

Integration von Erkennung und Interpretation gesprochener Sprache

Gernot A. Fink, Franz Kummert, Gerhard Sagerer

Universität Bielefeld, Technische Fakultät, Angewandte Informatik,
Postfach 100 131, 33501 Bielefeld

1 Einführung

Die Mehrzahl der heutigen sprachverstehenden Systeme sind aus zwei sehr unterschiedlichen Verarbeitungskomponenten aufgebaut, einer meist statistischen Spracherkennungskomponente und einer in der Regel wissensbasiert arbeitenden Komponente zur Interpretation. Die Kommunikation zwischen diesen Verarbeitungseinheiten ist stark eingeschränkt. In der Regel werden von der Spracherkennung nur Worthypothesenmengen berechnet und an die Interpretationskomponente weitergereicht *ohne weitere Interaktion*.

Wir stellen im folgenden ein Konzept zur Integration von Spracherkennung und Sprachverstehen vor, das die Interaktion dieser Komponenten entscheidend verbessert. Es stellt eine gemeinsame Wissensbasis und eine kompatible Wissensstruktur bereit und leistet eine integrierte Steuerung der Analyseaufgabe.

2 Integration

Grundlage für die von uns angestrebte Integration von Spracherkennung und -interpretation ist die Möglichkeit, komplexe sprachliche Einheiten direkt als strukturiertes Modell der Erkennungskomponente zu repräsentieren und von dieser verarbeiten zu lassen [3]. Um die Kompatibilität der in der Erkennung und bei der Analyse eingesetzten linguistischen Restriktionen zu gewährleisten, werden diese automatisch aus der linguistischen Wissensbasis extrahiert. Ihr struktureller Aufbau bleibt dabei erhalten [1].

Die so gewonnenen Modelle können wie ein Sprachmodell dazu benutzt werden, den Erkennungsprozeß zu steuern. Darüberhinaus stellen sie jedoch sicher, daß bereits die Erkennungsergebnisse wieder der in der linguistischen Wissensbasis vorliegenden Struktur genügen [2]. Diese können dann *ohne weitere Suche* auf partielle Interpretationen abgebildet werden.

3 Robuste inkrementelle Verarbeitung

Für die Interpretation einer Äußerung schlagen wir ein Verfahren vor, das diese von links nach rechts abarbeitet und dabei zwischen Prädiktions- und integrierten Erkennungs- und Interpretationsphasen wechselt. Dadurch erfolgt sowohl die Spracherkennung als auch die linguistische Analyse inkrementell.

In einer Prädiktionsphase wird aufgrund des aktuellen Analysezustandes eine Menge von linguistischen Konstituenten bestimmt, die als mögliche Erweiterungen der bislang berechneten Interpretation in Frage kommen. Diese werden an

die Erkennungskomponente übergeben, die auf der Basis der korrespondierenden strukturierten Sprachmodelle einen weiteren Teil der Äußerung abarbeitet. Die strukturierten Erkennungsergebnisse werden auf partielle Interpretationsergebnisse abgebildet und in die bereits vorliegende Interpretation integriert.

Solange die Äußerung nicht vollständig abgearbeitet ist, wird ein weiterer Prädiktionsschritt angestoßen. Andernfalls werden die noch unverbundenen Teile der aktuellen Interpretation zu einer Gesamtinterpretation zusammengefaßt.

Erst in letzter Zeit ist die Spracherkennungsforschung auf das Problem eingegangen, daß nicht notwendigerweise alle vom Benutzer geäußerten Wörter der Erkennungskomponente bekannt sein müssen. Es wurden erste Verfahren zur Detektion sogenannter "unbekannter Wörter" entwickelt.

In einem Erkennungssystem, das auf Konstituentenmodellen beruht, stellt sich dieses Problem in erweiterter Form. Zum einen können unbekannte Wörter verwendet werden, die gebildeten Konstituenten aber in ihrer Struktur bekannt sein. Zum anderen können aber auch Äußerungsstrukturen auftreten, die durch die linguistische Wissensbasis *nicht* abgedeckt werden.

Daher ist auch ein Modell für unbekannte Wörter bzw. Konstituenten Teil jeder Prädiktionsmenge [4]. Dadurch ist das System in der Lage, Teile der Äußerung, die nicht durch seine sprachliche Kompetenz abgedeckt werden, als unbekannt zu klassifizieren und die Interpretation am Ende dieses unbekanntem Bereichs wieder aufzunehmen.

4 Zusammenfassung

Es wurde ein integriertes inkrementelles Verfahren zur Spracherkennung und -interpretation vorgestellt. Linguistisches Wissen wird in Form von automatisch erzeugten Modellen direkt im Erkennungsprozeß angewandt, um strukturierte Erkennungsergebnisse zu erzeugen. Diese können sofort die aktuelle linguistische Interpretation erweitern. Um den Irregularitäten gesprochener Sprache Rechnung zu tragen, ist ein Modell für unbekannte Konstituenten Teil jeder Erkennungsaufgabe.

References

1. G.A. Fink, F. Kummert, and G. Sagerer. Automatic Extraction of Language Models from a Linguistic Knowledge Base. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, *Signal Processing VI: Theories and Applications*, volume 1, pages 547–550. Elsevier Science Publishers, Amsterdam, 1992.
2. G.A. Fink, F. Kummert, G. Sagerer, and E.G. Schukat-Talamazzini. Speech Recognition Using Semantic Hidden Markov Networks. In *Proc. European Conf. on Speech Technology*, pages 1571–1574, 1993.
3. G.A. Fink, F. Kummert, G. Sagerer, E.G. Schukat-Talamazzini, and H. Niemann. Semantic Hidden Markov Networks. In *Proc. Int. Conf. on Spoken Language Processing*, volume 2, pages 919–922, Banff, Canada, 1992.
4. A. Jusek, H. Rautenstrauch, G. A. Fink, F. Kummert, G. Sagerer, J. Carson-Berndsen, and D. Gibbon. Detektion unbekannter Wörter mit Hilfe phonotaktischer Modelle. In *Mustererkennung 94, 16. DAGM-Symposium Wien, Informatik aktuell*. Springer-Verlag, Berlin, 1994. erscheint.