Building of conversational agents using goal-oriented dialogue managers

Author : Saran Karthikeyan
Supervisors : Jun.-Prof. Ingo Siegert, Otto-von-Guericke Universität Magdeburg
Dipl. Inf. Marcus Petersen, regiocom SE

Agenda

General outline of the Master Thesis work

Introduction and Importance

Interactive Conversational Agents-Importance

- **Interactive conversations** To perform specific tasks: Single- or Multi-domain (Prototype)
- **Emotional labor** Pressure on call-center workforce, Customers from various locations
- **Business Importance** Cost-effectiveness, and Service quality ≅ Revenue
- **Technical Challenges** Hold longer conversation, Direction flow to achieve tasks
- **Research on market analysis** Forecast of Compound Annual Growth Rate of 31.6 % *

- Natural interactions
- Flexibility
- Responsive service
- Clarity

- Reliable service
- Co-operative principles
- Responsiveness

Hello!

Hello! Good Morning. How may I help you?

Could you look for a flight from ***New York to Berlin***?

Do you have any preferred day or time?

Please book for ***Friday at 10:00***

I have a **2** choices available. The ***Lufthansa*** is ***expensive*** and ***Emirates*** in the ***moderate*** price range.

Please book me a flight in ***Lufthansa***

Ok, I have booked a flight on ***Friday morning***, and the reference number is ***LUF569NB***

Thank you for the Service!

Thank you. Have a nice day! See you Again.

Figure 1: A sample interaction in a conversation of **flight-booking** domain

*https://www.reportsanddata.com/press-release/global-chatbot-market

Overview and design objectives

working principle of Interactive Conversational agents

Aspects in design techniques

- **Dataset selection** Suitable domain-specific corpus data and customization
- **Natural Language Processing** Natural Language Understanding (NLU), Dialogue Management (DM), Natural Language Generation (NLG)
- **Custom strategy for DM** custom architecture with Regiocom SE requirements and research standards
- **Business Logic Integration** Including API results and technical error handling of API
- **Live-user Evaluation** User Satisfaction measure

- Utterances
- Intents
- Entities (slots)
- Semantic frames

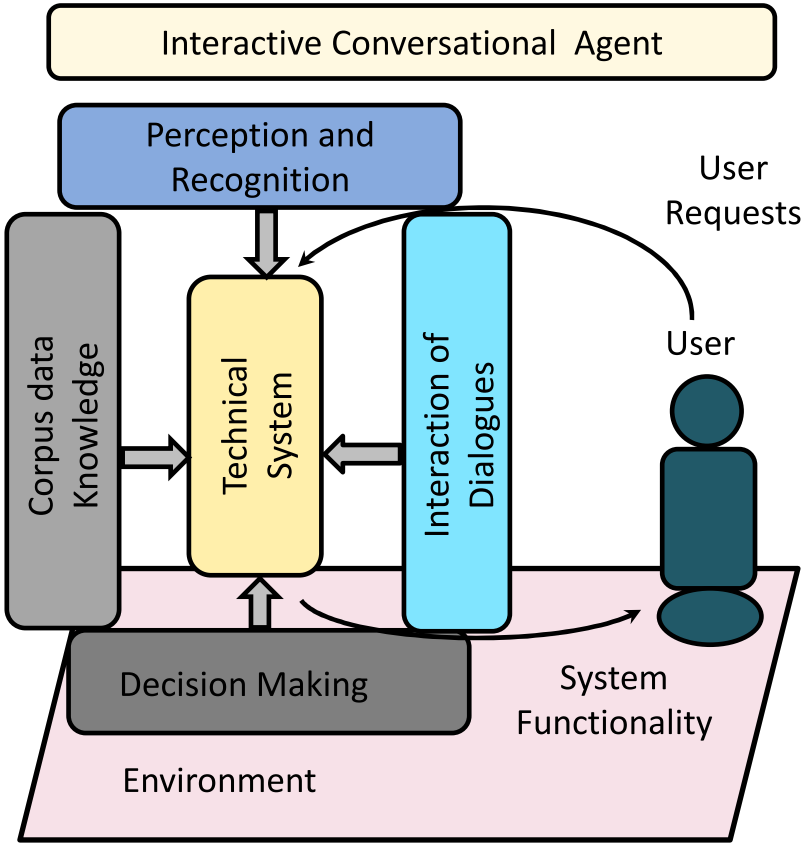

Figure 2: An Illustration of building components of the Interactive Conversational Agents [1]

Software frameworks and related works

Brief outline of overall process in existing approaches

- **Natural vs. Unnatural corpus** Wizard-of-Oz method to collect data for naturalness, Intimation of recording the conversations
- **Human-human conversations** Natural conversation between the humans, less prone to errors in terms of perception
- **Annotation of the corpus data** Annotation of the contextual information, Automation of the entire process
- **Dataset selection** MultiWOZ corpus has selected after analyzing SFX, and WOZ2.0 corpus [2]

+
- Context-rich
- Multiple intents for naturalness
- Availability of intents-slots annotations
- Suitable for Multi-domain settings

Rasa open source conversational AI framework

- **NLU** Rasa NLU for intent classification and entity recognition, Rasa Dialogue State Tracking (DST) [3]

  ! minimum two train samples for each intent
- **DM** Knowledge-driven approach (Deterministic rules and domain-expert)
- **NLG** Template-driven approach (standard templates for each system action)

  + Error-free ! Lacks flexibility

Statistical Techniques
Machine Learning for automation (DM, NLG)
Flexibility, Naturalness

Software frameworks and related works

State-of-Art approaches and challenges inciting a new approach

Technical Background	Advantages	Limitations
a) Ranking scheme for the system actions using LSTM cell by cosine similarity Responses using standard templates [4]	Execution of multiple system action Handling uncooperative Behaviour Attention mechanism for previous predictions	No continual learning Execution of multiple system action is possible in a sequential manner (i.e.) Responses for each system action sequentially using standard templates
b) A centralized agent with Multi-domain learning for joint observations, Multi-task learning and Roleplay [5]	End-to-end learning with less human effort Evaluation using simulator or human	Low success rate

!

- **Catastrophic Forgetting** Diminishment of previously learned representation (LSTM tends to forget data)
- **Continual/Incremental learning** Single or Multi-domain adaptation problem
- **Handling multiple system actions simultaneously** Naturalness in the automated responses
- **Reduction of human effort** Automation of DM and NLG

