Robuste Verarbeitung fehlerhafter Segmentierungsergebnisse[1]

F. Kummert, G. A. Fink, G. Sagerer

Universität Bielefeld, AG Angewandte Informatik,
Postfach 100131, 4800 Bielefeld 1

1 Motivation

Ein Hauptproblem bei der Verarbeitung fehlerhafter Segmentierungsergebnisse ist die Entscheidung darüber, wann eine ausreichende Interpretation vorliegt und die Analyse abgebrochen werden kann. Ein solches Analyseergebnis soll einerseits einen möglichst großen Teil der Segmentierungsresultate interpretieren. Andererseits sollte keine vollständige und extrem zeitraubende Suche im Zustandsraum zur Erzeugung der gewünschten Interpretation nötig sein. Es ist praktisch unmöglich ein solches Verhalten mit à priori definierten, statischen Abbruchkriterien zu erreichen. Eine mögliches statisches Kriterium wäre z.B. zu fordern, daß nur ein gewisser Prozentsatz der Segmentierungsdaten interpretiert werden muß. Setzt man diese Schranke hoch an, so lassen sich bei nur wenig fehlerbehafteten Daten gute Ergebnisse erzielen, die in der Regel auch eine gute Ausnützung der bereitgestellten Information gewährleisten. Treten in den Basisdaten jedoch viele **nicht behebbare** Fehler auf, kann keine Interpretation mehr erzeugt werden, die den Anforderungen genügt. Wird der Schwellwert niedriger angesetzt, gewinnt der Analyseprozeß an Robustheit, verliert aber unter Umständen Teile der Interpretation, wenn strukturelle Restriktionen innerhalb der Analyseergebnisse eine Interpretation der Daten über den gewählten Prozentsatz hinaus nicht erfordern.

An einem Beispiel aus der Sprachverarbeitung soll diese Problematik veranschaulicht werden. In einem sprachverstehenden System, das Dialoge aus dem Bereich Zugauskunft führen kann [Nie92, Sag88], wird als Kriterium für eine erfolgreiche Interpretation neben der strukturellen Konsistenz eine ausreichende Übereckung des Sprachsignals durch interpretierte Worthypothesen gefordert. Betrachten wir die folgenden beiden Dialoge, bei denen als Segmentierungsergebnis die beste Wortkette der akustischen Analyse verwendet wird. Erkannte gesprochene Wörter sind fett gedruckt.

Dialog A:

Benutzer: **ich möchte morgen früh nach Dresden fahren**
System: *Sie möchten morgen früh von Bielefeld nach Dresden fahren?*
Benutzer: wie **ja** am

[1] Diese Arbeit wurde vom Bundesministerium für Forschung und Technologie (BMFT) unter der Nummer 01IV102A0 gefördert.

Dialog B:

Benutzer: **ich möchte** uns acht den Ulm **nach Nürnberg fahren**
System: *Wann möchten sie fahren?*
Benutzer: **um achtzehn Uhr** den
System: ...

Fordert man eine hohe Überdeckung des Sprachsignals, so wird die erste Äußerung in Dialog A vollständig interpretiert. Bei der stark gestörten Bestätigung des Benutzers ist jedoch keine entsprechende Analyse möglich. Ebensowenig kann die erste Äußerung in Dialog B analysiert werden, da die Zeitangabe *"um achtzehn Uhr"* aus den Ergebnissen der Worterkennung nicht rekonstruiert werden kann. Eine gewählte niedrige Überdeckung erlaubt zwar in solchen Fällen eine Interpretation, es wird aber evtl. vorhandene Information nicht verarbeitet. Die Struktur der Interpretation einer Anfrage an das System fordert z.B. keine Zeitangabe, weshalb Dialog A um eine Rückfrage des Systems erweitert würde. Trotzdem stellt dieses Verfahren keineswegs immer einen Fortgang des Dialogs sicher. Extreme Störungen oder das Fehlen strukturell wichtiger Teile der Interpretation können auch hier zum Scheitern führen.

