

(12)

Patentschrift

(21) Anmeldenummer: A 2025/2003 (51) Int. Cl.⁷: **G10L 15/20**
(22) Anmeldetag: 2003-12-16
(42) Beginn der Patentdauer: 2006-01-15
(45) Ausgabetag: 2006-11-15

(56) Entgegenhaltungen:
DE 10010232A1
US 20020087307A1

(73) Patentinhaber:
SIEMENS AG ÖSTERREICH
A-1210 WIEN (AT).

(72) Erfinder:
TSCHIRK WOLFGANG
WIEN (AT).

(54) VERFAHREN ZUR OPTIMIERUNG VON SPRACHERKENNUNGSPROZESSEN

(57) Die Erfindung betrifft ein Verfahren zur Optimierung von Spracherkennungsprozessen, bei dem bei jedem Erkennungsvorgang zu jedem Wort der Gesamtmenge (V) der von dem Spracherkennungsprozess erfassten Wörter eine Trefferwahrscheinlichkeits - Hypothese ermittelt wird und bei dem aus der Gesamtmenge eine erste Teilmenge (S) ausgewählt wird, welche einen in der augenblicklichen Situation für diesen Erkennungsvorgang zulässigen Wortschatz umfasst, wobei eine zweite Teilmenge (E) von Wörtern ausgewählt wird, welche den Wortschatz der ersten Teilmenge und zusätzliche zufällig ausgewählte Wörter der Gesamtmenge umfasst und wobei die zu den Wörtern der zweiten Teilmenge gebildeten Hypothesen nach der ermittelten Trefferwahrscheinlichkeit gereiht werden und aus einer vorbestimmten Anzahl (H) der erstgereihten Hypothesen der wahrscheinlichste Treffer ermittelt wird.

$$\begin{aligned} S_0 &\subseteq S_1 \subseteq \dots \subseteq V, \\ 1 &\leq S \leq H, \leq V \\ 1 &\leq H \leq E, S-1 \end{aligned} \quad \begin{aligned} (6) \\ (7) \\ (8) \end{aligned}$$

$$u_1 = \frac{c_1}{(T-1)(1-r_1-r_2)} \quad (9)$$

$$u_2 = \frac{c_2}{1-c_1-r_1} \quad (10)$$

$$u_3 = \frac{c_3}{2(1-r_2)} \quad (11)$$

$$p_i^S = \frac{(S-1)u_1}{1+(E-1)u_1+u_2} \quad \text{if } i > 0 \quad (12)$$

$$p_i^{S,0} = \begin{cases} 1 & \text{if } i = 0 \\ \frac{(E-S-1)u_1}{1+(E-1)u_1+u_2} & \text{if } i > 0 \end{cases} \quad (13)$$

$$p_i^E = \frac{u_2}{1+(E-1)u_1+u_2} \quad \text{if } i > 0 \quad (14)$$

$$q_i^S = \frac{S u_2}{1+(E-1)u_1+u_2} \quad \text{if } i > 0 \quad (15)$$

$$q_i^{S,0} = \begin{cases} 1 & \text{if } i = 0 \\ \frac{(E-S-1)u_2}{1+(E-1)u_1+u_2} & \text{if } i > 0 \end{cases} \quad (16)$$

$$C = \sum_{i=1}^H \prod_{j=0}^{i-1} p_j^{S,0} \quad (17)$$

$$R = \sum_{i=1}^H p_i^E \prod_{j=0}^{i-1} p_j^{S,0} - \prod_{i=1}^H p_i^{S,0} \quad (18)$$

$$A = \sum_{i=1}^H q_i^S \prod_{j=0}^{i-1} q_j^{S,0} \quad (19)$$

$$(E, H) \mapsto \arg \max_{(E, H)} (C, R, A) \quad (20)$$

Technisches Gebiet

Die Erfindung betrifft ein Verfahren zur Optimierung von Spracherkennungsprozessen, bei dem bei jedem Erkennungsvorgang zu jedem Wort der Gesamtmenge der von dem Spracherkennungsprozess erfassten Wörter eine Trefferwahrscheinlichkeits - Hypothese ermittelt wird und bei dem aus der Gesamtmenge eine erste Teilmenge ausgewählt wird, welche einen in der augenblicklichen Situation für diesen Erkennungsvorgang zulässigen Wortschatz umfasst.