Implementation of the Proposed work

Architecture Overview-rasa open source

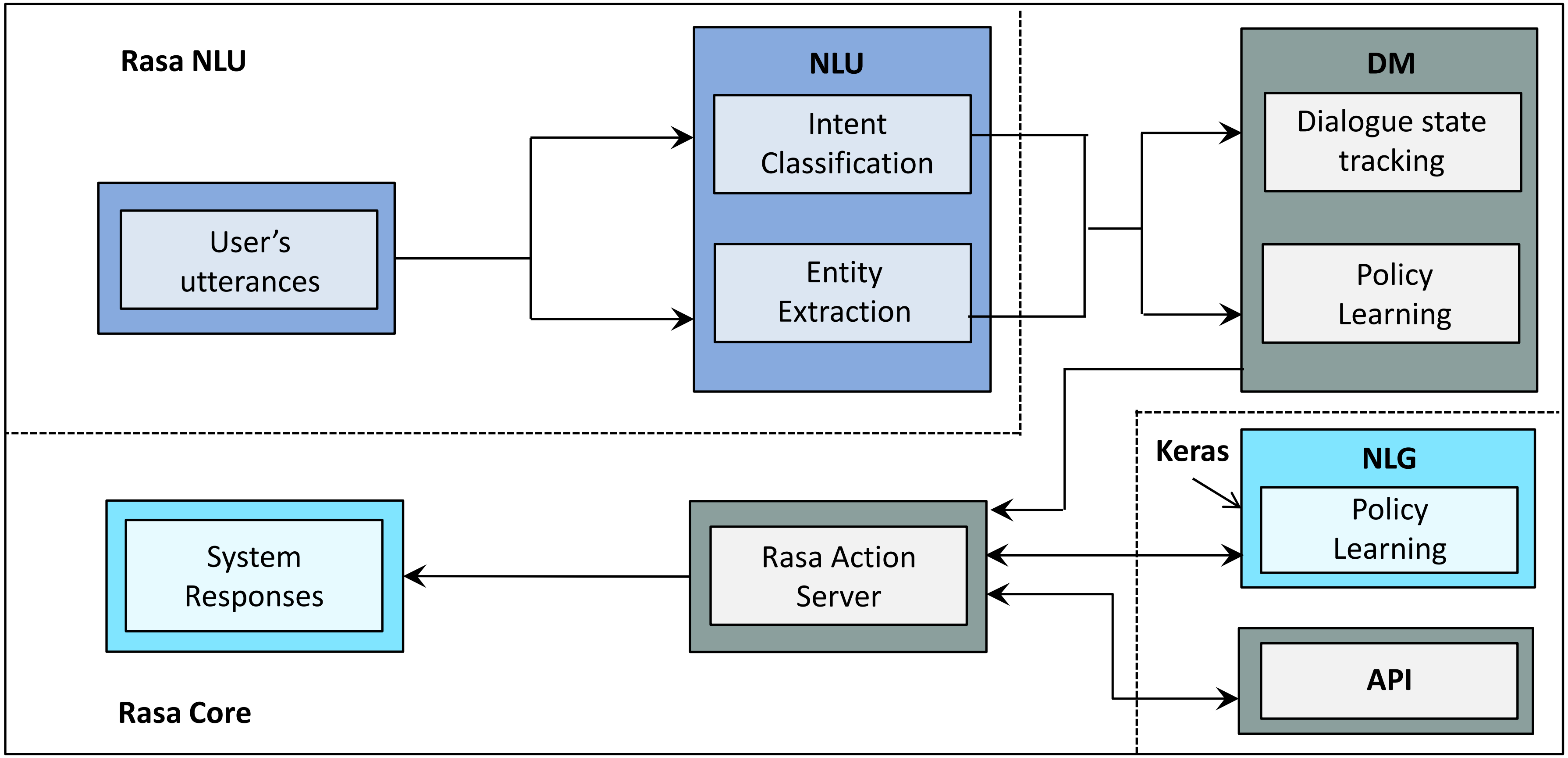

Figure 6: Architecture overview to build Interactive Conversational agent
Using rasa open source

Implementation of the proposed work

Customization of Dataset-Automated approach

- **NLU** user utterances and corresponding annotations
  DM Semantic frames of user and system utterances
  NLG system utterances and respective annotations
- **Multiple intents** Handled as a single string
- **Annotation** Start and end values of each slots
- **Utterances without annotation** Utterances without contextual information has regarded 'out of scope'
- **Multiple possibility of system action** Each user input has multiple possible system action

!

- Multiple intent variants for a similar context
- Imbalance in data distribution (DM)
- Corpus data specific to API database (DM)
- Improper de-lexicalization due to lack of annotation (NLG)

Table 2: Metric information in restaurant domain

Metric	MultiWOZ (Restaurant)
Dialogues	**11537**
Total Turns	**5759**
Avg. no. of turns	**4.3928**
Slots	**11**