Um diesem Problem zu begegnen wurde ein Verfahren zur Verarbeitung partieller Interpretationen entwickelt, das mit einem dynamischen Kriterium arbeitet, um über die Vollständigkeit von Interpretationen zu entscheiden. Es erlaubt zum einen die vollständige Verarbeitung aller überhaupt interpretierbaren Segmentierungsergebnisse und stellt zum anderen sicher, daß Fehlersituationen erkannt werden und danach mit Teilinterpretationen unterschiedlichen Vollständigkeitsgrades weitergearbeitet werden kann.

2 Verarbeitung partieller Interpretationen

Um in einem Musteranalysesystem die Verarbeitung partieller Interpretationen zu ermöglichen, müssen die folgenden Probleme gelöst werden:

- Analyseabbruch: Wann kann aufgrund der Geschichte der bisherigen Analyse entschieden werden, daß keine bessere Interpretation gefunden werden kann?
- Wiederaufsetzen: Welche der gewonnenen Zwischenergebnisse kommen als mögliche partielle Interpretationen in Frage und welches davon ist das beste?

Für die Beurteilung des **Analyseabbruchs** ist es wichtig, von einer lokalen Bewertung von Teilergebnissen weg zu einer Bewertung des Fortgangs der Analyse zu kommen. Geht man davon aus, daß der Suchraum der Analyse als Baum vorliegt, so gilt es zu verhindern, daß extensive Tiefen- oder Breitensuche durchgeführt wird, ohne daß neue Information gewonnen wird.

Welche Zwischenergebnisse für ein **Wiederaufsetzen** in Betracht kommen ist natürlich stark vom Anwendungsbereich abhängig. Allgemein erscheint es jedoch sinnvoll folgende Eigenschaften von partiellen Interpretationen zu fordern:

- Vollständigkeit bei elementaren Strukturen: Dies ist wichtig, damit die Interpretation ein gewisses Maß an inhaltlicher Relevanz erlangt.
- Teilweise Beziehung zum Analyseziel: Kann eine solche - wenn auch schwache - Beziehung nicht hergestellt werden, so ist es kaum möglich, sie durch besondere Maßnahmen zu erzwingen.

- Weiterverarbeitbarkeit: Diese Forderung kann unter Umständen eine Konsequenz der beiden zuerst genannten sein, soll aber hier explizit erhoben werden, da es je nach Ziel der Analyse unterschiedliche Formen von Teilinterpretationen sein können die eine erfolgreiche Weiterverarbeitung ermöglichen oder auch nur erleichtern.

Für die Abwägung zwischen verschiedenen partiellen Interpretationen ist es wichtig ein Maß für die **Größe** einer Interpretation bereitzustellen. Dies kann z.B. über die Anzahl der verarbeiteten Segmentierungsergebnisse oder auch über die Komplexität der erzeugten Ergebnisstrukturen definiert sein. Bei ähnlich großen Interpretationen können zur weiteren Differenzierung Bewertungsmaße wie **Qualität**, **Sicherheit** oder **Relevanz** herangezogen werden.

3 Anwendung in der Sprachverarbeitung

3.1 linguistische Wissensbasis

Das semantische Netz, das das Wissen für das Erkennen und Verstehen von Äußerungen und für das Führen eines Auskunftsdialogs beinhaltet, umfaßt die folgenden vier Abstraktionsebenen [Kum91, Mas89]:

- Die *Syntaxebene* enthält Konzepte, die zum einen syntaktische Konstituenten wie Verbalgruppe oder Präpositionalgruppe und zum anderen spezielle Zeitangaben wie Datum oder Uhrzeit modellieren.
- Die *Semantikebene* beruht auf Fillmore's Tiefenkasus Theorie [Fil68]. Hierbei wird angenommen, daß ein Verb für eine gewisse Bedeutung Leerstellen eröffnet, denen eine funktionale Rolle (Tiefenkasus) zugeordnet wird. Diese Theorie kann auch auf Nomina übertragen werden, so daß in dieser Ebene Konzepte für die Bedeutung von Verben und Nomina und für deren funktionale Rollen existieren.
- Die Konzepte der *Pragmatikebene* repräsentieren zulässige Benutzeranfragen wie Fahrplanauskunft und anwendungsabhängige Begriffe wie Abfahrtsort (pragmatische Bestimmung) oder "mit einem Zug fahren".
- Die *Dialogebene*, die auf [Mas89] basiert, enthält Konzepte, die die Benutzer- und Systemdialogschritte definieren. Von Benutzerseite sind aktuell die Schritte Informationsfrage, Ergänzung, Korrektur und Bestätigung erlaubt. Das System kann Nachfragen, eine Fahrplanauskunft geben, sich eine Interpretation bestätigen lassen oder um die Wiederholung einer Äußerung bitten.

