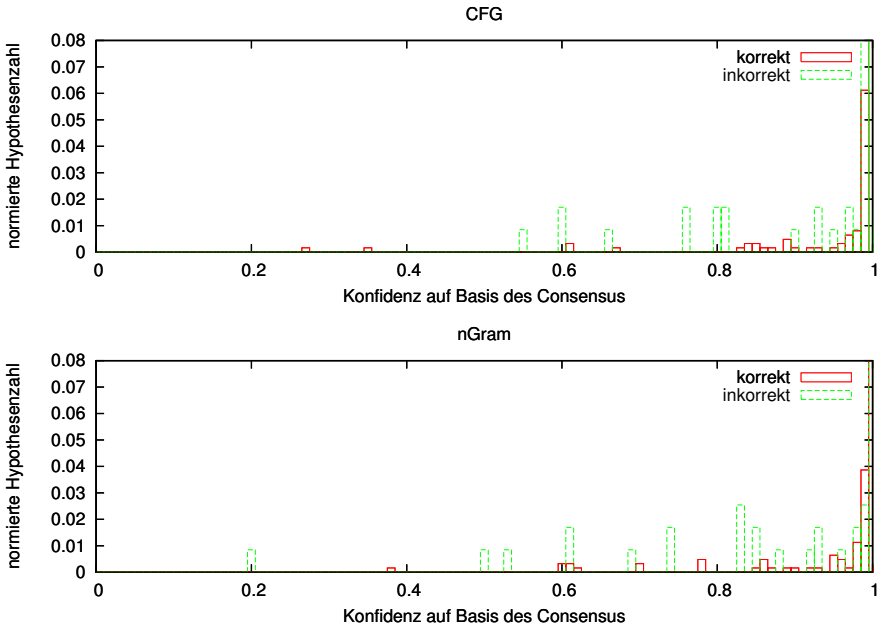


Abbildung 11.4: Die Abbildung zeigt die auf dem Consensus basierten Konfidenzwerte für korrekt erkannte Hypothesen (rot) und inkorrekt erkannte Hypothesen (grün). Für die Bestimmung der Konfidenz auf Basis des Consensus in der oberen Abbildung wurde ein CFG- Sprachmodell verwendet und für die untere Abbildung ein nGram- Sprachmodell.

Kapitel 12

Bestimmung semantisch falsch verstandener Teilsätze

Die Bestimmung von falschen Satzteilen erfolgt auf den falsch erkannten Hypothesen. Um die Aufgabe unabhängig von der Aufgabe der Bestimmung von falsch verstandenen Sätzen bearbeiten zu können, wird die Analyse nicht auf den falsch bestimmten sondern auf den falschen Hypothesen stattfinden.

Die Analyse setzt auf den Wörtern auf. Neben den Konfidenzmaßen werden auch verschiedene Lernmethoden untersucht werden, um in einem Satz, die falsch erkannten Wörter zu extrahieren. Aber zuerst soll ein Blick auf die Konfidenzen geworfen werden.

12.1 Konfidenzmaße

12.1.1 CFG- basierter Spracherkenner

Die in dem Abschnitt vorgestellten Ergebnisse wurden mit einem CFG- basierten Spracherkenner erreicht. Ein auf einer kontextfreien Grammatik basierter Spracherkenner ist weniger gut für die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit und auf Basis des Consensus geeignet, weil die Kosten für den Wechsel eines Pfades auf Grund der längeren Historie viel höher sind. Durch die stark eingeschränkten Pfade bei einem CFG- basierten Spracherkenner sind die Unterschiede zwischen den Pfaden groß und der korrekte Pfad vielleicht gar nicht enthalten.

Konfidenz auf Basis der a posteriori Wahrscheinlichkeit

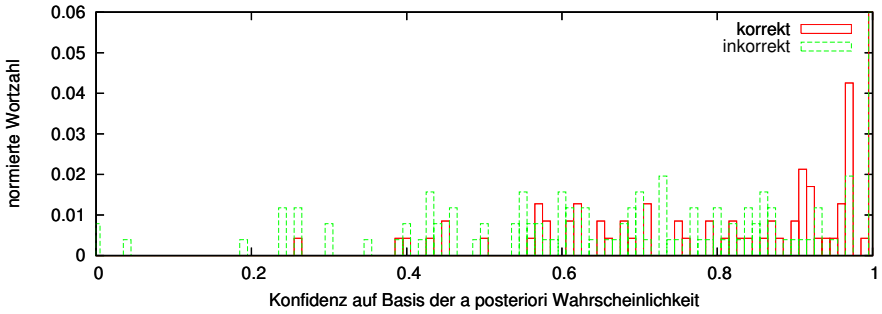
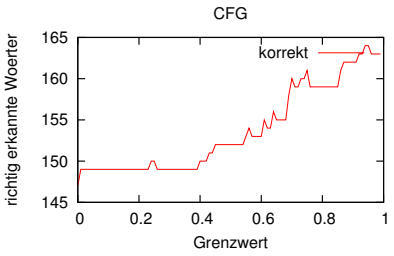


Abbildung 12.1: In der Abbildung sieht man, wie sich die Konfidenz des CFG- basierten Spracherkenners für falsch erkannte Wörter (grün) und richtig erkannte Wörter (rot) für die semantisch inkorrekten Hypothesen verteilen. Die obere Abbildung zeigt die Verteilung für die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit.
Die nebenstehende Abbildung untersucht die Trennung der Konfidenzen auf Basis der a posteriori Wahrscheinlichkeit. Es werden die verschiedenen Schwellwerte untersucht und bestimmt, wieviele der falsch erkannten Wörter eine geringere Konfidenz auf Basis der a posteriori Wahrscheinlichkeit haben und wie viele der richtig erkannten Wörter eine höhere Konfidenz aufweisen.



Die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit ist ein bekanntes Maß für die Bestimmung des Vertrauens in die Erkennung eines Wortes durch den Spracherkenner. Darum gehörte die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit zu den Merkmalen, durch die eine gute Erkennung von falsch verstandenen Wörtern durch den Spracherkenner erreicht werden sollte.