Implementation of the proposed work

Custom architecture for Dialogue Management

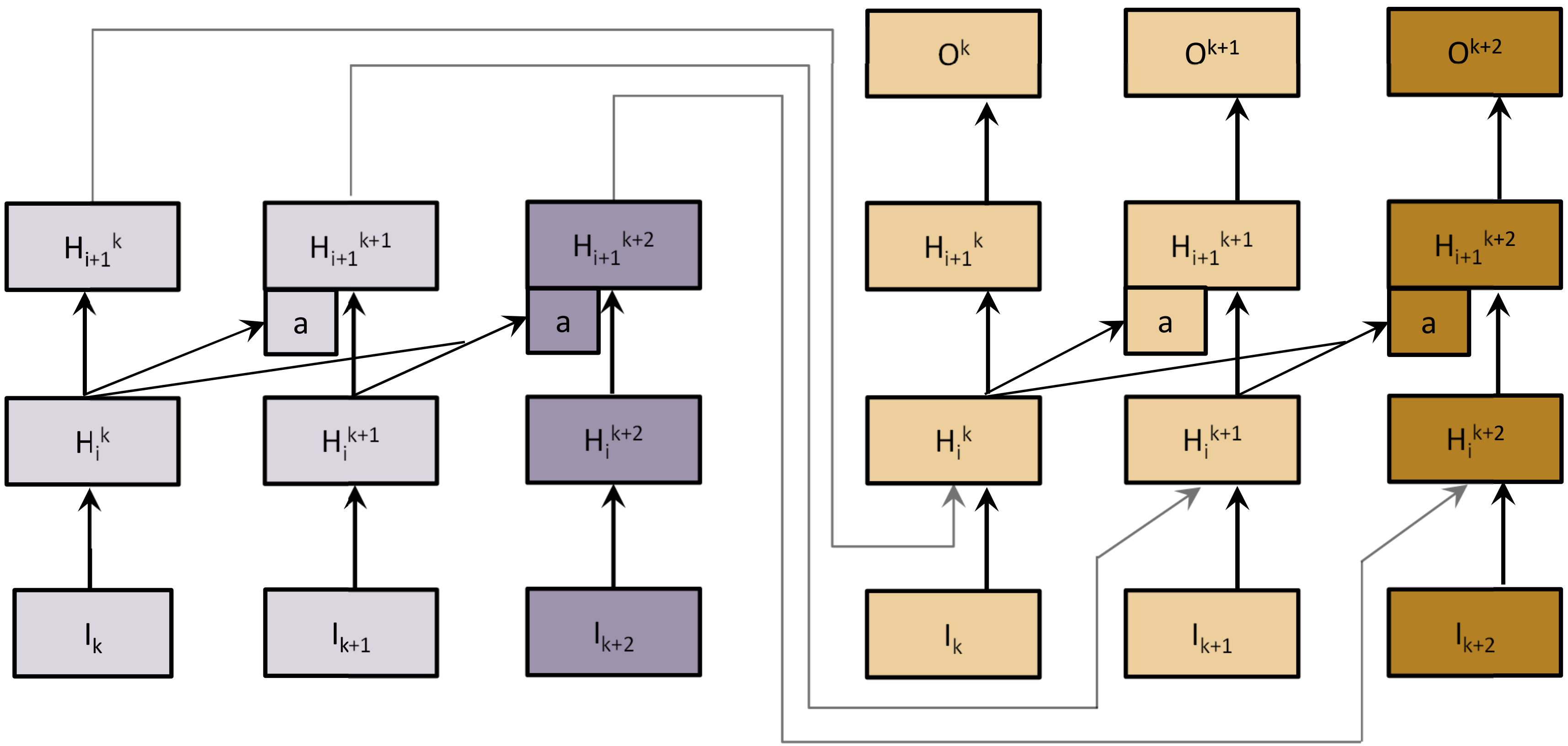

Figure 8: Custom architecture for DM-Progressive sequence-to-sequence learning
Light shaded region: Frozen layers, a: Adapter function, H: LSTM *

* Inspired from [7] [8]

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

sntregiocom

Implementation of the proposed work

Business Logic Integration

- **DST** Tracking of all dialogues for entire interaction
- **Proposed DM model** Prediction of a suitable system action for the perceived intents
- **API handling** checks the API database for the user specifications, checks the availability for single and multiple slots specifications
- **Automated response selection** Each system action has many possible candidates, selection of an apt response by checking availability of relevant slots
- **Integration of API and DST results** Including the random selection of API results and respective slots in the selected candidate response

Technical error handling
To intimate the user about the problem that has occurred in API

There is a **PRICE** price **FOOD** restaurant named **NAME** at **ADDR**. There are **CHOICE** other choices for your specifications. Would you like to choose another one?

There is a **moderate** price **European** restaurant named **Zizzi Cambridge** at **18 Westhill road**. There are **5** other choices for your specifications. Would you like to choose another one?

Figure 9: Handling of API results in the Selected candidate response

- Inference Filtering techniques to select an apt candidate response
- Replacement of API results and slots values in the response

Implementation of the proposed work

Architecture Overview of developed Interactive Conversational agents

Layer	Konversa1	Konversa2
Structured response layer	Output texts	Output texts
API Handling	Including API results and slots value in the selected response	Including API results and slots value in the selected response
NLG	Automated language generator for predicted system action	Automated language generator for predicted system action
Decision layer	No assistance has provided with any set of rules The system has learnt by itself from the trained data — Trained DM model	If system action is apt (Y / N) — Rules for an apt action
Core layer	Dialogue manager for policy learning to predict system actions	Dialogue manager for policy learning to predict system actions
NLU	Entity extraction — Intent classification	Entity extraction — Intent classification
Input layer	Input texts	Input texts

Figure 10: Architecture overview of Automatized (Konversa1) and semi-automatized (Konversa2) Interactive conversational Agents

Results and discussion on live-user evaluation

Analysis on evaluation techniques

- **Usability test** complies ISO standards
- **Efficiency Measurement** robustness, learnability, flexibility, Questionnaire for User Interface Satisfaction (QUIS 7.0) [9]
- **Quality Measures** Overall Reaction, Effectiveness in the task, System Capabilities, Learnability, Visuals and Displays, Assistance and Feedback

Attraktdiff Test

- **Hedonic quality** stimulation, Identification Creativity, Innovativeness, Novelty Connectivity, Professionality, presentability
- **Pragmatic quality** Interaction behavior Complexity, predictability
- **Attractiveness** pleasantness, appeal

