

INTENTS IN SPRACHDIALOGEN: EINE PRAXISPERSPEKTIVE

Benjamin Weiss¹, Stefan Hillmann², Sebastian Möller²

*¹Hanyang Universität Seoul, ²Technische Universität Berlin
benniweiss@hanyang.ac.kr*

Kurzfassung: Für den englischsprachigen SGD Datensatz wurden die originale Annotation von Dialogakten und Intents aus der Dialog State Tracking Challenge 8 mit einer auf CUED basierenden, schlankeren Annotation anhand von automatischen Erkennungsergebnissen verglichen. Dazu wurden die beiden Intenterkennung DIET (Rasa, mit drei verschiedenen Embeddings) und ClassifierDLApproach (SparkNLP, mit einem weiteren Embedding) verwendet. Ergebnisse zeigen die höchste Leistung für DIET mit ConveRT Embedding für die Original- und Zielannotation. Die auf CUED basierende Zielannotation ist nicht nur schlanker, und damit universeller und einfacher einzusetzen, sondern zeigt auch für den verwendeten Datensatz eine leicht bessere Erkennung der Dialogakte, was insgesamt für die Verwendung von CUED spricht.

1 Einleitung

Bei der Entwicklung eines sprachgetriebenen Dialogsystems oder Dienstes stellen das Sprachverstehen und das Dialogmanagement zentrale Komponenten dar, die nicht als fertiges Modul übernommen werden können, sondern jeweils spezifisch auf die Domäne und den Aufgabenbereich angepasst werden müssen. Dazu müssen konkrete Nutzereingaben funktional abstrahiert (Sprachverstehen) und interpretiert (User State Tracking) werden, um sinnvolle Systemreaktionen auslösen zu können. Bei der praktischen Entwicklung dieser Komponenten wird als Abstraktion üblicherweise mit sogenannten „Intents“ gearbeitet - also spezialisierte Label von Nutzeräußerungen, die als funktionale Zielklassen des automatischen Sprachverstehens und als bedeutender Teil des Dialogzustandes verwendet werden. Um Äußerungen zu repräsentieren, kommen zu diesen äußerungsbezogenen Intents bei informations- und aufgabenorientierten Systemen noch Entitäten hinzu, die einzelne Informationen beschreiben. Abgesehen von vollständigen „end-to-end“ Ansätzen werden „Intents“ oder verwandte Repräsentationen sowohl in industriellen, als auch in wissenschaftlichen Systemen verwendet. Ziel dieses Beitrages ist die Bereitstellung eines praktisch motivierten Vorgehens zur effizienten Intentdefinition für aufgabenorientierte Sprachdienste, dessen Effektivität anhand der Leistung von Sprachverstehenskomponenten belegt wird.

Mit Beispieläußerungen für verwendete „Intents“ wird ein Klassifizierer trainiert, der üblicherweise nicht die textuelle, sondern eine semantisch-positionelle Repräsentation (Embeddings) als Zahlenvektor verwendet, um auf ähnliche Eingaben abstrahieren zu können. Gängige Beschreibungen und Beispiele von Dialogintents zeigen diese als konkrete Nutzerabsichten einer einzelnen Äußerung, wie etwa das Buchen eines Fluges, dem Erzählen eines Witzes oder der Frage, ob ein Nutzer eine Krankheit (bspw. Covid-19) hat.^{1,2}

¹IBM Watson: <https://cloud.ibm.com/docs/assistant?topic=assistant-intents>

²Amazon Alexa: <https://developer.amazon.com/de/designing-for-voice/what-users-say/#erkenne-die-intents>

Allerdings kann die Nutzerabsicht auch noch abstrakter beschrieben werden: Als allgemeine, kommunikative Funktion, wie sie etwa mit *Dialogakten* beschrieben werden kann: FRAGE, AUFFORDERUNG, BESTÄTIGUNG, ABLEHNUNG, KORREKTUR und so weiter. Die eigentliche spezifische Funktion ergibt sich aus der Verbindung solcher Dialogakte mit konkreten Informationen, wie etwa BITTE mit der Aktivität „buchen“ und der Entität „Flug(ticket)“ (vgl. etwa Jokinen und McTear [1]).