Setzt man den empirisch bestimmten Grenzwert 0.94 an, so werden, wie man in der Abbildung 12.1 sehen kann, aus den 331 Testdaten 196 richtig und 135 falsch eingeschätzt. Damit werden $59,2\%$ der Wörter in die korrekte Klasse eingeordnet.

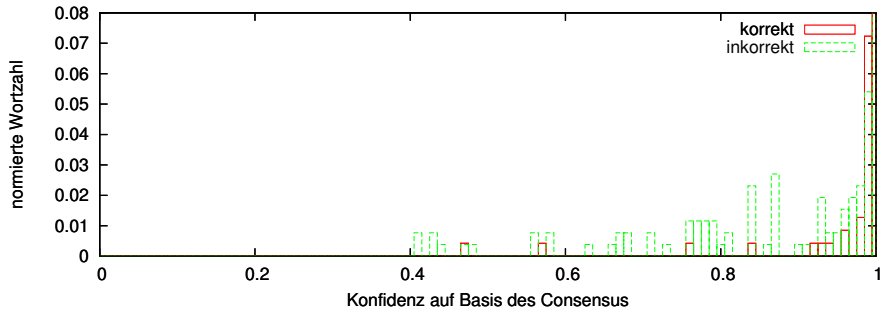
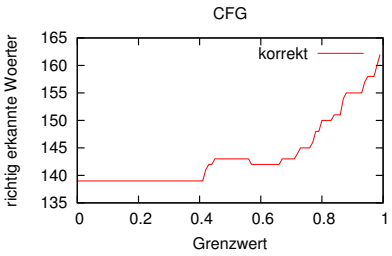


Abbildung 12.2: Die Abbildungen untersuchen die Konfindenzen auf Basis des Consensus für den CFG- basierten Spracherkenner. Die obere Abbildung zeigt die Verteilung der Konfidenzen der Wörter für vom Spracherkenner korrekt erkannte Wörter in rot und für inkorrekt erkannte Wörter in grün.
Die nebenstehende Abbildung zeigt, wie für die verschiedenen Grenzwerte die Wortzahl variiert, welche richtig den korrekt bzw. falsch erkannten Wörtern vom CFG- basierten Spracherkenner zugeordnet wird. Es wird davon ausgegangen, daß alle Wörter falsch erkannt wurden, für die sich die Konfidenz auf Basis des Consensus unter dem Grenzwert befindet. Wörter, die eine höhere Konfidenz auf Basis des Consensus haben, werden als vom CFG- basierten Spracherkenner korrekt erkannt angenommen.



Konfidenz auf Basis des Consensus

Die Konfidenz auf Basis des Consensus gehört zu den Maßen, welche das Vertrauen in die Erkennung eines Wortes durch den Spracherkenner wiedergibt. Die Leistung der Konfidenz auf Basis des Consensus bleibt aber unter den Leistungen der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit.

Von 331 Wörtern wurde 195 Wörtern die richtige Erkennungsleistung zugeordnet (Abbildung 12.2). Damit wurde mit $58,9\%$ eine leicht schlechtere Zuordnung zwischen falsch und richtig erkannten Wörtern von der Konfidenz auf Basis des Consensus erreicht als die $59,2\%$ der korrekten Zuordnungen der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit.

Interessant ist, daß der Grenzwert, welcher sowohl für die Konfidenz auf Basis des Consensus als auch für die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit sehr hoch liegt. Der beste Grenzwert hat sich bei einer Konfidenz von $0,99$ für den Consensus ergeben.

12.1.2 nGram- basierter Spracherkenner

Konfidenz auf Basis der a posteriori Wahrscheinlichkeit

Die erwartete bessere Verteilung der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit für den nGram- basierten Spracherkenner, so daß falsch erkannte Wörter auch einen schlechten Konfidenzwert zugeordnet bekommen, während korrekt erkannte Wörter eine hohe Konfidenz auf Basis der a posteriori Wahrscheinlichkeit haben, läßt sich in der Verteilung der Konfidenzen auf Basis der a

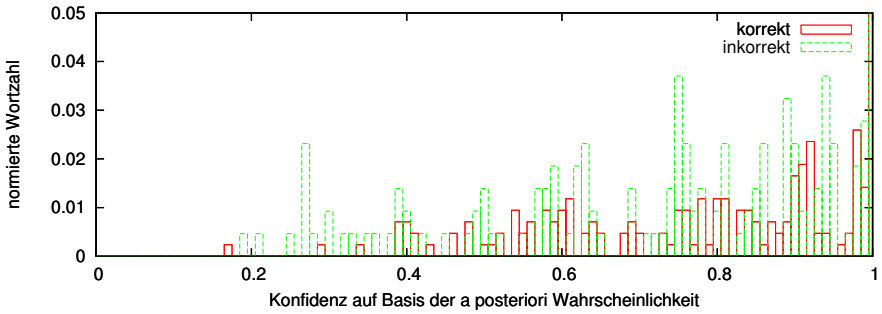
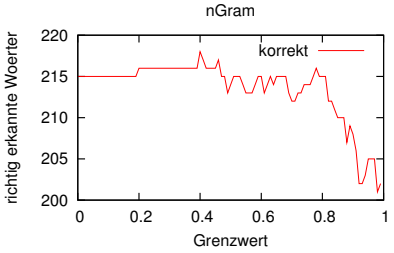


Abbildung 12.3: Das obere Schaubild zeigt die Konfidenzen auf Basis der a posteriori Wahrscheinlichkeit des nGram- basierten Spracherkenners für richtig (rot) und falsch erkannte Wörter (grün).
Die nebenstehende Abbildung untersucht das Trennergebnis für verschiedene Schwellwerte. Wörter, welche Konfidenzen auf Basis der a posteriori Wahrscheinlichkeit unter dem Schwellwert aufweisen, werden als falsch erkannte Wörter betrachtet und Wörter mit einer höheren Konfidenz als richtig erkannt.



posteriori Wahrscheinlichkeiten in der Abbildung 12.3 gut erkennen. Die Überschneidungen der zwei Kurven sind zwar vorhanden, aber es läßt erkennen, daß die Mehrzahl der falsch erkannten Wörter eine geringe Konfindenz auf Basis der a posteriori Wahrscheinlichkeit aufweisen.

Die besten Erkenungsleistungen werden bei einem Schwellwert von 0.4 erreicht. Dieser Grenzwert ist viel niedriger als beim CFG- basierten Spracherkenner. Für den gegebenen Schwellwert werden 218 untersuchte semantische Hypothesen korrekt eingeordnet. Dies sind $65,9\%$ der untersuchten Wörter. Damit ist die erreichte Rate besser als die $59,2\%$ durch den CFG- basierten Spracherkenner.

Konfidenz auf Basis des Consensus

Die Konfidenz auf Basis des Consensus ordnet von den 331 untersuchten Wörtern 223 Wörter in die richtigen Klassen ein. Damit erreicht die Konfidenz auf Basis des Consensus beim nGram- basierten Spracherkenner die besten Trennergebnisse. Immerhin $67,4\%$ der Wörter bekommen einen größeren Consensuswert als 0.98 zugeordnet, wenn sie vom nGram- basierten Spracherkenner korrekt erkannt wurden und einen kleineren Consensuswerte, wenn sie falsch erkannt wurden.

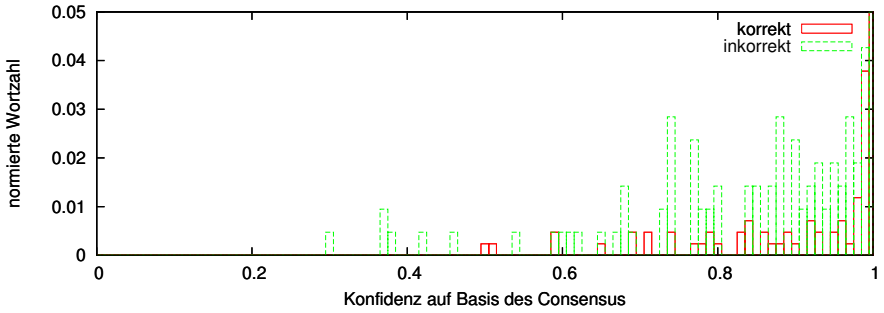
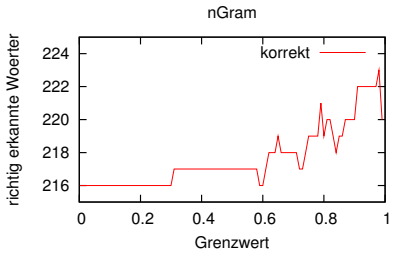


Abbildung 12.4: Die obere Abbildung veranschaulicht die Verteilung der Konfidenzen auf Basis des Consensus für richtig (rot) und falsch erkannte Wörter (grün).
Die nebenstehende Abbildung zeigt, wie für die verschiedenen Grenzwerte die Wortzahl variiert, welche richtig den korrekt bzw. falsch erkannten Wörtern vom nGram- basierten Spracherkenner zugeordnet wird. Es wird davon ausgegangen, daß alle Wörter falsch erkannt wurden, für die sich die Konfindenz auf Basis des Consensus unter dem Grenzwert befindet. Wörter, die eine höhere Konfindenz auf Basis des Consensus aufweisen, werden als vom Spracherkenner korrekt erkannt angenommen.



12.2 Lernverfahren

Die Ergebnisse, welche durch lineare Trennung der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit oder auf Basis des Consensus erreicht wurden, weisen ein großes Verbesserungspotential auf. Aus diesem Grund wurde nach einer anderen Möglichkeit gesucht, eine Aussage darüber zu treffen, ob ein Wort vom Spracherkenner richtig erkannt wurde oder nicht. Die Ergebnisse, welche mit den verschiedenen Lernmethoden erreicht wurden, sollen in diesem Kapitel vorgestellt werden.

Im Kapitel Merkmale wurden bereits folgende Merkmale vorgestellt:

- Wortlänge
- a posteriori Score
- Konfidenz auf Basis der a posteriori Wahrscheinlichkeit
- Konfidenz auf Basis des Consensus

Aus diesen Merkmalen konnten alle Lernverfahren wählen. Teilweise wurden alle Merkmale verwendet, um eine Aussage auf möglichst vielen Kriterien treffen zu können. Teilweise mußte aber auch der Merkmalsraum auf Grund der begrenzten Trainingsdaten beschränkt werden.

Bevor auf die verschiedenen Lernmethoden eingegangen wird, soll auf die verwendeten Kriterien für die Bewertung der Ergebnisse der Lernverfahren eingegangen werden.

12.2.1 Bewertung

Die Lernmethoden werden anhand von verschiedenen Kriterien beurteilt. Diese sollen im Folgenden kurz vorgestellt werden:

- **korrekt erkannt:** In der Spalte 'korrekt erkannt' befindet sich die Anzahl der korrekt klassifizierten Wörter.
- **inkorrekt erkannt:** Die Spalte 'inkorrekt erkannt' enthält die Anzahl der Wörter, welche nicht korrekt klassifiziert wurden. Die Anzahl der korrekt und inkorrekt erkannten Wörter aufsummiert ergibt die Anzahl der Testdaten.
- **Rate (%):** Die Rate gibt an, wieviele Daten prozentual korrekt erkannt wurden. Sie bestimmt sich aus dem Verhältnis aus korrekt erkannten Daten zu allen Testdaten.
- **Recall:** Der Recall gibt an, wieviele der Daten, welche als korrekt erkannt hätten gefunden werden können, auch gefunden wurden. Der Recall bestimmt sich wie folgt:

$$\text{Recall} = \frac{\text{korrekt erkannte Daten}}{\text{korrekt erkannte Daten} + \text{inkorrekt nicht erkannte Daten}}$$

- **Precision:** Die Precision gibt an, wieviele der gefundenen Daten auch korrekt gefunden wurden. Die Precision wird wie folgt angegeben:

$$\text{Precision} = \frac{\text{korrekt erkannte Daten}}{\text{korrekt erkannte Daten} + \text{inkorrekt erkannte Daten}}$$

- **F-Measure:** Der F-Measure gewichtet in einem Maß den Recall und die Precision. Recall und Precision bestimmen den F-Measure wie folgt:

$$\text{F-Measure} = \frac{(b^2+1)\text{Precision Recall}}{b^2(\text{Precision} + \text{Recall})}$$

  Der Parameter b bestimmt die Gewichtung von Precision und Recall. Mit $b = 1$ herrscht ein Gleichgewicht. Mit steigendem b gewinnt der Recall an Bedeutung, während für Werte kleiner als 1 die Precision stärker gewichtet ist. Für die Ergebnisse sollen Precision und Recall zu gleichen Teilen eingehen, so daß b auf 1 gesetzt wird.

