



Inkrementelle Spracherkennung aus mehreren kombinierten Quellen

Abschlussarbeit (Master, evtl. Bachelor)

Motivation: Interaktive Anwendungen erfordern Spracherkennungssysteme, die möglichst verzögerungsfrei (und schon während der Eingabe) möglichst genaue und detaillierte Ergebnisse liefern, beispielweise um Reaktionen schon initiieren zu können, während die Spracheingabe noch läuft (sog. *inkrementelle Spracherkennung* [1]).



Kurzfassung: Ein im Fachbereich bereits vorhandener inkrementeller Spracherkennung [1] entspricht in seiner Erkennungsleistung nicht mehr dem Stand der Technik, wie sie kommerzielle Erkennung (zum Beispiel von Google) erreichen. Googles inkrementeller Erkennung [2] erzeugt allerdings zu wenige Zwischenergebnisse, denen zudem ein präzises Zeit-Alignment der Wörter fehlt. Dadurch sind modalitätsübergreifende Verstehenskomponenten (zum Beispiel “lege dieses und jenes von da nach da” in Verbindung mit Zeigegestenerkennung) kaum umsetzbar.

Ziel der Abschlussarbeit ist es, auf Basis der Kombinationstechnik in [3, 4] und Methoden des Forced-Alignment eine inkrementelle Spracherkennungslösung zu entwickeln, die die Erkennungsleistung des State-of-the-Art mit detaillierten Timing-Informationen erweitert. Der Arbeitsprozess könnte sich in folgende Schritte gliedern:

- Aneignung von Wissen über spontane, gesprochene Sprache und Spracherkennung (z.B. [5]),
- Einarbeitung in den Umgang mit der freien Spracherkennung Sphinx-4 [6], die Nachverarbeitung für Google-Erkennungsergebnisse [3, 4], und in inkrementelle Verarbeitung [7, 8],
- Auswertung erster Erfahrungen mit der inkrementellen Nutzung der Google-ASR [9],
- Implementierung einer Komponente die das Zeitalignment für inkrementelle Google-Ergebnisse möglichst verzögerungsfrei erstellt,
- wenn möglich eine Kombination der Zeitalignmentkomponente mit Sphinx-basierter Spracherkennung um Verzögerungen der Google-ASR zu überbrücken,
- vergleichende Analyse der Erkennungsqualität sowie der Zeitigkeit [10] der entwickelten Methode mit den zugrundeliegenden Standardverfahren.

Die erfolgreiche Bearbeitung des Themas erfordert die Fähigkeit sich in teils umfangreiche vorhandene Softwareprojekte und Algorithmen einzuarbeiten, gute Programmierkenntnisse in JAVA sowie Kenntnisse auf der Kommandozeile. Wissen aus den Lehrveranstaltungen Sprachsignalverarbeitung, Computerlinguistik und/oder Maschinelles Lernen ist von Vorteil, aber nicht notwendig.

Die Abschlussarbeit kann auf Deutsch oder Englisch erarbeitet und verfasst werden; relevante Literatur ist fast ausschließlich auf Englisch verfügbar. Bei jeder Abschlussarbeit streben wir die Veröffentlichung der Forschungsergebnisse auf einer internationalen oder nationalen Konferenz an.

Stichworte: Inkrementelle Verarbeitung, Spracherkennung, Sprachdialogverarbeitung, Mensch-Maschine-Interaktion, Integration komplexer Software-Systeme

Literatur

- [1] Timo Baumann, Michaela Atterer und David Schlangen. „Assessing and Improving the Performance of Speech Recognition for Incremental Systems“. In: *Proceedings of NAACL-HLT 2009*. Boulder, USA, 2009, S. 380–388.
- [2] Ian McGraw und Alexander Gruenstein. „Estimating Word-Stability During Incremental Speech Recognition“. In: *Proceedings of Interspeech*. ISCA. Portland, USA, Sep. 2012.
- [3] Johannes Twiefel. „Development and Evaluation of Semantically Constrained Speech Recognition Architectures“. Master’s thesis. Universität Hamburg, Dept. of Informatics, März 2014.
- [4] Johannes Twiefel, Timo Baumann, Stefan Heinrich und Stefan Wermter. „Improving Domain-independent Cloud-based Speech Recognition with Domain-dependent Phonetic Post-processing“. In: *Twenty-Eighth AAAI Conference on Artificial Intelligence*. Québec City, Canada, 2014.
- [5] Daniel Jurafsky und James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2. Aufl. Pearson International, 2009.
- [6] Paul Lamere, Philip Kwok, William Walker, Evandro Gouvea, Rita Singh und Peter Wolf. „Design of the CMU Sphinx-4 decoder“. In: *Proceedings of EUROSPEECH*. ISCA. Geneva, Switzerland, Sep. 2003, S. 1181–1184.
- [7] David Schlangen und Gabriel Skantze. „A General, Abstract Model of Incremental Dialogue Processing“. In: *Proceedings of the EACL*. Athens, Greece, 2009, S. 710–718.
- [8] Timo Baumann. „Incremental Spoken Dialogue Processing: Architecture and Lower-level Components“. Diss. Universität Bielefeld, Germany, Mai 2013. URN: urn:nbn:de:hbz:361-25819101.
- [9] Svenja Neef. „Portierung und Bewertung von InproTK auf Google Android“. Bachelor’s thesis. Universität Hamburg, Dept. Informatik, 2014.
- [10] Titus von der Malsburg, Timo Baumann und David Schlangen. „TELIDA: A Package for Manipulation and Visualisation of Timed Linguistic Data“. In: *Proceedings of SigDial 2009*. London, UK, 2009.

Kontakt

Timo Baumann (baumann@inf...), Prof. Wolfgang Menzel



URL dieses Dokuments: