BMBF

Verbmobil

Verbundvorhaben

Beurteilung von Spracherkennerergebnissen auf Phonemebene

Michael Lehning

TU Braunschweig

Vm

Report 69

Juli 1995

Juli 1995

Michael Lehning

Institut für Nachrichtentechnik
Technische Universität Braunschweig
Schleinitzstr. 22
38092 Braunschweig

Tel.: (0531) 391 - 2453
Fax: (0531) 391 - 8218
e-mail: m.lehning@tu-bs.de

Gehört zum Antragsabschnitt: 14.4 Transkription und teilweise Segmentation auf Wortebene

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 N 0 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

Beurteilung von Spracherkennerergebnissen auf Phonemebene *

Michael Lehning
Institut für Nachrichtentechnik, TU Braunschweig

1 Einleitung

Bei der Bewertung von Verbundworterkennern kann als einfachstes Fehlermaß die *Satzfehlerrate* herangezogen werden, die sich aus dem Verhältnis der fehlerhaften Sätze zur Gesamtanzahl aller Sätze ergibt. Die so ermittelte Fehlerrate hat aber nur eine geringe Aussagekraft, da aus ihr kein Rückschluß auf die Art der Fehler gezogen werden kann. Als weitaus günstiger hat sich die Beurteilung auf *Wortebene* herausgestellt. Bei diesem Verfahren [PALLETT1988] [PALLETT1989] werden die Referenz- und Hypothesesätze mit dem Verfahren der Dynamischen Programmierung [SANKOFF1983] gegeneinander ausgerichtet. Bei kleinen Fehlerraten hat sich diese Art der Ausrichtung bewährt. Bei höheren Fehlerraten kann die Beurteilung auf Wortebene zu einer schlechten Ausrichtung führen, da kein Wissen über die *phonetische Ähnlichkeit* zwischen den Wörtern für den Algorithmus zur Wortausrichtung benutzt wird. Bei den im folgenden vorgestellten Verfahren wird die Ausrichtung auf *Phonemebene* vorgenommen. Die Notation auf Phonemebene wird regelbasiert [WOTHKE1991] oder per Aussprachelexikon aus dem orthographischen Text gewonnen.

2 Beurteilung auf Wortebene

Bei der Ausrichtung auf Wortebene wird mittels der Dynamischen Programmierung (DP) die Referenz- zur Hypotheseäußerung ausgerichtet. Als Distanzmaß wird das Levenstheinmaß gewählt. Bei diesem Distanzmaß werden die Einfügungen, Auslassungen und Vertauschungen von Wörtern mit festzusetzender Gewichtung gewertet. Als Ergebnis wird die Ausrichtung geliefert, bei der sich die größtmögliche Übereinstimmung zwischen Referenz- und Hypothesesatz ergibt. Die Auslassung und Einfügung werden meistens gleich gewertet. Das Fehlermaß für eine Ersetzung wird kleiner als die Summe aus Auslassungs- und Einfügungsfehler gewertet (anderenfalls würde nie ein Vertauschungsfehler auftreten, da statt einer Ersetzung vom System eine Auslassung mit anschließender Einfügung erkannt wird).

*Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministers für Forschung und Technologie unter dem Förderkennzeichen 01IV103C gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor

Beispiel:

Referenz:	Dies	ist	ein	Test	für	ein	***	System
Hypothese:	Die	ist	***	Test	für	ein	sie	System
Auswertung:	S		D				I	

Erläuterung: S: Ersetzung (Substitution) D: Auslassung (Deletion) I: Einfügung (Insertion)

3 Beurteilung auf Phonemebene

Wie oben erwähnt kann es bei Sätzen mit hoher Fehlerrate bei einer Ausrichtung, die lediglich auf dem Levenstheinmaß für Wortauslassungen, Wortvertauschungen und Worteinfügungen basiert, zu einer augenscheinlich unsinnigen Ausrichtung kommen:

Beispiel: Ausrichtung auf Wortebene

Referenz:	Wo	hält	mein	Zug
Hypothese:	Fällt	dein	Flug	aus
Auswertung:	S	S	S	S

Als menschlicher Beobachter würde man sicher folgende Ausrichtung bevorzugen:

Referenz:	Wo	hält	mein	Zug	***
Hypothese:	***	Fällt	dein	Flug	aus
Auswertung:	D	S	S	S	I

Die zuletzt angeführte Ausrichtung würde bei einer Ausrichtung auf Phonemebene geliefert werden. Bei einer auf der Levensthein-Distanz beruhenden Ausrichtung auf Phonemebene, bei der die **Phonem**auslassungen, -einfügungen und -vertauschungen gezählt werden, ergibt sich nämlich folgende Ausrichtung:

Referenz:	v	o:	h	E	l	t	m	aI	n	t	s	u:	k	*	*
Hypothese:	*	**	f	E	l	t	d	aI	n	f	l	u:	k	aU	s
Auswertung:	D	D	S				S			S	S			I	I

Aus der Ausrichtung auf Phonembasis kann in einer nachgeschalteten Verarbeitungsstufe die Wortausrichtung gewonnen werden [PICONE1990], indem einer Wortgrenze im Hypothesestring jeweils die nächstliegende Wortgrenze im Referenzstring zugeordnet wird.

3.1 Phonemausrichtung auf der Basis von Verwechslungsmatrizen

Bei der Phonemausrichtung auf Basis des Levenstheinmaßes werden korrespondierende Phoneme bzw. Phonemketten gut gegeneinander ausgerichtet. Probleme treten aber bei der Ausrichtung von Phonemen auf, die nicht identisch, aber lautverwandt sind. Durch das Levenstheinmaß wird der Ähnlichkeitsgrad zwischen den verschiedenen Phonemen nicht berücksichtigt. Eine aus phonetischer Sicht sinnvollere Ausrichtung kann durch die Benutzung von *Kostentabellen für die Phonemvertauschung* gewonnen werden [PICONE1986]. Um das Einfügen und Auslassen von Phonemen zu modellieren, wird in der Verwechslungstabelle ein spezielles Phonem für die *Auslassung* (Nullphonem) eingeführt. Die Kostenfunktion wird somit wie folgt definiert:

$$C(i,j) = D(i,j) + min\,\{C(i-1,j) + D(0,j), C(i-1,j-1), C(i,j-1) + D(i,0)\}$$

Erläuterung:

$C(i,j)$	Kostensumme bis zum Phonem i in der Referenzphonemkette bzw. bis zum Phonem j in der Hypothesenphonemkette (entnommen aus der Verwechslungsmatrix)
$D(i,j)$	Kosten zwischen dem Phonem i in der Referenzphonemkette und dem Phonem j in der Hypothesenphonemkette
$D(0,j)$	Kosten für die Einfügung des Phonems j in der Hypothesenphonemkette
$D(i,0)$	Kosten für das Auslassen des Phonems i der Referenzphonemkette in der Hypothesenphonemkette

3.2 Ausrichtung der Phonemkette mit Hilfe phonologischer Merkmale

Da häufig keine Kostenmatrix zur Schätzung der Kosten für die Phonemvertauschungen zur Verfügung steht, kann eine Abschätzung der phonetischen Ähnlichkeit durch Vergleich der phonologischen Merkmale der Phoneme gewonnen werden [PALLETT1990]. Bei diesem Ansatz wird die Wortausrichtung wie oben aus der phonetischen Ausrichtung gewonnen. Die Kosten für die Ausrichtung zwischen zwei Phonemen werden durch einen auf der DP basierenden Vergleich der binären phonologischen Phonemmerkmalsvektoren [TERNES1987] [TRUBETZKOY1977] gewonnen. Bei diesem Vergleich werden die Kosten für eine Ersetzung extrem hoch gewählt, so daß keine Vertauschung von Merkmalen vorgenommen wird. Außerdem werden die Merkmale in einer stets gleichbleibenden Reihenfolge angegeben. Auslassungen und Einfügungen von Merkmalen werden jeweils mit dem Gewicht *1.0* gewichtet. Die auf diese Weise gewonnenen Kosten werden von der Merkmalsebene über die Phonemebene bis zur Wortebene hochgereicht. Das sich so ergebende Distanzmaß kann als Summe der Phonemmerkmale interpretiert werden, die geändert werden müßten, um die eine Phonemrepräsentation in die andere zu überführen. Als binäre Merkmale können nach [TERNES1987] z.B. folgende Parameter für das Konsonantensystem des Deutschen herangezogen werden: *konsonantisch (K), dauernd (d), anterior (a), koronal (k), nasal (n), lateral (l), stimmhaft (s), verzögert (v)* bzw. für das Vokalsystem: *hoch (Ho), niedrig (N), hinter (Hi), rund (R), lang(L).* Dies Verfahren weist gegenüber der Phonemausrichtung mittels Levenstheindistanz zwei Vorteile auf:

- Verwandte Laute liefern bei der Vertauschung wie gewünscht ein geringes Fehlermaß. Beispiel:

Phonem	Phonologische Merkmale (s.o.)	Ausrichtung der Merkmale	Erläuterung
$[m]$	K a n s	K a * n s	konsonantisch(K), anterior(a), nasal(n), stimmhaft(s)
$[n]$	K a k n s	K a k n s	konsonantisch(K), anterior(a), koronal(k), nasal(n), stimmhaft(s)
			Kosten des Vergleiches: 1

- Das Verfahren der DP kann für alle Stufen der Ausrichtung (Wort-, Phonem- und Merkmalsebene) benutzt werden.

4 Zusammenfassung

Mit den beschriebenen Algorithmen ist es möglich, die Bewertung von Erkennungsergebnissen von Verbundworterkennern zu beurteilen. Es wurde gezeigt, daß die Ausrichtung auf Phonemebene eine aus phonologischer Sicht sinnvollere Ausrichtung zwischen Referenz- und Hypothesenwortkette ermöglicht. Über die eigentliche Auswertung von Erkennerergebnissen hinaus ist es auch denkbar, mit den beschriebenen Verfahren die Komplexität eines Lexikons zu bewerten und daraus Fehlervorhersagen bei Änderung bzw. Vergrößerung des Lexikons vorzunehmen.

Literatur

[PALLETT1988] D. S. Pallett: *Scoring Continuous Speech in the DARPA Speech Recognition Program: Part II.* Proc. of AVIOS '88, American Voice Input/Output Society Conference, San Francisco, CA, 3.-5. October 1988

[PALLETT1989] D. S. Pallett: *Benchmark Tests for DARPA Resource Managment Database Performance Evaluations.* IEEE Proc. International Conference on Acoustics, Speech and Signal Processing, Glasgow, 1989, 536-539

[PALLETT1990] D. S. Pallett, W. M. Fisher, J. G. Fiscus: *Tools for the Analysis of Benchmark Speech Recognition Tests.* IEEE Proc. International Conference on Acoustics, Speech and Signal Processing, 1990, Albuquerque, 97-100 (S2.16)

[PICONE1986] J. Picone, K. M. Goudie-Marshall, G. R. Doddington, W. Fisher: *Automatic Text alignment for Speech System Evaluation.* IEEE Transactions on ASSP, Vol. ASSP-34, No. 4, 780-784, August 1986

[PICONE1990] J. Picone, G. R. Doddington, D.S. Pallett: *Phone- Mediated Word Alignment for Speech Recognition Evaluation.* IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.38, No.3, March 1990, pp 559-562

[SANKOFF1983] D. Sankoff, J.B. Kruskal (eds.): *Time Warps, String Edits, and Macromolecules.* Addison-Wesley Publishing Company, Inc. (Reading, MA) 1983.

[TERNES1987] Elmar Ternes: *Einführung in die Phonologie.* Wissenschaftliche Buchgesellschaft Darmstadt, 1987

[TRUBETZKOY1977] N.S. Trubetzkoy: *Grundzüge der Phonologie [1. Aufl. Travaux du Cercle Linguistique de Prague 7 (1939)]* . Göttingen, 1977

[WOTHKE1991] K. Wothke: *Letter-to-Phone Rules for German.* IBM Report, Document No. TR 75.91.04, 20. Februar 1991, Heidelberg Scientific Center, IBM Deutschland

Dieser Report erschien als Beitrag im Tagungsband:
Fortschritte der Akustik, DAGA 93- Frankfurt a.M, Teil B, 972 - 975