Objective assessment
Cronbach's alpha, consensus Measure [11]

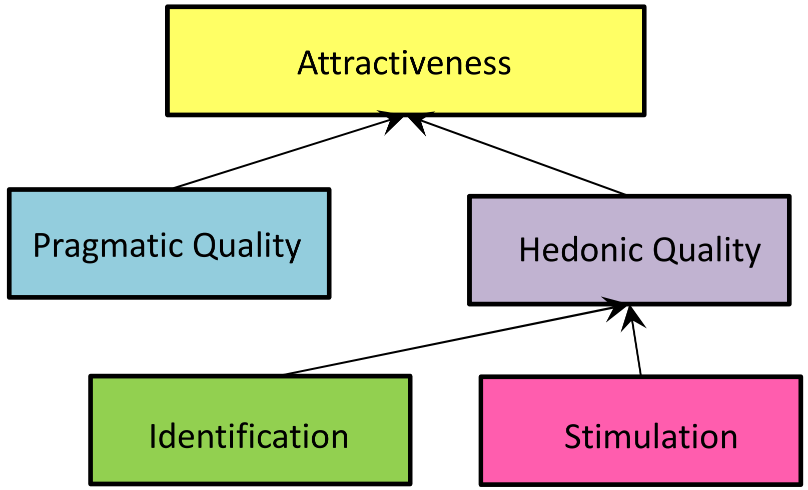

Figure 11: An overview of the Attraktdiff evaluation Model to measure attractiveness [10]

Results and discussion on live-user evaluation

Dialogue statistics and Attraktdiff Live-user evaluation results

Predicted label

True Label	BS	GG	IB	OS	RI	RR
BS	1379	0	0	0	0	0
GG	0	21	0	0	0	0
IB	0	0	248	3	0	2
OS	0	0	0	457	36	11
RI	0	0	0	21	2577	7
RR	0	0	0	13	7	978

Figure 12: Confusion Matrix for perceived intents

Setup and survey procedure

- **Total Participants** 7 professional (P), 7 Naive (N)
- **Konversa1** 6P:7N, **Konversa2** 7P:6N
- **Male-Female ratio** 4:2 in P, 6:1 in N
- Design techniques introduction and clarification

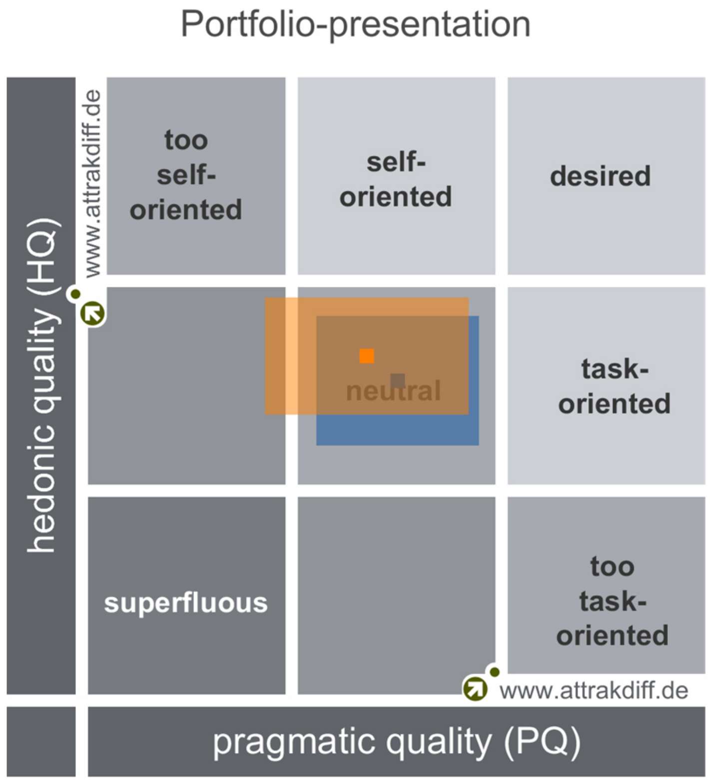

Figure 13: Portfolio of subjective assessment of two products using Attraktdiff test *

* http://www.attrakdiff.de/

Results and discussion on live-user evaluation

Live-user evaluation results- Usability Test

Table 4: Results based on average scale
OR-Overall Reaction, **TE**-Task Effectiveness,
SC-System Capabilities, **LA**-Learnability,
VD-Visuals and Display, **AF**-Assistance and Feedback