12.2.2 Entscheidungsbäume

Merkmale

Die Entscheidungsbäume benötigen verschiedene Kriterien, auf Grund derer der ID3 Algorithmus eine Strukturierung der Daten vornimmt. Um den Algorithmus die komplette Freiheit bei der Wahl des Kriteriums zu lassen, können alle beschriebenen Kriterien verwendet werden.

Konfiguration

Bei den genannten Kriterien handelt es sich nicht um diskrete sondern um reelwertige Kriterien. Darum wird die reelwertige Erweiterung von ID3 verwendet. Die untersuchten Konfigurationen sind in der Tabelle 12.1 dargestellt.

Variante	Spracherkenner	Beschneidung	Ballung
1	CFG- basiert	ja	nein
2	CFG- basiert	ja	ja
3	CFG- basiert	nein	nein
4	nGram- basiert	ja	nein
5	nGram- basiert	ja	ja
6	nGram- basiert	nein	nein

Tabelle 12.1: Konfiguration für Entscheidungsbäume

In der Spalte 'Spracherkenner' kann abgelesen werden, welche Art Spracherkenner verwendet wird. Die Spalte 'Beschneidung' gibt Auskunft darüber, ob nach Aufbau des kompletten Entscheidungsbaumes eine Beschneidung des Baumes stattgefunden hat. Beim Beschneiden werden einzelne Äste im Baum gekürzt, so daß die Generalisierungsfähigkeit erhöht wird. Unter 'Ballung' versteht man das vorherige Zusammenfassen von Klassen in sinnvolle Einheiten. Sinnvolle Einheiten wären z.B. die a posteriori Wahrscheinlichkeit auf 0.1 genau zu runden, um die Anzahl der Knoten im Entscheidungsbaum zu begrenzen.

Ergebnisse

Variante	korrekt erkannt	inkorrekt erkannt	Rate (%)	Recall	Precision	F-Measure
1	135	142	40,8	0,82	0,45	0,64
2	131	146	39,6	0,98	0,45	0,72
3	178	153	53,8	0,74	0,57	0,66
4	195	136	58,9	0,93	0,59	0,76
5	217	114	65,6	0,85	0,66	0,76
6	**234**	**97**	**70,7**	**0,77**	**0,74**	**0,76**

Tabelle 12.2: Ergebnisse der Entscheidungsbäume

Die Ergebnisse des Entscheidungsbaumes sind in der Tabelle 12.2 dargestellt. Die dritte und die sechste Variante beider Spracherkenner ohne Beschneidung erreichen die besten Ergebnisse. Ein Auszug des verwendeten Baums ist in Abbildung 12.5 angedeutet, da hier die verschiedenen Merkmale und ihre Ordnung zu sehen sind. Sowohl der CFG- basierte Spracherkenner als auch der nGram-basierte Spracherkenner wählen als erstes Kriterium die Konfidenz auf Basis des Consensus.

Die erste Variante mit Beschneidung und ohne Ballung liefert für den CFG-basierten Spracherkenner die schlechtesten Ergebnisse. Durch die Ballung werden fast alle Zweige des Baumes entfernt, so daß das erste Kriterium, die Konfidenz auf Basis des Consensus, das entscheidende Entscheidungskriterium ist. Der trainierte Baum tendiert dazu alle Hypothesen als korrekt zu betrachten, was die hohen Recallwerte zeigen. Interessant ist, daß für den nGram- basierten Spracherkenner es geschickter ist, wenn vorher eine Ballung der Daten vorgenommen wird, während beim CFG-basierten Spracherkenner die Ballung zu einer Verschlechterung der Klassifikationsergebnisse führt.

Das semantische Lernverfahren funktioniert sehr gut. Besonders die Ergebnisse basierend auf dem nGram Sprachmodell erreichten eine gute Einordnung der erkannten Wörter. Also werfen wir nun einen Blick auf die statistischen Lernverfahren, um zu schauen, ob sie vergleichbare Ergebnisse erreichen.

12.2.3 Neuronale Netze

Merkmale

Die Wahl der Merkmale muß beim neuronalen Netz gut gewählt werden, denn im Gegensatz zum Entscheidungsbaum und zu den Bayes Klassifikatoren gilt hier nicht: Je mehr Merkmale gegeben sind, desto besser ist das Ergebnis. Dies liegt daran, daß für das Training entsprechend mehr Trainingsdaten notwendig wären. Da diese aber begrenzt sind, wurde ein wenig mit der Wahl der Merkmale experimentiert.

Die untersuchten Merkmalskonfigurationen sind in der Tabelle 12.3 dargestellt.