Da diese verschiedenen Ansätze durchaus die Leistung der Spracherkennung beeinflussen können, sollen im Folgenden einige kurze Definitionen zu den wichtigsten Fachbegriffen genannt werden, um dann den Vorschlag einer einheitlichen Annotation von Sprachdaten zu begründen. Abschließend wird ein Evaluierungsexperiment vorgestellt, dass die Leistung zweier Spracherkennungsansätze für für Originalannotation eines frei zugänglichen Datensatzes und die vorgeschlagene Zielannotation vergleicht.

2 Intents und Dialogakte

Grundsätzlich kann eine sprachliche Äußerung als kommunikative und damit soziale Handlung betrachtet werden. Mit dem Konzept von *Sprechakten* werden Dimensionen solchen Handelns bezeichnet. Im engeren Sinne bezeichnen Sprechakte (der illokutionäre Akt) das Vollziehen von sprachlichem Handeln in Kategorien wie Drohen, Fragen, Bitten, Informieren, Empfehlen und so weiter [2]. Der konkrete Inhalt dieser Handlung wird hierbei nicht betrachtet, womit auch die genaue Intention des Äußernden nicht erfasst wird. Allerdings formalisieren Sprechakte einen wichtigen Aspekt der sprachlichen Bedeutung und Intention.

Für den Bereich der Dialogformalisierung und Annotation von Mensch-Mensch-Dialogen, als auch von Mensch-Computer-Interaktion, haben sich so genannte *Dialogakte* etabliert. Dabei sind Dialogakte durchaus von Sprechakten inspiriert, jedoch umfangreicher, da sie neben der illokutionären Handlung bspw. auch semantische Informationen und den Dialogkontext explizit umfassen, womit sie auch die kommunikative Funktion abbilden sollen. Beispiele solcher Dialogakte zur Erfassung der kommunikativen Funktion sind oft als dialogische Paare definiert, wie etwa ANGEBOT–ANNAHME, FRAGE–ANTWORT/BESTÄTIGUNG. Damit repräsentiert die kommunikative Funktion, zusammen mit dem semantischen Inhalt, die Intention des Akteurs für diese Äußerung. Verschiedene Taxonomien von Dialogakten sind verbreitet, manche davon aufgaben- und domänenspezifisch, wie etwa für das Projekt „Verbmobil“ (bspw. INIT_DATE, SUGGEST_SUPPORT_DATE, ACCEPT_DATE Alexandersson et al. [3]).

Zur Lösung des Problems verschiedener verbreiteter Annotationsschemata wurde ein allgemeiner ausgerichteter Standard vorgeschlagen, der unterschiedliche, sich ergänzende Dimensionen von Dialogakten definiert (etwa zur Aufgabe, Feedback, Turn-Management oder Kommunikationsmanagement), deren Inventar erweiterbar ist [4]. Dabei wird als zentrale Eigenschaft von Dialogakten definiert, dass sie den Informationszustand der Akteure aktualisieren.

Allerdings sind neben diesen allgemeinen (akademischen) Aktivitäten zu Dialogakten weiterhin aufgaben- und domänenspezifische Definitionen in der kommerziellen Praxis für sprachbasierte Dienste sehr verbreitet. Dazu hat insbesondere das aktuelle Geschäftsmodell von Sprachplattformen beigetragen, die für Chatbots oder Smart Speaker einfach trainierbare Sprachverstehtenmodelle anbieten, mit denen Entwickler bzw. Designer konkrete Nutzeräußerungen erkennen können. Für diese Art von Äußerungsannotation hat sich der Begriff *Intent* durchgesetzt, bspw. für: „[...] *query to book a flight, a statement to provide a departure time, or a command to set an alarm.*“ [5, p. 165]. Allerdings zeigt sich in der Praxis nach subjektiver Erfahrung der Autoren eine uneinheitliche Verwendung von Intents, was natürlich der diversen Unterschiede in der Erfahrung der Entwickler und Domäne der Anwendung, aber auch an der immer noch recht neuen Technologie liegen kann.

Während also Dialogakte, zumindest die jüngeren Ansätze, abstrakte und allgemeingültige kommunikative Funktionen erfassen wollen, vermischen Intents eher funktionale und semantische Informationen (bspw. SET_ALARM, vgl. auch Beispiele in Tab. 1). Stark verkürzt repräsentieren Dialogakte die kommunikative Funktion als Äußerungskategorie, während relevante semantische Informationen – auch zur Domäne und Aufgabe – als Attribut-Wert-Paare abgebildet (bspw. COMMAND(set alarm)) werden.

Diese Unterscheidung erscheint zunächst profan. Allerdings zeigen sich durchaus praktisch relevante Unterschiede: Zum Einen kann für unerfahrene Entwickler/Designer der Begriff *Intent* suggerieren, dass zum Training des Sprachverstehens alle Varianten einer Nutzerintention einem Intent zugeordnet werden, unabhängig von der sprachlichen Oberflächenform und der kommunikativen Funktion. Die kann jedoch aus Klassifikationssicht zu einer schlechteren Separierbarkeit führen, bspw. wenn die beiden intentional ähnlichen Äußerungen *Wo finde ich ein gutes italienisches Restaurant?* und *Ich habe Lust auf Pizza.* gemeinsam kategorisiert werden, da sich diese beiden Äußerungen semantisch und syntaktisch kaum ähneln. Diese Aussage widerspricht durchaus typischen Beispielen, die sich an Entwickler/Designer für kommerzielle Sprachverstehensdienste wenden. Konzeptuell klarer wäre es, hier zwei verschiedene Kategorien zu definieren, die dann beide auf die selbe Funktion verweisen.

Zum Anderen ist das kollaborative Erstellen und Pflegen von Sprachmaterial für Trainingszwecke auf einfache und eindeutige Klassen angewiesen, um konsistente Daten zu erhalten. Eine einheitliche und erprobte Definition oder sogar Taxonomie von allgemeinen Dialogakten mit wenigen Klassen sollte zu einheitlicherer und einfacherer Annotation (bei mehreren Annotatoren) führen. Dies gilt gerade bei wenig geschulten Personen, bspw. beim Einsatz von Crowdsourcing. Zudem sind bereits für einige Sprachen Datensätze vorhanden, die zur Vergrößerung der Trainingsmenge hinzugezogen werden können. Eine solche wird bzw. im vorliegenden Fall genutzt (Abschnitt 3).

Ein konkreter Vorschlag für aufgabenorientierte Sprachdialogsysteme zu einer minimal benötigten Menge von generischen Dialogakten heißt „CUEd Standard Dialogue Acts“³ [6], im Folgenden kurz CUEd. Dieser Standardisierungsvorschlag beschränkt auch das Set an Parameterkombinationen, die als Attribut-Wert-Paare semantische Informationen, bzw. Entitäten, für jeden Dialogakt erlaubt sind. CUEd definiert die folgenden vier Klassen und allgemeine Dialogakte:

- Provide
 - INFORM: Informationen bereitstellen
- Query
 - REQUEST: Werte erfragen
 - REQALTS: alternative Werte/Ergebnisse erfragen oder aktuelle Suchbedingungen verändern
 - REQMORE: nach einer Antwort weitere Informationen erfragen oder eine solche Suche vorschlagen
- Confirmation
 - CONFIRM: Bestätigung (mit Wert)
 - CONFREQ: Bestätigung und REQUEST:
 - SELECT: Auswahl aus zwei Möglichkeiten
 - AFFIRM: Zustimmung („ja“)
 - NEGATE: Ablehnung („nein“), ggf. mit korrektem Wert und ggf. weiterer Bedingung
 - DENY: explizite Ablehnung eines Wertes mit weiterem Parameter
- Housekeeping
 - HELLO: Begrüßung
 - SILENCE: Stille
 - THANKYOU: Nutzer bedankt sich

³CUEd: Cambridge University Engineering Department

- ACK: back-channel
- BYE: Nutzer beendet das Gespräch verbal
- HANGUP: Nutzer beendet das Gespräch durch auflegen
- REPEAT: Aufforderung zur Wiederholung
- HELP: Aufforderung zur Hilfe
- RESTART: Aufforderung zum Neustart
- NULL: keine Aktion