Stand der Technik

Beim Einsatz automatischer Spracherkennungssysteme, beispielsweise zur Umwandlung gesprochener Kommandos in elektrische Steuerungsbefehle, wird der Anwender mit dem Problem konfrontiert, dass der Erkennungsvorgang mit einer bestimmten Wahrscheinlichkeit fehlerhafte Ergebnisse liefern wird. Zu diesen fehlerhaften Ergebnissen gehören: das Verwechseln von Befehlen, das fälschliche Rückweisen von Befehlen und das fälschliche Akzeptieren von Störsignalen als Befehle.

Die jeweiligen Wahrscheinlichkeiten für ein Auftreten eines der genannten Fehler hängen voneinander ab, eine geringe Falschrückweisungsrate bedingt meist eine hohe Falschakzeptanzrate und oft auch eine höhere Verwechslungsrate, umgekehrt führt die Forderung nach einer geringen Falschakzeptanzrate auch zu einer höheren Falschrückweisungsrate.

Je nach Anwendungsfall soll nun das Verhältnis der genannten Fehlerarten zueinander optimiert werden. So besteht insbesondere bei Steuerungsaufgaben in lauter Umgebung die Forderung, dass lediglich Kommandos des Benutzers zu einem Steuerungsbefehl führen und die Umgebungsgeräusche mit hoher Zuverlässigkeit zurückgewiesen werden. Hier wird im Interesse einer geringen Falschakzeptanzwahrscheinlichkeit auch eine höhere Falschrückweisungsrate akzeptiert während bei anderen Anwendungen, bei denen der Komfort des Benutzers im Vordergrund steht, die Falschrückweisungsrate niedrig sein soll und dafür eine höhere Falschakzeptanzrate in Kauf genommen wird.

Derartige Spracherkennungssysteme sind beispielsweise aus der DE 100 10 232 A1 bekannt. Diese Schrift beschreibt ein Spracherkennungsverfahren, bei dem ein gesprochener Satz mit vorbestimmten Sätzen eines besonderen Vokabulars verglichen wird, um den Satz dieses Vokabulars zu ermitteln, der phonetisch dem zu erkennenden Satz am ähnlichsten ist. Damit ist dieses Verfahren besonders auf die Bedürfnisse beim automatisierten Lehren einer Fremdsprache abgestimmt.

Aus der US 2002/0087307 A1 ist weiterhin ein Spracherkennungsverfahren bekannt, bei dem der Spracherkennungsvorgang zweifach durchgeführt wird, wobei verschiedene Umgebungsgeräuschmodelle angewendet werden. Die beiden Ergebnisse werden dann miteinander verglichen und die übereinstimmenden Treffer weiterverarbeitet.

Darstellung der Erfindung

Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren anzugeben, mit dem die Eigenschaften eines Spracherkennungsverfahrens hinsichtlich seiner Fehlerarten anwendungsbezogen optimiert werden können.

Erfindungsgemäß wird diese Aufgabe gelöst mit einem Verfahren der eingangs genannten Art, bei dem eine zweite Teilmenge von Wörtern ausgewählt wird, welche den Wortschatz der ersten Teilmenge und zusätzliche zufällig ausgewählte Wörter der Gesamtmenge umfasst und bei dem die zu den Wörtern der zweiten Teilmenge gebildeten Hypothesen nach der ermittelten Trefferwahrscheinlichkeit gereiht werden und aus einer vorbestimmten Anzahl der erstgereihten Hypothesen der wahrscheinlichste Treffer ermittelt wird.

Die Erfindung ermöglicht den optimierten Einsatz eines Spracherkennungssystems mit einer konstanten Erkennungsrate. Durch die geeignete Wahl der zweiten Teilmenge und der Anzahl der erstgereihten Hypothesen kann das Verhältnis der oben genannten Fehlerarten jeder Situation angepasst werden.

Vorteilhafte Ausgestaltungen der Erfindung ergeben sich aus den Unteransprüchen.

Günstig ist es dabei insbesondere, wenn die Größe der zweiten Teilmenge und die Anzahl der erstgereihten Hypothesen, aus denen der wahrscheinlichste Treffer ermittelt wird, mittels Optimierungsverfahren für jeden Erkennungsvorgang festgelegt wird.

Vorteilhaft ist es weiterhin, wenn für jeden Erkennungsvorgang ein eigenes Optimierungskriterium gewählt wird.

Günstig ist es auch, wenn eines der Wörter jeder Menge nicht einem Befehl, sondern der Gesamtheit der möglichen Störsignale entspricht.

Kurzbeschreibung der Zeichnung

Die Erfindung wird anhand einer Figur näher erläutert, welche die wesentlichen Formeln der mathematischen Grundlagen des erfindungsgemäßen Verfahrens darstellt.

Ausführung der Erfindung

Nach dem Stand der Technik wird bei einem Verfahren zur Optimierung von Spracherkennungsprozessen bei jedem Erkennungsvorgang zu jedem Wort der Gesamtmenge V der von dem Spracherkennungsprozess erfassten Wörter, deren Anzahl gleich V sei und die durch ein Umgebungsgeräusch-Muster zu einer Menge V_0 ergänzt wird, eine Trefferwahrscheinlichkeits-Hypothese ermittelt. Das wahrscheinlichste Ergebnis, der Treffer, wird nun entweder aus der Gesamtzahl der Hypothesen oder aus einer ersten Teilmenge S_0 dieser Hypothesen, die S Wörter und ein Umgebungsgeräusch-Muster enthält, ermittelt, wie beispielsweise auch in dem in W. Tschirk, „Neural Net Speech Recognizers. Voice Remote Control Devices for Disabled People,“ e&i Artificial Intelligence 7/8/2001, pp. 367-370, 2001, beschriebenen System.

Zum Beispiel werden bei einem Spracherkennungssystem, welches zur Steuerung der Beleuchtung, der Heizung und des Telephonapparates in einer Wohnung herangezogen wird, nachdem mit einem ersten Kommando die Auswahl der „HEIZUNG“ erfolgt ist, beim nächsten Schritt nur mehr die Wörter „WÄRMER“ oder „KÄLTER“ akzeptiert, nicht aber beispielsweise die Wörter „HELLER“ oder „DUNKLER“ welche in dieser Situation keinen sinnvollen Steuerbefehl ergeben.

Die Wörter „WÄRMER“ und „KÄLTER“ bilden daher in dieser Situation mit dem Muster „Umgebungsgeräusch“ die erste Teilmenge S_0 der Wahrscheinlichkeits-Hypothesen.

Erfindungsgemäß wird nun eine zweite Teilmenge E von E Wörtern ausgewählt, welche den Wortschatz der ersten Teilmenge und zusätzliche zufällig ausgewählte Wörter der Gesamtmenge V_0 umfasst und mit dem Muster eines „Umgebungsgeräusch“ ergänzt zu E_0 .

Die bei einem Erkennungsvorgang zu den Wörtern der zweiten Teilmenge E_0 gebildeten Hypothesen werden nach der ermittelten Trefferwahrscheinlichkeit gereiht und aus einer vorbestimmten Anzahl H der erstgereihten Hypothesen wird der wahrscheinlichste Treffer ermittelt.

Durch geeignete Wahl der Anzahl E der Wörter von zweiter Teilmenge E bzw. ergänzter zweiter Teilmenge E_0 und der vorbestimmten Anzahl H der erstgereihten Hypothesen kann nun die Eigenschaft des Spracherkennungsverfahrens hinsichtlich des Verhältnisses von Falschrückweisungsrate R , Falschakzeptanzrate A und Verwechslungsrate C an die jeweilige Situation

angepasst werden.