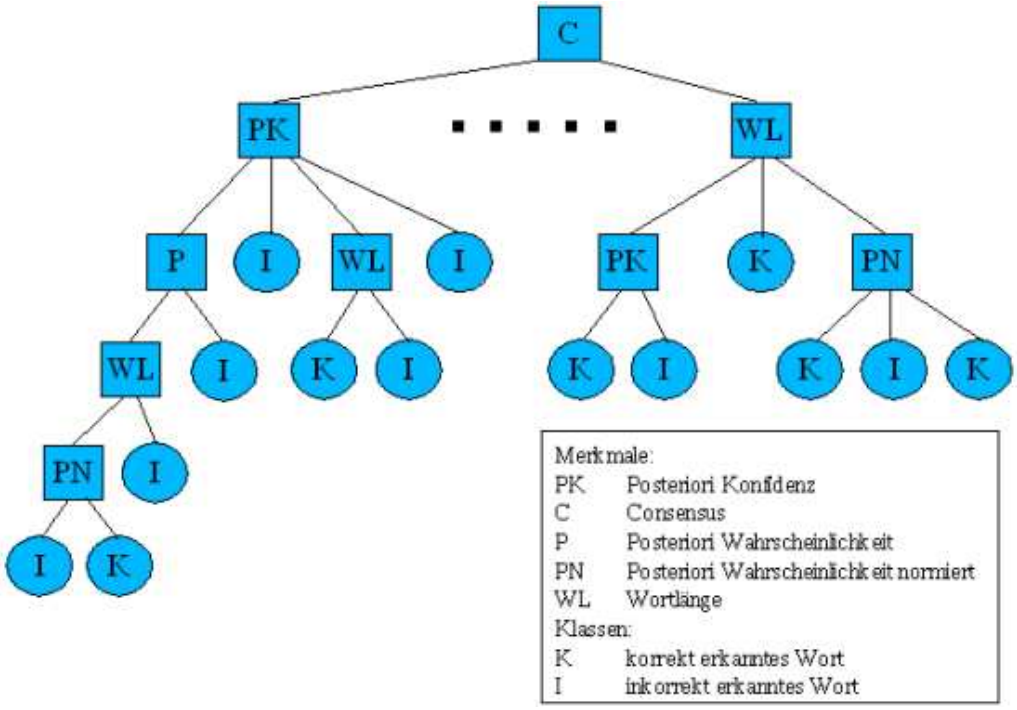


Abbildung 12.5: In der Abbildung sieht man einen Ausschnitt aus dem nGram- basierten Entscheidungsbaum ohne Beschneidung. Die Wahl des ersten Kriteriums ist auf die Konfidenz auf Basis des Consensus gefallen. Die Wahl des Merkmals auf der zweiten Ebene ist sehr unterschiedlich. Teilweise reichte bereits die erste Ebene aus, um alle bekannte Trainingsbeispiele als korrekt bzw. inkorrekt zu erkennen.
Die Tiefe des Baumes variiert stark. Teilweise wurde der komplette Merkmalsraum verwendet und es konnte nicht allen Trainingsbeispielen, welche diese Kriterien erfüllen, eine eindeutige Klasse zugeordnet werden. Die durchschnittliche Baumtiefe liegt bei drei Ebenen.

Konfiguration

Alle untersuchten neuronalen Netze haben eine Eingabeeinheit, eine versteckte Einheit und eine Ausgabeeinheit. Die Anzahl der Neuronen für die Eingabeeinheit ist bestimmt durch die Anzahl der Merkmale. Die Anzahl der Ausgabeneuronen ist 1, weil nur zwei Klassen unterschieden werden sollen. Ein hoher Ausgabewert (1) soll für ein korrekt erkanntes Wort stehen, während ein kleiner Ausgabewert (0) ein falsch erkanntes Wort signalisieren soll.

Die variable Größe ist die Anzahl der Neuronen in der versteckten Schicht. Die Anzahl der versteckten Schicht wurde durch Ausprobieren bestimmt. Die Neuronenanzahl wurde zwischen 1 und Anzahl der Merkmale variiert. Dabei wurde bei Netzen mit gleicher Klassifikationsleistung das Netz mit der geringsten Anzahl an versteckten Neuronen bevorzugt.

Die Bestimmung des Abbruchkriteriums erfolgte mittels Kreuzvalidierungsmenge. Durch die Verwendung von Validierungsdaten kann auf einfache Weise das Overfitting der Netze auch bei höherer Anzahl an versteckten Neuronen verhindert werden.

Merkmal / Variante	1	2	3	4	5	6
CFG- basierter Spracherkenner	x	x	x			
nGram- basierter Spracherkenner				x	x	x
Konfidenz: a Score	x		x	x		x
Konfidenz: a posteriori Vorgängerwort	x			x		
Konfidenz: a posteriori Nachfolgerwort	x			x		
Konfidenz: Consensus	x	x	x	x	x	x
Konfidenz: Consensus Vorgängerwort	x	x		x	x	
Konfidenz: Consensus Nachfolgerwort	x	x		x	x	
Wortlänge			x			x
a posteriori Wahrscheinlichkeit			x			x

Tabelle 12.3: Merkmalswahl für neuronale Netze

Ergebnisse

Variante	Neuronen in versteckter Schicht	korrekt erkannt	inkorrekt erkannt	Rate (%)	Recall	Precision	F-Measure
1	1	86	66	56,6	0,01	0,44	0,23
2	3	87	65	57,2	0,03	0,67	0,35
3	2	86	66	56,6	0,01	0,44	0,23
4	**3**	**128**	**77**	**62,4**	**0,67**	**0,76**	**0,72**
5	2	144	61	70,2	1,00	0,69	0,85
6	2	86	110	42,0	0,48	0,44	0,46

Tabelle 12.4: Ergebnisse der neuronalen Netze

Die Ergebnisse der neuronalen Netze, welche in der Tabelle 12.4 gegenübergestellt sind, variieren stark zwischen nGram- basiertem Spracherkenner und CFG- basiertem Spracherkenner. Das CFG- Sprachmodell hat einen sehr niedrigen Recallwert. Dies dedeutet, daß viele korrekt verstandene Wörter als falsch erkannt betrachtet werden. Beim nGram- basierten Spracherkenner finden wir das Gegenteil. Der Recallwert von 1.00 signalisiert uns, daß alle Wörter als korrekt erkannt angenommen wurden.

Am Besten schneidet die Variante 4 des nGram- basierten Spracherkenners ab, welche die Konfidenz auf Basis des Consensus und die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit für letztes, dieses und nächstes Wort betrachtet. Die benötigte Anzahl an versteckten Neuronen liegt bei 3. Hier werden für $62,4\%$ der untersuchten Wörter die richtige Einordnung durch das neuronale Netz vorgenommen und der Recall und die Precision ist ausgeglichen, was für einen guten Kompromiss aus alle richtig erkannte Wörter finden und falsche Wörter zurückweisen steht.

Die Variante 6, welche auf den ersten Blick die besten Ergebnisse unter Verwendung aller vorgestellten Merkmale erreicht, wird nicht als das beste Klassifikationsergebnis betrachtet, da der Recall von 1.00 zeigt, daß die Tendenz gegeben ist, alle Wörter als korrekt erkannt zu bestimmen.

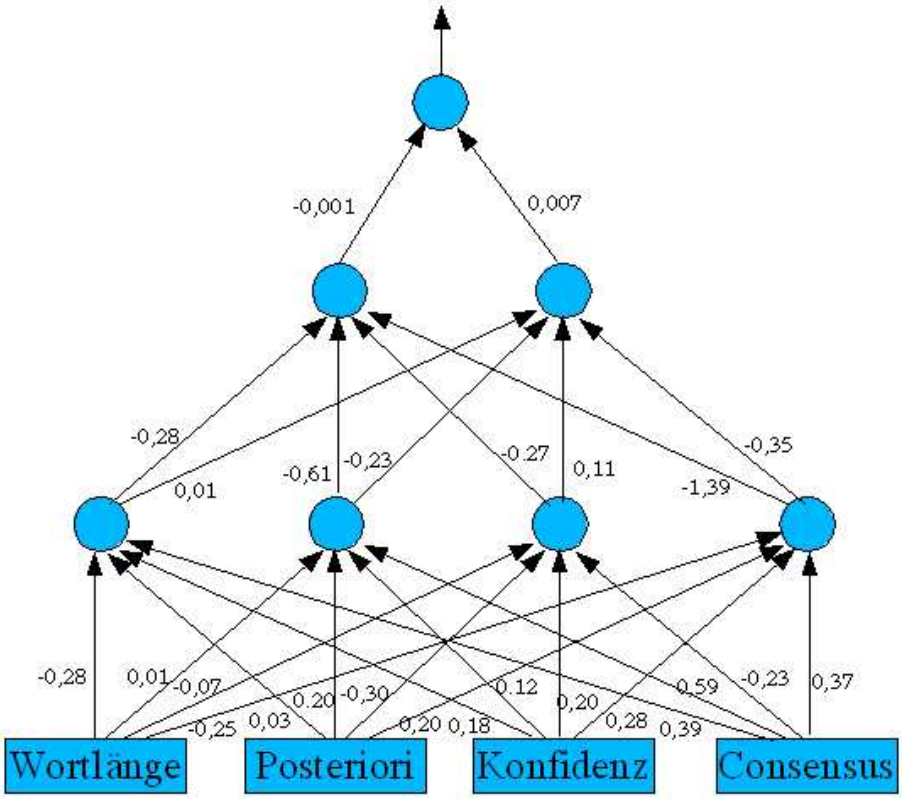


Abbildung 12.6: Die Abbildung zeigt das neuronale Netz mit all seinen Gewichten. Die verwendete Eingabe ist die Wortlänge, der a posteriori Score, die Konfidenz auf Basis des a posteriori Wahrscheinlichkeit und der Konfidenz auf Basis des Consensus.
Die Gewichte der neuronalen Netze sind nur schwer zu interpretieren. Aber unter der Annahme, daß eine lineare Propagierungsfunktion verwendet worden wäre, hat die Konfidenz auf Basis des Consensus den größten Einfluss ($-1,8E^{-3}$) auf das Ergebnis und die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit den kleinsten Einfluss ($-0,3E^{-3}$). Dieses Verhältnis läßt sich zwar nicht direkt auf die nicht lineare Interpretation durch die Sigmoidfunktion übertragen, aber eine Tendenz ist zu erkennen.

12.2.4 Support Vector Machines

Merkmale

Für die Support Vector Machines wurden alle Merkmale verwendet, obwohl dafür genügend Trainingsdaten vorhanden sein müssen, um den Merkmalsraum trainieren zu können.

Konfiguration

Das entscheidende Kriterium beim Einsatz von Support Vector Machines ist die Wahl des verwendeten Kernels. Drei Kernel sollen für den Vergleich verschiedener Support Vector Machines vorstellt werden. Die Kernel sind in der Tabelle 12.5 gegenübergestellt.

Der lineare Kernel nimmt eine Trennung durch eine Hyperebene vor. Der Gauss Kernel unterstellt den Daten eine Normalverteilung und verwendet zur

Variante	Spracherkenner	Kernel
1	CFG- basiert	Linear
2	CFG- basiert	Gauss
3	CFG- basiert	Polynomial
1	nGram- basiert	Linear
2	nGram- basiert	Gauss
3	nGram- basiert	Polynomial

Tabelle 12.5: Konfiguration für Support Vector Machines

Trennung eine Gaussglocke. Der Polynomialkernel kann eine Trennung durch eine Funktion beliebigen Grades vornehmen. Der Grad der Funktion muß aber im Voraus als Parameter angegeben werden.

Ergebnisse

Variante	korrekt erkannt	inkorrekt erkannt	Rate (%)	Recall	Precision	F-Measure
1	195	136	58,9	1,00	0,59	0,80
2	195	136	58,9	1,00	0,59	0,80
3	195	136	58,9	1,00	0,59	0,80
4	195	136	58,9	1,00	0,59	0,80
5	195	136	58,9	1,00	0,59	0,80
6	195	136	58,9	1,00	0,59	0,80

Tabelle 12.6: Ergebnisse des Support Vector Lernens

Die Support Vector Machines erreichen, wie man in der Tabelle 12.6 sehen kann, schlechte Ergebnisse. Es ist aber auffällig, daß es zwischen den verschiedenen Kerneln keine Unterschiede in den Ergebnissen gibt. Der Recall ist für alle Kernel bei 1.00. Dies bedeutet, daß keine vom Spracherkenner falsch bestimmten Wörter in die Klasse der korrekt erkannten Wörter klassifiziert wurden. An der dagegen niedrigen Precision sieht man aber, daß einige vom Spracherkenner falsch erkannten Daten als korrekt erkannt eingestuft wurden.

Betrachtet man die Daten näher, so ergibt sich, daß alle Daten als korrekt klassifiziert betrachtet werden. Damit bringen uns die Support Vektoren keinen weiteren Mehrwert. Sie werden darum nicht weiter betrachtet

12.2.5 Bayes Klassifikator

Merkmale

Für den Bayes Klassifikator wird der komplette vorgestellte Merkmalsraum verwendet. Dies ist sinnvoll, weil die Merkmale voneinander unabhängig betrachtet werden. Dadurch muß trotz höherer Anzahl an Merkmalen kein umfangreicherer Merkmalsraum trainiert werden.

