

Musterlösungen zum 2. Übungsblatt

Sebastian Padó

12. Januar 2005

Aufgabe 1.1b

Laute lassen sich durch die Verteilung der Schallenergie über einzelne Frequenzbänder anschaulich charakterisieren und unterscheiden. Ein Oszillogramm (Schallenergie-Zeit-Diagramm) zeigt nur die Gesamt-Schallenergie für jeden Zeitpunkt an. Das Spektrogramm (Frequenz-Zeit-Diagramm) zeigt durch den Schwärzungsgrad (bzw. die Färbung) die Energie für die einzelnen Frequenzen an, was die Unterscheidung der Laute wesentlich vereinfacht.

Aufgabe 1.1c

Sprachexterne Einflüsse sind vermutlich nicht so schlimm – sie können durch bessere Ausrüstung minimiert werden. Was die beiden anderen Probleme betrifft, kann für beide sinnvoll argumentiert werden.

- Kontinuität des Signals: verhindert sauberer Trennung einzelner Laute und ist dadurch Fehlerquelle, weil einzelne Laute nicht unabhängig voneinander klassifiziert werden können. Das Problem ist signalinhärent (Menschen sprechen nunmal so), man kann es also nicht einfach “lösen”.
- Varianz der Realisierung: Die Abbildung von Sprache zu Schrift ist stark personenabhängig und selbst für einzelne Benutzer nicht eindeutig; dies bedeutet, dass einmal trainierte Modelle für andere Sprecher, andere Umstände etc. schlechter funktionieren. Auch dieses Problem kann nicht einfach umgangen werden, es sei denn, man zwingt alle Sprecher, langsam, deutlich und hochdeutsch zu reden, was den Komfort des Systems deutlich beeinträchtigt.

Aufgabe 1.1d

- Fabien Delacroix: Ausländischer Name, Sprache (Französisch) muss für korrekte Aussprache erkannt werden
- Team: ebenso, englisch (oder Ausnahmenlexikon)

- DoCoMo: Akronym, benötigt besondere Ausspracheregeln (vergleiche USA mit UNO).
- beendete: richtige morphologische Analyse (be-end-e-te) ist nötig, um das “ee” nicht als langes “e” auszusprechen
- 13.: erstens muss die Zahl per Lexikon in ein Wort übersetzt werden, und dann muss die Deklinationseendung (die dreizehnte) aus dem Kontext (hier dem bestimmten Artikel) erschlossen werden
- Tour de Japon: ausländischer Name, s.o.
- 11.: Zahl, s.o.
- 48:34:12: Zeitangabe, muss (vermutlich per Sonderregel) in “48 Stunden, 34 Minuten und 12 Sekunden” übersetzt werden
- ca.: Abkürzung, muss ergänzt werden (Zusatzlexikon)
- 11,000: muss als eine Zahl erkannt werden, dann s.o.
- E.: muss als Währungseinheit erkannt und zu “Euro” ergänzt werden

Aufgabe 1.1e

Concept-to-speech: Hier wird die Äusserung direkt aus einer Bedeutungsrepräsentation erzeugt. Der Vorteil davon ist, dass Bedeutungsrepräsentationen eines gegebenen Systems typischerweise ziemlich gleich aussehen: bei einem Fluginformationssystem hat man vermutlich für jeden Flug die Start- und Zielflughäfen und -zeitpunkte. Diese Information kann in ziemlich genau vorhersagbarer Weise ausgedrückt werden. Anders formuliert: weil der Autor des Systems ziemlich genau weiss, was sein System sagen muss, kann er die Generierungskomponente auf eine natürliche Aussprache der relevanten Sätze optimieren (z.B. “Ihr Flug startet um X Uhr in Y und kommt um Z Uhr an.”). Man kann also erwarten, dass die erreichte Qualität gut bis sehr gut ist.

Im Gegensatz dazu weiss der Autor eines TTS-Systems nicht, was er als Eingabe bekommen wird. Wie in der Vorlesung besprochen, benötigt man ziemlich tiefes Wissen über die Struktur von Sätzen (insbesondere die Syntax), um eine natürliche Aussprache (Prosodie!) erzeugen zu können. Ein TTS-System muss also sehr viel generellere Eingaben verarbeiten können – und wenn die Eingabesätze ungrammatisch sind, hat es trotzdem verloren. Diese Allgemeinheit ist aber gleichzeitig sein Vorteil: es ist universeller einsetzbar als ein CTS-System, das auf eine spezielle Domäne optimiert ist.

Man kann sich den Effekt auch am “Schichtenmodell-Würfel” klar machen (siehe auch Frage 2). Die Invariante war, dass Systeme immer komplexer werden, je mehr von dem Würfel sie abzudecken versuchen. Ein CTS-System hat Zugang zu tieferer Information (weiter unten) und muss dafür Abstriche in Domäne und Umfang machen (nur kleiner Ausschnitt); ein TTS-System muss sich mit flacherer Information (nur oben), kann dafür aber mehr Domänen und einen grösseren Sprachumfang abdecken.