3 Anwendung auf SGD Datensatz

Um den potentiellen Nutzen einer einheitlichen und kleinen Menge an Dialogakten für die Erkennungsleistung eines Klassifikators abzuschätzen, wurde auf einen frei zugänglichen Datensatz verschiedene Verfahren zur Dialogaktklassifikation und Entitätenextraktion angewendet. Bei dem Korpus handelt es sich um den in Englisch vorliegenden Schema-Guided-Dialog⁴ (SGD) [7]. Dieser enthält über 16.000 aufgabenorientierte Dialoge in 16 Domänen,⁵ die von jeweils zwei Personen mit Hilfe eine Dialogsimulation auf Textebene geführt würden. Dieser Datensatz wurde auch 2020 bei der Dialog State Tracking Challenge 8⁶ verwendet. Dort wird „Intent“ als aufgabenorientierte Nutzerabsicht verwendet, was dazu führt, dass ein ganzer Dialog, oder zumindest mehrere konsekutive *Turns* zu einem Intent zählen. Die Annotation nutzt bereits relativ allgemeine Dialogakte, weist aber im Original folgende Unterschiede zu CUED auf:

- multiple Dialogakte, die einen einzelnen Hauptdialogakt zugewiesen wurden
- eine eigene Intentdefinition mit Bezug zur Aufgabe (bspw. INFORM_INTENT, deren Parameterwerte die eigentlichen Aufgaben benennen), die im Original als separate Klasse der Äußerung angegeben ist und der von uns verwendeten CUED Bedingung allerdings den bestehenden Dialogakten (INFORM) zugeordnet wird

Die Transformation der Originalannotation zur CUED-Bedingung ist in Tab. 1 abgebildet. Damit ergibt sich für diesen speziellen Vergleich der originalen SGD Annotation mit dem hier angewendeten CUED eine Vereinfachung durch den Wegfall (als unnötig angesehen) multipler Dialogakte, sowie der Verringerung ihrer Anzahl durch das Wegfallen von SELECT sowie der drei aufgabenbezogenen „Intents“.

4 Evaluation

Um die an CUED angelehnte sparsamere Dialogaktmenge mit der Originalannotation zu vergleichen, wurden diese beiden Bedingungen mit verschiedenen, freien und (nach Eigenaussage) industriell ausgerichteten Sprachverstehensansätzen trainiert und getestet:

- RASA NLU mit dem DIET Erkener⁷ und drei verschiedenen Embeddings BERT, GloVe und ConveRT
- SparkNLP von John Snow Labs⁸ mit dem Universal Sentence Encoder

Beide Ansätze wurden jeweils für die Klassifikation von Dialogakten und den „Intents“ genannten aufgabenbezogenen Dialogakten, sowie bei Rasa-DIET zusätzlich für die Erkennung von Entitäten, verwendet. Dabei wird erwartet, dass die geringere Menge an Dialogakten durch ihre sinnvolle, semantisch-funktionale Definition zu einer Erhöhung der Erkenerleistung führt.

⁴<https://github.com/google-research-datasets/dstc8-schema-guided-dialogue>

⁵Alarm, Banks, Buses, Calendar, Events, Flights, Homes, Hotels, Media, Messaging, Movies, Music, Payment, RentalCars, Restaurants, RideSharing, Services, Trains, Travel, Weather