Konfiguration

Da es sich bei den Merkmalen um kontinuierliche Merkmale handelt, können nicht einzelne Ausprägungen der Merkmale gezählt werden, sondern es muß eine Verteilung geschätzt werden. Als Verteilungen bieten sich die Normalverteilung oder Mixturen von Gaussverteilungen an. Unter Mixturen von Gaussverteilungen versteht man die Verwendung von mehreren Normalverteilungen, um eine beliebige Abbildung zu repräsentieren.

Variante	Spracherkenner	Verteilung	Anzahl Verteilungen
1	CFG- basiert	Normalverteilung	1
2	CFG- basiert	Mixturen von Gaussverteilungen	2
3	CFG- basiert	Mixturen von Gaussverteilungen	3
4	CFG- basiert	Mixturen von Gaussverteilungen	4
5	CFG- basiert	Mixturen von Gaussverteilungen	5
6	nGram- basiert	Normalverteilung	1
7	nGram- basiert	Mixturen von Gaussverteilungen	2
8	nGram- basiert	Mixturen von Gaussverteilungen	3
9	nGram- basiert	Mixturen von Gaussverteilungen	4
10	nGram- basiert	Mixturen von Gaussverteilungen	5
11	nGram- basiert	Mixturen von Gaussverteilungen	6
12	nGram- basiert	Mixturen von Gaussverteilungen	7
13	nGram- basiert	Mixturen von Gaussverteilungen	8

Tabelle 12.7: Konfiguration für Bayes Klassifikatoren

Für den vorgestellten Bayes Klassifikator naiver Bayes könnte die in der Tabelle 12.7 vorgestellten Konstellationen untersucht werden.

Ergebnisse

Gute Ergebnisse erzielt auch der Bayes Klassifikator (siehe Tabelle 12.8). Zur Approximation der Verteilung der Daten werden 4 bzw. 7 Normalverteilungen benötigt. Damit können fast 70% der Daten in die richtige Klasse eingeordnet werden. Dies war nicht zu erwarten, wenn man sich die Verteilungen der Merkmale betrachtet. Man sieht, daß sich die Kurven für korrekt und falsch erkannte Wörter nicht wirklich unterscheiden. Selbst diese feinen Nuancen versteht der Bayes Klassifikator zu nutzen, um seine Entscheidung zu treffen.

Variante	korrekt erkannt	inkorrekt erkannt	Rate (%)	Recall	Precision	F-Measure
1	142	135	51,3	0,55	0,45	0,50
2	133	144	48,0	0,92	0,45	0,69
3	133	144	48,0	0,82	0,44	0,68
4	173	104	62,1	0,58	0,56	0,57
5	154	123	55,6	0,54	0,48	0,51
6	218	113	65,9	0,73	0,70	0,72
7	194	137	58,6	0,84	0,61	0,73
8	181	150	54,7	0,75	0,59	0,67
9	211	120	63,7	0,64	0,71	0,68
10	219	112	66,1	0,76	0,69	0,78
11	225	106	68,0	0,73	0,73	0,73
12	**228**	**103**	**68,9**	**0,79**	**0,73**	**0,76**
13	225	106	68,0	0,73	0,73	0,73

Tabelle 12.8: Ergebnisse des Bayes Klassifikators

Kapitel 13

Schlußfolgerung

13.1 Vergleich der Bewertungsmethoden

Variante	korrekt erkannt	inkorrekt erkannt	Rate (%)	Recall	Precision	F-Measure
Konfidenz: a posteriori	623	116	84,3	0,84	1,00	0,92
Konfidenz: Consensus	**629**	**110**	**85,1**	**0,88**	**0,96**	**0,92**

Tabelle 13.1: Gegenüberstellung der Ergebnisse der falsch erkannten Sätze

falsch erkannte Sätze Die Tabelle 13.1 zeigt einen Vergleich von der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit und der Konfidenz auf Basis des Consensus bei der Erkennung von falsch erkannten Sätzen. Über 85% der untersuchten Hypothesen werden richtig als korrekt oder inkorrekte semantische Hypothesen erkannt.

Die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit erreicht etwas schlechtere Ergebnisse als die Konfidenz auf Basis des Consensus. Die 85% Marke wird von der Konfidenz auf Basis der a posteriori Wahrscheinlichkeit nicht erreicht. Der Precision Wert von 1.00 zeigt uns, daß die Konfidenz auf Basis der a posteriori Wahrscheinlichkeit kaum eine falsch verstandene Hypothese als korrekt annimmt. Dies ist hilfreich für die weitere Verarbeitung, da keine richtig erkannten Hypothesen aus der normalen Verarbeitung des Dialogsystems herausgenommen werden.

falsch erkannte Satzteile In der Tabelle 13.2 sind die Besten jeder Methode zur Erkennung von falsch erkannten Teilsätzen noch einmal gegenübergestellt. Am besten schneidet der Entscheidungsbaum ab. Er nimmt ein Training auf semantischer Ebene vor.

Die neuronalen Netze erreichten einen sehr guten Recall Wert. Dies wird durch die höhere Anzahl an korrekt erkannten Wörtern in den Trainingsdaten erreicht. Durch die gierige Suche bestimmen die korrekt erkannten Wörter darum die Gewichte stärker.

Das Support Vector Lernen extrahiert aus den Daten keine weiteren Informationen. Durch die gierige Suche wird der minimalste Fehler erreicht, wenn

Variante	korrekt erkannt	inkorrekt erkannt	Rate (%)	Recall	Precision	F-Measure
Konfidenz: a posteriori	218	113	65,9	0,70	0,73	0,72
Konfidenz: Consensus	223	108	67,4	0,76	0,66	0,71
Entscheidungsbaum	234	97	70,7	0,77	0,74	0,76
neuronales Netz	132	73	64,4	0,83	0,71	0,77
Support Vector Machine	195	136	58,9	1,00	0,59	0,80
Bayes	228	103	68,9	0,79	0,73	0,76

Tabelle 13.2: Gegenüberstellung der Ergebnisse der falsch erkannten Satzteile

alle Worte als korrekt erkannt eingeordnet werden. Dies ist aber nicht Ziel der Klassifikation.