Aufgabe 2a

- Sprachtiefe: die verschiedenen linguistischen Ebenen von Phonemen bzw. Buchstaben (am einfachsten) bis zu Semantik und Kontext (am komplexesten)
- Sprachumfang: Verschiedene Stile (Zeitungsstil, Umgangssprache, gesprochene Sprache, Juristendeutsch, Poesie), von denen jeder besondere Präferenzen für bestimmte sprachliche Strukturen hat
- Domänen: Verschiedene Gegenstandsbereiche (z.B. Wirtschaft, Psychologie, etc.), wo denen jeder ein eigenes Vokabular und oft auch besondere Präferenzen für bestimmte sprachliche Strukturen hat

Aufgabe 2b

Was die Sprachtiefe angeht, so ist tiefe Verarbeitung generell schwierig, weil Bedeutung und Kontext sehr stark mit Weltwissen interagieren und daher schwer unabhängig davon formalisierbar sind¹.

Generell gesprochen, handelt es sich hier um ein Problem der *Robustheit* oder *Generalisierbarkeit*: Es ist sehr schwer, Modelle zu finden, die immer funktionieren, aber machbar, Modelle zu finden, die meistens funktionieren. Wenn man sich daher in einer Dimension auf ein kleineres Gebiet einschränkt, kann man vereinfachende Annahmen machen und sich deshalb in den anderen Dimensionen weiter ausbreiten.

Wenn man sich z.B. auf eine bestimmte Domäne beschränkt, hat man nur mit einem bestimmten Vokabular zu tun und kann detaillierte Bedeutungen für diese Wörter angeben, was tiefe Verarbeitung möglich macht. Beispielsweise haben die Wörter "Bank" und "Schuld" in der Domäne "Wirtschaft" eine klar definierte Bedeutung (Geldinstitut und geliehenes Geld). Will man dieses Modell nun domänenunspezifisch machen, hat das Modell das Problem, dass "Bank" nun auch "Sitzgelegenheit" und "Schuld" auch "moralische Verantwortung" bedeuten kann.

Ähnlich kann man vereinfachende Annahmen über (z.B.) die syntaktische Struktur machen, wenn man den Sprachumfang eines Modells einschränkt; andererseits ist heute fast unmöglich, ein Modell zu bauen, das domänenunabhängige tiefe Verarbeitung mit vollem Sprachumfang macht.

Aufgabe 2c

- Spracherkennung: flache Verarbeitung (Phonetik, evtl. Morphologie), aber nach Möglichkeit domänenunabhängig und voller Sprachumfang
- Dialogsystem: möglichst tiefe Verarbeitung (Bedeutungsrepräsentationen nötig), aber domänenspezifisch (Verbmobil: Terminvereinbarungen) und vermutlich auch mit begrenztem Sprachumfang

¹ Andererseits ist die Formalisierung von Weltwissen ein ziemlich hoffnungsloses Unternehmen.

- Sprachsynthese: mitteltiefe Verarbeitung (benötigt Syntax für Prosodie), weitgehend domänenunabhängig (grosses Lexikon vorausgesetzt), aber vermutlich mit begrenztem Umfang (kann lange, komplexe Sätze nicht gut aussprechen)

Aufgabe 3a

3 Möglichkeiten:

- Ingenieurwissenschaftliches Interesse: Konstruktion sprachverarbeitender Computer
- Linguistisches Interesse: Interesse an (formalen) Modellen von Sprache
- Psychologie: Interesse an menschlichem Sprachverstehen

Natürlich überschneiden sich diese Motivationen jeweils (Details sollten klar sein).

Aufgabe 3b

- Die Modellierung menschlicher Probleme hat zwei potentielle Zwecke: erstens sollen sprachverarbeitende Computer mit Menschen kommunizieren (bzw. von Menschen generierte Texte analysieren); dazu ist es hilfreich, wenn sie über ein Modell der menschlichen Sprachverarbeitung verfügen (Beispiel: damit das Dialogsystem keine Holzwegsätze verwendet). Zweitens ist Sprache ein von Menschen erzeugtes System – insofern kann die Erforschung des menschlichen Sprachverstehens potentiell Hinweise auf die Plausibilität bzw. Unplausibilität verschiedener Sprachmodelle geben.
- Es ist sinnvoll, vom Menschen unabhängige Sprachmodelle zu untersuchen, weil der Computer eine andere “Architektur” hat als der Mensch: während menschliches Denken hochparallel und assoziativ funktioniert, sind Computer sehr gut in einfachen, seriellen Berechnungen. Modelle, die effiziente menschliche Sprachverarbeitung beschreiben, sind also hochwahrscheinlich sehr ineffizient für Computer; es ist also sinnvoll, abstrakte Modelle von Sprache zu untersuchen, die keine Annahmen über die Verarbeitung machen.

Aufgabe 3c

- Modelle des menschlichen Sprachverstehens: Performanzmodelle
- Abstrakte Modelle der Sprache: Kompetenzmodelle