3.2 Verarbeitungsstrategie

Um die Vorerwartungen der linguistischen Wissensbasis möglichst umfassend zu nutzen, wird eine Hypothese nicht aufgrund einer sequentiellen Abarbeitung des Sprachsignals (Links-Rechts-Analyse) erweitert, sondern aufgrund von strukturellen Beziehungen. Dies bedeutet, daß für die Erweiterung einer Hypothese nicht die aktuelle Überdeckung des Sprachsignals mit Worthypothesen entscheidend ist, sondern die Analyse durch Vorerwartungen, die im semantischen Netz modelliert sind, gesteuert wird. Dadurch wird eine Worthypothese in jedem noch nicht überdeckten Abschnitt des Sprachsignals akzeptiert, sobald sie den Anforderungen der Wissensbasis genügt.

Ziel der linguistischen Analyse ist die Instantiierung eines Konzepts, das eine zulässige Benutzeranfrage repräsentiert, z.B. Informationsfrage, Korrektur. Wegen der unsicheren Worterkennung müssen dabei auch fehlerhafte Segmentierungsergebnisse verarbeitet werden. Wie in Abschnitt 1 dargelegt, ist dafür die Verwendung eines dynamischen Abbruchkriteriums wünschenswert. Zum einen wird die Anzahl der benötigten Dialogschritte gering gehalten, da im Erfolgsfall das gesamte Sprachsignal interpretiert wird, und somit Ergänzungsfragen sich erübrigen. Zum anderen vermeidet man ein Fehlschlagen des Dialogs, da auch sehr fehlerhafte Segmentierungsergebnisse toleriert werden.

Die bei der Verwendung eines dynamischen Abbruchs in Abschnitt 2 dargelegten Probleme (Analyseabbruch, Wiederaufsetzen) wurden folgendermaßen gelöst:

Eine (Teil-)Interpretation wird in dieser Anwendung als gültig für einen vorzeitigen Abbruch bezeichnet, falls

- sie mindestens auf einer Worthypothe basiert und
- nur vollständige syntaktische Konstituenten enthält.

Da während der linguistischen Analyse alle vollständigen Konstituenten einem Pragmatikkonzept zugeordnet werden, stellt die zweite Bedingung sicher, daß jede gültige Hypothese eindeutig interpretiert werden kann. So kann die Interpretation "muß ... nach Dresden" als Verbindungswunsch für eine Fahrt nach Dresden interpretiert werden. Die Hypothese "muß ... Hannover nach Dresden" jedoch ist nicht eindeutig interpretierbar, da sich die zwei Interpretationen "von Hannover nach Dresden" und "über Hannover nach Dresden" anbieten.

Eine gültige (Teil-)Interpretation I_1 ist besser als I_2, falls

- die akustische Qualität $Q(I_1)$ besser als $Q(I_2)$ ist (um statistische Schwankungen des Qualitätsmaßes aufzufangen, wird der Vergleich auf Intervallen durchgeführt) oder
- $Q(I_1)$ liegt im gleichen Intervall wie $Q(I_2)$ und I_1 überdeckt einen größeren Bereich des Sprachsignals.

Die Analyse einer Äußerung wird abgebrochen, falls

- zumindest eine gültige (Teil-)Interpretation existiert und
- nach dem Auffinden der aktuell optimalen (Teil-)Interpretation mehr als n unzulässige Suchbaumknoten erzeugt wurden.

Da ein Suchbaumknoten als unzulässig bewertet wird, falls die zugeordnete Teilinterpretation linguistisch sinnvoll nicht mehr erweitert werden kann, ist die Anzahl der unzulässigen Suchbaumknoten ohne neue gültige Teilinterpretation ein gutes Kriterium für den vorzeitigen Abbruch der Analyse.

4 Ergebnisse

Die dargestellten Ergebnisse wurden unter den folgenden Rahmenbedingungen erzielt.

- Für die Worterkennung wurde das ISADORA-System [Sch91] verwendet.
- Es wurde ohne Sprachmodell gearbeitet. Damit ist die Perplexität gleich der Lexikongröße von 1081.

- Für das Training des Akustik-Moduls wurden je 500 domänenspezifische Sätze von 4 männlichen Sprechern verwendet.
- Die Worterkennung erreicht bei diesem 4-Sprecher System eine Wortakkuratheit von 74.6% und eine Satzerkennungsrate von 35%.
- Die linguistische Analyse arbeitet mit den Worthypothesen, die sich aus den 10 besten Wortketten ergeben.
- Für jeden Dialog spricht der Testsprecher (einer der Trainingssprecher) die erste Äußerung ins Mikrophon und der Analyseprozeß beginnt. Gemäß der Interpretation des Systems und der daraus resultierenden Systemantwort wird eine weitere Äußerung gesprochen. Dies wird solange wiederholt bis der Dialog erfolgreich mit einer Fahrplanauskunft beendet ist oder fehlschlägt.
- Die Tests wurden auf einer DEC-Station 5200 mit 32MB Hauptspeicher und 25 Mips durchgeführt.

Um die Güte des dynamischen Abbruchkriteriums beurteilen zu können, wurden die 13 Dialoge (von insgesamt 50) verwendet, die in einer früheren Testreihe unter ansonsten identischen Bedingungen fehlschlugen. Nach der Integration des dynamischen Abbruchs konnten davon 7 Dialoge erfolgreich geführt werden. Die durchschnittliche Zahl der Benutzerdialogschritte liegt bei 2.3 und die linguistische Analyse benötigt 31.4 Sekunden je Äußerung für die linguistische Analyse. Damit ergibt sich, daß die Verwendung eines dynamischen Abbruchkriteriums und die Interpretation von Teilergebnissen die Robustheit beim Führen eines Auskunftsdialogs stark erhöht.

Literaturverzeichnis

[Fil68] C. Fillmore: *A Case for Case*, in E. Bach, R. T. Harms (Hrsg.): *Universals in Linguistic Theory*, Holt, Rinehart and Winston, New York, 1968, S. 1–88.

[Kum91] F. Kummert: *Flexible Steuerung eines sprachverstehenden Systems mit homogener Wissensbasis*, Dissertation, Technische Fakultät der Universität Erlangen-Nürnberg, 1991.

[Mas89] M. Mast: *Entwicklung und Realisierung eines Dialogmoduls für ein System zum Verstehen kontinuierlich gesprochener Sprache*, Arbeitsbericht der DFG, Lehrstuhl für Informatik 5 (Mustererkennung), Erlangen, 1989.

[Nie92] H. Niemann, G. Sagerer, U. Ehrlich, G. Schukat-Talamazzini, F. Kummert: *The Interaction of Word Recognition and Linguistic Processing in Speech Understanding*, in P. Laface, R. DeMori (Hrsg.): *Speech Recognition and Understanding*, NATO ASI Series F 75, Springer, Berlin, Heidelberg, 1992, S. 425–453.

[Sag88] G. Sagerer, F. Kummert: *Knowledge Based Systems for Speech Understanding*, in H. Niemann, M. Lang, G. Sagerer (Hrsg.): *Recent Advances in Speech Understanding and Dialog Systems*, NATO ASI Series F, Vol. 46, Springer-Verlag, Berlin, 1988, S. 421–458.

[Sch91] E. G. Schukat-Talamazzini, H. Niemann: *Das ISADORA-System – ein akustisch-phonetisches Netzwerk zur automatischen Spracherkennung*, in B. Radig (Hrsg.): *Mustererkennung 1991*, Bd. 290 von *Informatik Fachberichte*, Springer Verlag, Berlin, Heidelberg, New York, 1991, S. 251–258.