Verfahren wie Bayes Klassifikator, Konfidenz auf Basis der a posteriori Wahrscheinlichkeit und Konfidenz auf Basis des Consensus, welche eine Abstraktion der Daten vornehmen, um zu bestimmen, ob ein Wort richtig oder falsch erkannt wurde, erreichen Raten um die 65%.

Falsch erkannte Sätze können mit einer Rate von 85% bestimmt werden. Verbindet man die Erkennung von falsch erkannten Hypothesen und die darin enthaltenen falsch erkannten Wörter, dann können um die 60% der Wörter richtig als vom Spracherkenner korrekt bzw. inkorrekt erkannt werden.

13.2 Ergebnisse

Diese Studienarbeit hat zwei Aspekte, in welcher Form ein Konfidenzmaß das Dialogsytem untestützen kann, untersucht:

- Wahl der Hypothese des Spracherkenners
- Extraktion falsch erkannter Wörter

Die Bestimmung einer Hypothese des Spracherkenners nicht nur auf der a posteriori Wahrscheinlichkeit, sondern auch auf der Konfidenz, hat keine sinnvollen Ergebnisse erzielt. Darum braucht dieser Ansatz nicht weiter verfolgt zu werden.

Die Extraktion von falsch erkannten Wörtern wurde erfolgreich untersucht. Mit einer Wahrscheinlichkeit von 60% können falsch erkannte Wörter bestimmt werden. Auf dieser Grundlage kann das Dialogsystem gezielt nach alternativen Wörtern im Worthypothesengraphen oder besser im Konfusionsnetzwerk suchen. Wird ein Wort gefunden, welches in den semantischen Kontext besser paßt, dann kann dieses Wort vom Dialogsystem weiter verarbeitet werden.

Wenn kein passenderes Wort im Konfusionsnetzwerk gefunden wird, dann kann, wenn das als falsch erkannte Wort kein sinntragendes Wort ist, die weitere Verarbeitung des Dialogsystems ohne Berücksitigung des falsch erkannten Wörters durchgeführt werden. Ansonsten kann das sinntragende Wort einem Parameter der Aktion zugeordnet werden und gezielt nach dieser fehlenden Information im Klärungsdialog gefragt werden.

13.3 Ausblick

Diese Studienarbeit hat sich mit der Erkennung von falsch erkannten Hypothesen und Teilhypothesen beschäftigt. Die Integration dieser Erkenntnis in das Dialogsystem gilt es im Weiteren vorzunehmen. Es müssen sinntragende von Füllwörtern unterschieden werden und eine Zuordnung von falsch erkannten Wörtern zu den Parametern getroffen werden. Erst dann kann gemessen werden, in wie weit die bestimmten Informationen das Dialogsystem verbessern können.

Literaturverzeichnis

[Bohus02] *Integrating Multiple Knowledge Sources for Utterance-Level Confidence Annotation in the CMU Communicator Spoken Dialog System*, Technical Report CS-190, Carnegie Mellon University, Pittsburgh, PA, 2002

[Callan03] *Neuronale Netze im Klartext*, Robert Callan, Pearson Studium, 2003

[Cristianini00] *An Introduction to Support Vector Machines and other kernel-based learning methods*, Nello Cristianini, John Shawe- Taylor, Cambridge University Press, 2000

[Fink03] *Mustererkennung mit Markov-Modellen*, Gernot A. Fink, Teubner, 2003

[Fuegen04] *Tight Coupling of Speech Recognition and Dialog Management*, Christian Fügen, Hartwick Holzapfel, Alexander Waibel, Universität Karlsruhe, Institut für Logik, Komplexität und Deduktionssysteme, 2004

[Hatzen00] *Integrating recognition confidence scoring with language understanding and dialogue modeling*, Timothy J. Hazen, Theresa Burianek, Joseph Polifroni and Stephanie Seneff, In Proceedings of the International Conference on Spoken Language Processing, Beijing, October, 2000

[Kemp97] *Estimating confidence using word lattice*, Thomas Kemp, Thomas Schaaf, In: Proceedings of Eurospeech 97, 5th European Conference on Speech Communication and Technology, Rhodos, Greece 1997. Universität Karlsruhe; Institut für Logik, Komplexität und Deduktionssysteme, 1997

[Mangu99] *Finding consensus in speech recognition: word error minimization and other applications of confusion networks*, Lidia Mangu, Erik Brill, Andreas Stolcke, IBM Watson Research Center, Microsoft Research, SRI International, 1999

[Metze04] *Speech Recognition for Multimodal Interfaces*, Florian Metze, http://isl.ira.uka.de/multimodalCourse/, Interactive System Labs, 2004

[Mitchell97] *Machine Learning*, Tom M. Mitchell, Microsoft Research, 1997

[Rogina95] *The Janus speech recognizer*, Ivica Rogina, Alexander Waibel, Universität Karlsruhe, Institut für Logik, Komplexität und Deduktionssysteme, 1995

[SanSegundo01] *Confidence Measures for Spoken Dialogue Systems*, Ruben San-Segundo, Bryan Pellom, Kadri Hacioglu, Wayne Ward, ICASSP'2001, Mayo 5-11, Salt Lake City, Utah, USA, 2001

[Schölkopf00] *Statistical Learning and Kernel Methods*, Bernhard Schölkopf, Microsoft Research, 2000

[Vapnik00] *The Nature of Statistical Learning Theory*, Vladimir N. Vapnik, Springer, 2000

[Waibel04] *Vorlesungsunterlagen kognitive Systeme*, Alexander Waibel, http://wwwiaim.ira.uka.de/Teaching/VorlesungKogSys/, 2004

[Zöllner03] *Vorlesungsunterlagen machinelles Lernen*, Marius Zöllner, Regine Becher, http://wwwiaim.ira.uka.de/Teaching/VorlesungML/, 2003

Index