⁶<https://sites.google.com/dstc.community/dstc8/home>

⁷https://rasa.com/docs/rasa/reference/rasa/nlu/classifiers/diet_classifier/

⁸<https://www.johnsnowlabs.com/spark-nlp/>

Tabelle 1 – Transformation der originalen Annotation auf das CUED Schema

Annotation		
CUED	Original	Beispiel
AFFIRM	AFFIRM	yes that is it
	AFFIRM_INTENT	yes please play it
	AFFIRM_INTENT+INFORM	yes play it on the speaker in kitchen
	INFORM_INTENT+SELECT	sounds good reserve the car please
	INFORM+AFFIRM	perfect and i need a regular ride
	INFORM+INFORM_INTENT +SELECT	that flight will work i need tickets for three
	SELECT	that works for me
INFORM	INFORM	select the 12th of march
	INFORM_INTENT	i am searching for a doctor
	INFORM_INTENT+THANK_YOU	thanks a lot i would like to know my balance
	INFORM+INFORM_INTENT	can you find me a psychologist in san jose
	INFORM+INFORM_INTENT +THANK_YOU	thanks i need to make another payment for 65 bucks
NEGATE	INFORM_INTENT+NEGATE_INTENT	not right now i need some dates for events
	INFORM+INFORM_INTENT +NEGATE_INTENT	no find a one-way flight i want economy
	INFORM+NEGATE	no make it the 10th
	NEGATE	no
	NEGATE_INTENT	no check my balance if you can
	NEGATE+THANK_YOU	no thank you so much
REQALTS	INFORM+REQUEST_ALTS	are there any other tourist attraction type places around
	REQUEST_ALTS	are there any others
REGMORE	REQUEST	what is their phone number
	REQUEST+AFFIRM	yes that sounds good what will it cost me
BYE	NEGATE_INTENT+GOODBYE	no not now thank you
	NEGATE+GOODBYE	no that is all
	SELECT+GOODBYE	that is perfect that is all i needed
	THANK_YOU+GOODBYE	thanks that is all i need
THANKYOU	THANK_YOU	thanks a lot

DIET (Dual Intent and Entity Transformer) nutzt eine Transformer-Architektur, um gemeinsam Dialogakte und Entitäten zu Klassifizieren [8]. Die drei unterschiedlichen Embeddings sind das große kontextuelle Subwort-Embedding BERT [9], kontextunabhängiges Wort-Embedding GloVe (Global Vectors) [10], sowie das für Antwortgenerierung optimierte und als Wort-Embedding genutzte ConveRT (Conversational Representations from Transformers) [11].

SparkNLP unterstützt verschiedene vortrainierte Modelle Wort- und Satz-Embeddings zur Klassifikation von Texten. In dieser Arbeit haben wir zur Dialogakt- und Intent-Erkennung den Universal Sentence Encoder⁹ (USE) verwendet. USE ist eine Implementierung des durch Cer et al. beschriebenen Ansatzes der Verwendung eines Deep Averaging Network (DAN) für die Berechnung der Embeddings [12]. Der Vorteil vom DAN-Ansatz gegenüber einem Transformer sind die weniger komplexen mehrschichtigen feed-forward Netzwerke und der damit einhergehende lineare Zusammenhang zwischen der Länge der Eingabesequenz und dem Berechnungsaufwand für das entsprechende Embedding [12]. Der Klassifizierer von SparkNLP basiert auf ClassifierDLApproach¹⁰.

⁹https://nlp.johnsnowlabs.com/2020/04/17/tfhub_use.html

¹⁰<https://nlp.johnsnowlabs.com/api/index#com.johnsnowlabs.nlp.annotators.classifier>.

4.1 Methodik

Der SDG Korpus wurde textlich normalisiert, um das Format einer Sprachinteraktion anzunähern. So wurden schriftsprachliche Abkürzungen ausgeschrieben, ausschließlich Kleinbuchstaben verwendet, sowie Interpunktion entfernt (die in Tab. 1) gezeigten Beispiele wurden so erzeugt). Zudem wurden die Daten in die notwendigen Formate von Rasa NLU und SparkNLP konvertiert.

Ein eigener Split von 80% zum Training und 20% zum Test wurde mit `train_test_split` von `sklearn` durchgeführt, wodurch die Dialogakte anteilig in beiden Sets gleichmäßig verteilt wurden. Der DIET-Transformer wurde mit den Trainingsdaten für alle drei Embeddings mit jeweils der Standard-Pipeline für 10 Epochen trainiert, wobei die vortrainierten Embeddings verwendet wurden. Dieses Vorgehen trainiert gemeinsam Dialogakte und Entitäten, wobei die tatsächlichen Aufgaben der „Intents“ als Entitäten repräsentiert wurden. SparkNLP wurde auch mit Standardwerten, separat für Dialogakte und Aufgaben trainiert.