So kann beispielsweise in erfahrungsgemäß lauten Situationen die Falschakzeptanzrate besonders niedrig gewählt werden. Die mathematischen Grenzen für die Wahl der zweiten Teilmenge E_0 und deren Größe E sowie der vorbestimmten Anzahl H der erstgereihten Hypothesen sind in den Formeln 0, 1 und 2 der Figur definiert.

Im folgenden wird nun ein vorteilhaftes Optimierungsverfahren näher erläutert. Dazu werden die zu optimierenden Werte für die Größe E der zweiten Teilmenge E_0 und die vorbestimmte Anzahl H der erstgereihten Hypothesen als Funktionen von Falschrückweisungsrate R , Falschakzeptanzrate A und Verwechslungsrate C dargestellt.

Dazu werden die Eigenschaften des Spracherkennungsverfahrens im engeren Sinn zu einem Testvokabular mit einer bestimmten Anzahl T von Wörtern ermittelt. Als Ergebnis werden Testwerte c_T , r_T und a_T zu Falschrückweisungsrate, Falschakzeptanzrate und Verwechslungsrate erhalten.

Daraus lässt sich für den Spracherkennungsprozess ein charakteristisches Konstantentriple $U = (u_1, u_2, u_3)$ ableiten, welches den Spracherkennungsprozess unabhängig von der Größe des zu erkennenden Vokabulars beschreibt.

Die Beziehungen zwischen Konstantentriple $U = (u_1, u_2, u_3)$ und den Testwerten c_T , r_T und a_T zu Falschrückweisungsrate, Falschakzeptanzrate und Verwechslungsrate sind in den Gleichungen 3, 4 und 5 dargestellt.

Damit lassen sich für eine bestimmte Hypothese zu einem analysierten Merkmalsmuster, d.h. Wort oder Umgebungsgeräusch die in den Gleichungen 6 bis 10 dargestellten Aussagen treffen:

Wenn das analysierte Merkmalsmuster ein Wort darstellt, ist die Wahrscheinlichkeit, dass das Spracherkennungssystem dazu eine falsche in der Reihung der Hypothesen an i -ter Stelle aufscheinende Hypothese liefert (d.h. es als anderes Wort oder Umgebungsgeräusch falsch klassifiziert), welche

- ein Wort ergibt, das der ersten Teilmenge S angehört: p_i^S (Glg.6)
- ein Wort ergibt, das der zweiten Teilmenge E , aber nicht der ersten Teilmenge S angehört p_i^{ES} (Glg.7)
- kein Wort sondern ein Umgebungsgeräusch ergibt p_i^G (Glg.8)

Wenn das analysierte Merkmalsmuster ein Umgebungsgeräusch darstellt, ist die Wahrscheinlichkeit, dass das Spracherkennungssystem dazu als Hypothese an der i -ten Stelle der Reihung fälschlich ein Wort angibt, gleich

- q_i^S (Glg. 9) als der Wahrscheinlichkeit, dass ein Wort aus der ersten Teilmenge S fälschlich angegeben wird und
- q_i^{ES} (Glg. 10) als der Wahrscheinlichkeit, dass ein Wort fälschlich angegeben wird, welches der zweiten Teilmenge E , aber nicht der ersten Teilmenge S angehört.

Aus diesen Wahrscheinlichkeiten können nun gemäß den Gleichungen 11, 12, und 13 die Werte für Falschrückweisungsrate R , Falschakzeptanzrate A und Verwechslungsrate C ermittelt werden und für alle zulässigen Kombinationen von zweiter Teilmengengröße E und vorgegebener Anzahl von Hypothesen H gemäß den Gleichungen 1 und 2 Optimalwerte gemäß Gleichung 14.

Patentansprüche:

1. Verfahren zur Optimierung von Spracherkennungsprozessen, wobei bei jedem Erkennungsvorgang zu jedem Wort der Gesamtmenge (V) der von dem Spracherkennungsprozess erfassten Wörter eine Trefferwahrscheinlichkeits - Hypothese ermittelt wird und wobei aus der Gesamtmenge eine erste Teilmenge (S) ausgewählt wird, welche einen in der augenblicklichen Situation für diesen Erkennungsvorgang zulässigen Wortschatz umfasst, *dadurch gekennzeichnet*, dass eine zweite Teilmenge (E) von Wörtern ausgewählt wird, welche den Wortschatz der ersten Teilmenge und zusätzliche zufällig ausgewählte Wörter der Gesamtmenge umfasst und dass die zu den Wörtern der zweiten Teilmenge gebildeten Hypothesen nach der ermittelten Trefferwahrscheinlichkeit gereiht werden und aus einer vorbestimmten Anzahl (H) der erstgereihten Hypothesen der wahrscheinlichste Treffer ermittelt wird.
2. Verfahren nach Anspruch 1, *dadurch gekennzeichnet*, dass die Größe der zweiten Teilmenge (E) und die Anzahl der erstgereihten Hypothesen (H), aus denen der wahrscheinlichste Treffer ermittelt wird, mittels Optimierungsverfahren (14) für jeden Erkennungsvorgang festgelegt wird.
3. Verfahren nach Anspruch 2, *dadurch gekennzeichnet*, dass für jeden Erkennungsvorgang ein eigenes Optimierungskriterium gewählt wird.
4. Verfahren nach einem der Ansprüche 1 bis 3, *dadurch gekennzeichnet*, dass eines der Wörter jeder Menge nicht einem Befehl, sondern der Gesamtheit der möglichen Störsignale entspricht.

Hiezu 1 Blatt Zeichnungen



$$\mathbb{S}_0 \subseteq \mathbb{E}_0 \subseteq \mathbb{V}_0 \quad (0)$$

$$1 \leq S \leq E \leq V \quad (1)$$

$$1 \leq H \leq E - S + 1 \quad (2)$$

$$u_1 = \frac{c_T}{(T-1)(1-c_T-r_T)} \quad (3)$$

$$u_2 = \frac{r_T}{1-c_T-r_T} \quad (4)$$

$$u_3 = \frac{a_T}{T(1-a_T)} \quad (5)$$

$$p_i^S = \frac{(S-1)u_1}{1+(E-i)u_1+u_2} \quad \text{if } i > 0 \quad (6)$$

$$p_i^{E \setminus S} = \begin{cases} 1 & \text{if } i = 0 \\ \frac{(E-S-i+1)u_1}{1+(E-i)u_1+u_2} & \text{if } i > 0 \end{cases} \quad (7)$$

$$p_i^G = \frac{u_2}{1+(E-i)u_1+u_2} \quad \text{if } i > 0 \quad (8)$$

$$q_i^S = \frac{Su_3}{1+(E-i+1)u_3} \quad \text{if } i > 0 \quad (9)$$

$$q_i^{E \setminus S} = \begin{cases} 1 & \text{if } i = 0 \\ \frac{(E-S-i+1)u_3}{1+(E-i+1)u_3} & \text{if } i > 0 \end{cases} \quad (10)$$

$$C = \sum_{i=1}^H p_i^S \prod_{j=0}^{i-1} p_j^{E \setminus S} \quad (11)$$

$$R = \sum_{i=1}^H p_i^G \prod_{j=0}^{i-1} p_j^{E \setminus S} + \prod_{i=1}^H p_i^{E \setminus S} \quad (12)$$

$$A = \sum_{i=1}^H q_i^S \prod_{j=0}^{i-1} q_j^{E \setminus S} \quad (13)$$

$$(E, H)_{opt} = \underset{(E, H)}{\operatorname{argmin}} O(C, R, A) \quad (14)$$

Fig. 1