4.2 Ergebnisse

Die Klassifikationsergebnisse für das Testset werden anhand der gewichteten und ungewichteten Makro-Mittelwerte über alle Klassen anhand der F1-Scores ausgewertet (siehe Tab. 2). Für den USE-Erkennen wurde separat ein zweiter Dialogakterkennung für die aufgabenbezogenen Intents trainiert (Task), während diese in DIET als Entitäten erfasst wurde. Für SparkNLP wurde auf eine Entitätenerkennung verzichtet und die Labelhäufigkeiten ausbalanciert, weshalb die Anzahl der Beispiele für Tasks (8838) geringer als in den anderen Bedingungen ist. In fett sind die höchsten Werte markiert. Es zeigt sich, dass diese sowohl für die Original Annotation, als auch CUED mit recht schlanken ConveRT-Embeddings und dem DIET-Klassifizierer erreicht werden, was in Einklang mit anderen Vergleichen steht [8]. Wie erwartet, sind höhere F1-Werte durch CUED für Dialogakte erreicht worden. Allerdings sollte für eine praktische Abschätzung der Klassifikationsleistung diese Veränderung im Rahmen einer dialogischen Anwendung evaluiert werden.

Tabelle 2 – Erkennungsergebnisse DSTC8: Gewichtete (wa) und ungewichtete Mittelwerte (ua)

Annotation	Erkennung	Kategorie	Precision (wa/ua)	Recall (wa/ua)	F1 (wa/ua)	Support
Original	BERT+DIET	Dialogakte	.87/.76	.87/.71	.87/.73	32250
Original	BERT+DIET	Entitäten	.79/.68	.68/.43	.72/.50	117714
CUED	BERT+DIET	Dialogakte	.91/.88	.91/.85	.91/.86	31936
CUED	BERT+DIET	Entitäten	.75/.67	.71/.43	.71/.50	117694
Original	ConveRT+DIET	Dialogakte	.88/.76	.88/.75	.88/.75	32250
Original	ConveRT+DIET	Entitäten	.82/.68	.67/.46	.73/.52	117714
CUED	ConveRT+DIET	Dialogakte	.93/.88	.93/.89	.93/.88	31936
CUED	ConveRT+DIET	Entitäten	.78/.65	.70/.43	.72/.45	117694
Original	GloVe+DIET	Dialogakte	.87/.73	.87/.71	.87/.72	32250
Original	GloVe+DIET	Entitäten	.79/.68	.64/.40	.69/.46	121571
CUED	GloVe+DIET	Dialogakte	.91/.87	.91/.86	.91/.86	31936
CUED	GloVe+DIET	Entitäten	.75/.64	.71/.44	.71/.49	121551
Original	USE+ClassifierDLApproach	Dialogakte	.35/.16	.56/.24	.42/.19	41412
CUED	USE+ClassifierDLApproach	Dialogakte	.91/.90	.91/.88	.91/.89	41412
CUED	USE+ClassifierDLApproach	Task	.38/.20	.56/.31	.45/.24	8838

Für diese beste Bedingung (ConveRT+DIET) sind Abb. 1 die Verwechslungsmatrizen für die Originalannotation (links) und die CUED-Annotation (rechts) dargestellt. Auffällige Ver-

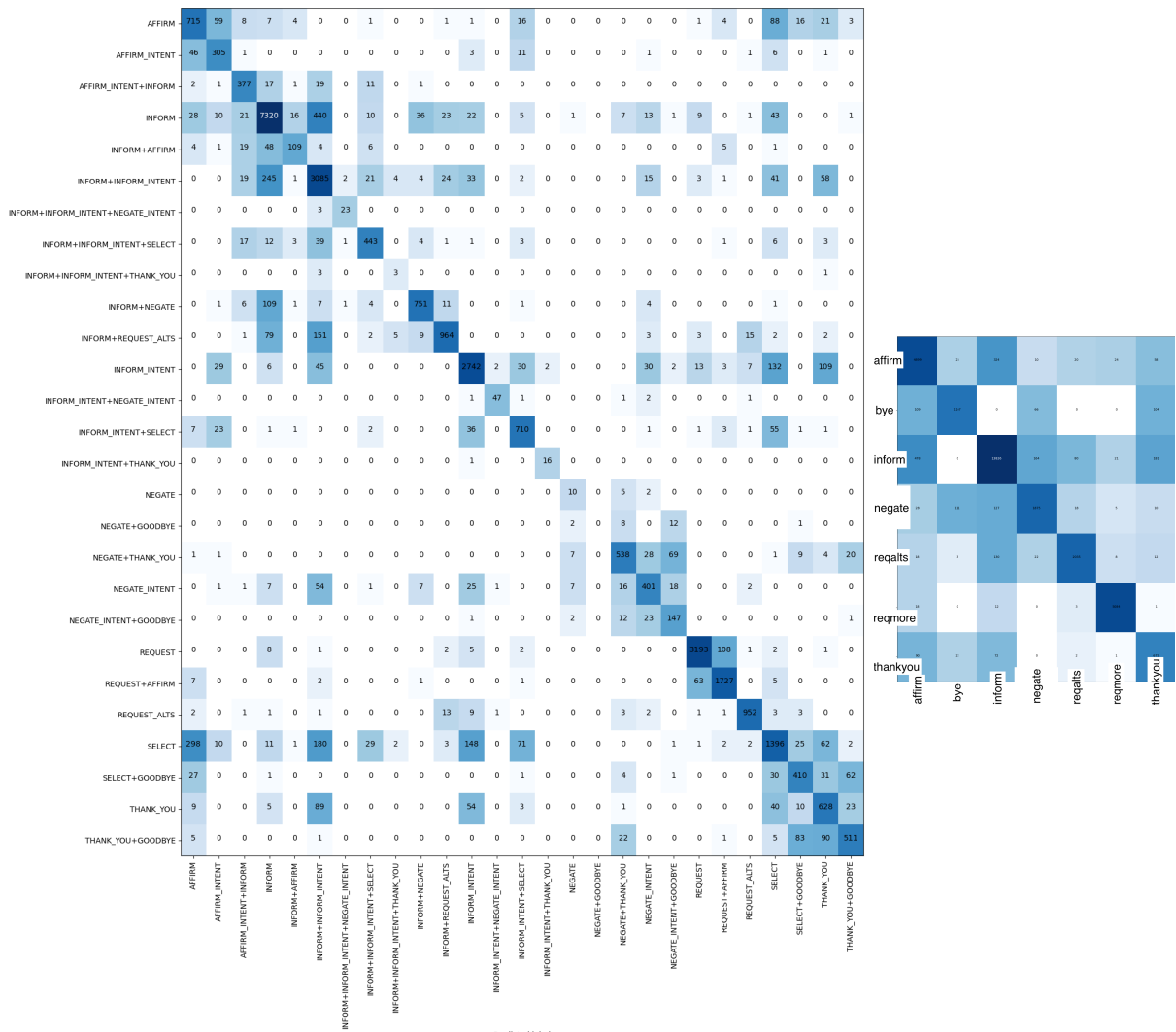


Abbildung 1 – Verwechslungsmatrizen der (links) original- und (rechts) CUED-annotierten Dialogakte for Convert-Embeddings mit DIET Klassifizierer.

wechslungen in den Originaldaten betreffen bspw. SELECT und AFFIRM, INFORM und INFORM+INFORM_INTENT oder REQUEST und REQUEST+AFFIRM, die jeweils im Mapping auf CUED zusammengelegt wurden. Für die CUED-Annotation fallen Verwechslungen zwischen INFORM und AFFIRM auf. Insgesamt sind für das Testset die korrekten Dialogaktklassen jedoch fast durchgehend am häufigsten zugewiesen worden,¹¹ was sich auch in den zufriedenstellenden Werten in Tab. 2 zeigt.

5 Diskussion und Fazit

Diese erste Evaluation eines auf CUED basierenden Ansatzes zur universellen und schlanken Dialogaktannotation anstatt komplexerer oder spezialisierter Dialogakte/Intents zeigt für einen großen und frei zugänglichen Datensatz vielversprechende Ergebnisse. Um ein valideres Bild der Leistung bei der Dialogakt-Klassifikation zu erhalten, sollte die DSCT8 Annotation des SDG-Datensatzes systematisch auf Konsistenz geprüft und ggf. korrigiert werden. Weitere praktische Gesichtspunkte müssen jedoch auch noch evaluiert werden. So muss als weiterer Schritt dieses Vorgehen unter Berücksichtigung von Daten mit noch konkreteren Intentannotationen, einer Restklasse für diese bzw. Dialogakte, sowie von Daten auf Deutsch ausgeführt werden. Insbesondere die potentielle Verlagerung von dialogisch bedeutenden Informationen fein-

¹¹Ausnahmen: Inform+Inform_Intent+Thankyou statt Inform+Inform_Intent sowie Negate+Goodbye

definierter und spezialisierter Intents auf Parameterwerte von allgemeinen Dialogakten muss anhand von gemeinsamen Accuracy-Maßen für Dialogakte und Entitäten bewertet werden. Zudem muss der hier postulierte Vorteil einfacherer Annotationen neuer und Transformationen bestehender Daten praktisch überprüft werden.

6 Danksagung

Wir bedanken uns bei dem Bundesministerium für Bildung und Forschung (BMBF), welches die hier beschriebenen Arbeiten durch Förderung des Projekts OKS - Optimierung konversationeller Schnittstellen (Förderkennzeichen 16SV8151) unterstützt.

Literatur

- [1] JOKINEN, K. und M. MCTEAR: *Spoken Dialogue Systems*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool, 2010.
- [2] SEARLE, J. R.: *Expression and Meaning*. Cambridge University Press, 1979.
- [3] ALEXANDERSSON, J., N. REITHINGER, und E. MAIER: *Insights into the dialogue processing of VERBMOBIL*. In *Proc. 5th Conference on Applied Natural Language Processing*, S. 33–40. 1997.
- [4] ISO 24617-2:2012: *Language resource management – Semantic annotation framework (SemAF), Part 2: Dialogue acts*. 2012.
- [5] MCTEAR, M., Z. CALLEJAS, und D. GRIOL: *The Conversational Interface*. Springer, 2006.
- [6] YOUNG, S.: *CUED standard dialogue acts*. Tech. Rep., Dialogue Systems Group, Cambridge University, 2009.
- [7] RASTOGI, A., X. ZANG, S. SUNKARA, R. GUPTA, und P. KHAITAN: *Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset*. In *Proc. Association for the Advancement of Artificial Intelligence (AAAI)*, S. 8689–8696. 2020.
- [8] BUNK, T., D. VARSHNEYA, V. VLASOV, und A. NICHOL: *DIET: Lightweight language understanding for dialogue systems*. In *arXiv preprint arXiv:2004.09936*. 2020.
- [9] DEVLIN, J., M.-W. CHANG, K. LEE, und K. TOUTANOVA: *BERT: Pre-training of deep bidirectional transformers for language understanding*. In *Proc. NAACL*, S. 4171–4186. 2019.
- [10] PENNINGTON, J., R. SOCHER, , und C. MANNING: *GloVe: Global vectors for word representation*. In *Proc. Empirical Methods in Natural Language Processing (EMNLP)*, S. 1532–1543. 2014.
- [11] HENDERSON, M., I. CASANUEVA, N. MRKŠIĆ, P.-H. SU, T.-H. WEN, und I. VULIĆ: *ConveRT: Efficient and accurate conversational representations from transformers*. In *Proc. Findings of the Association for Computational Linguistics (EMNLP)*, S. 2161–21743. 2019.
- [12] CER, D., Y. YANG, S.-Y. KONG, N. HUA, N. LIMTIACO, R. ST. JOHN, N. CONSTANT, M. GUAJARDO-CESPEDES, S. YUAN, C. TAR, B. STROPE, und R. KURZWEIL: *Universal Sentence Encoder for English*. In *Proc. of the 2018 Conf. on Empirical Methods in NLP: System Demonstrations*, S. 169–174. ACL, Brussels, Belgium, 2018. doi:10.18653/v1/D18-2029.