	Konversa1	Konversa2
Positive	LA, VD, AF	All Quality measures
Neutral	TE	Null
Negative	OR, SC	Null

Training Details of custom DM strategy

- **Epochs** 100
- **Batch Size** 240
- **Optimizer** ADADELTA

Table 5: Task success rate based on the subjective assessment using Consensus measure

Product	Success rate
Konversa1	0.63
Konversa2	0.67

Table 6: Assessment based on Cronbach's alpha *

Cronbach's alpha	Satisfactory	Good and excellent	Not satisfactory
Konversa1	-	OR, TE, SC,LA, AF	VD
Konversa2	LA	OR,TE, SC,AF	VD

*https://www.wessa.net/rwasp_cronbach.wasp

Conclusion and Future work

solved challenges and future work

- Misprediction in the system action
- Responses lack integrity

- **Service Reliability** 24/7 service
- **Co-operative principles**
  Quality, Quantity, Relation, Manner
- **Responsiveness**
- **Achievement of all objective aspects**

- Handles typo errors
- Flexibility in automated responses
- API-Database free corpus data
- Multi-domain adaptation

Design Complexities

- **Dataset** Improper grammatical structures
- **Decision-Making** Each user intent specific query has multiple possible system action, API handling for user expectations
- **Automated response** Availability of relevant slot information in the selected response

Future Work

- **Dataset Modifications** Grammatically structured utterances for a natural response
- **Suitable attention mechanism** attention to previous system action, present input for a coherent output

References

[1] Susanne Biundo and AndreasWendemuth: *An Introduction to Companion-Technology*, Cognitive Technologies, Springer International Publishing AG, pp. 1-15, 2017, DOI 10.1007/978-3-319-43665-4 1

[2] Pawe l Budzianowski, Tsung-HsienWen, Bo-Hsiang Tseng, Inigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gasic: *MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling*, Association for Computational Linguistics, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: 5016–5026, 2018, https://www.aclweb.org/anthology/D18-1547.pdf

[3] Ledell Wu, Adam Fisch, Sumit Chopra, Keith Adams, Antoine Bordes, and Jason Weston: *StarSpace: Embed Al l The Things!*, Thirty-Second AAAI Conference on Artificial Intelligence, Facebook AI Research, 2017, https://arxiv.org/pdf/1709.03856.pdf

[4] Vladimir Vlasov, Akela Drissner-Schmid, Alan Nichol: *Few-Shot Generalization Across Dialogue Tasks*, Rasa Opensource, 32nd Conference on Neural Information Processing Systems, abs/1811.11707, Montreal, Canada, 2018

References

[5] Sungjin Lee, Qi Zhu, Ryuichi Takanobu, Xiang Li, Yaoqin Zhang, Zheng Zhang, Jinchao Li, Baolin Peng, Xiujun Li, Minlie Huang, and Jianfeng Gao *ConvLab: Multi-Domain End-to-End Dialog System Platform*, Microsoft Research, Tsinghua University, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 1:64-69, 2019, https://arxiv.org/pdf/1904.08637.pdf

[6] Monika Schak, and Alexander Gepperth: *A Study on Catastrophic Forgetting in Deep LSTM Networks*, Part-II Proceedings in 28th International Conference on Artificial Neural Networks , University of Applied Sciences Fulda, Germany, September 17–19, 2019

[7] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le: *Sequence to Sequence Learning with Neural Networks*, Proceedings of the 27th International Conference on Neural Information Processing Systems - 2:3104–3112, 2014, https://arxiv.org/pdf/1409.3215.pdf

[8] Andrei A. Rusu, Neil C. Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, Raia Hadsell: *Progressive Neural Networks*, Google DeepMind, London, UK, 2016, https://arxiv.org/pdf/1606.04671.pdf

References

[9] Andrei Malchanau, Volha Petukhova and Harry Bunt: *Multimodal dialogue system evaluation: a case study applying usability standards*, 9th International Workshop on Spoken Dialogue System Technology, 579:145-159, ISBN: 978-981-13-9442-3, DOI: 10.1007/978-981-13-9443-0 13, 2019

[10] Martin Schrepp, Theo Held, Bettina Laugwitz: *The influence of hedonic quality on the attractiveness of user interfaces of business management software*, Interacting with Computers, pp. 1055-1069, 2006

[11] William J. Tastle a, Mark J. Wierman: *Consensus and dissention: A measure of ordinal dispersion*, International Journal of Approximate Reasoning, 45:531-545, 2007

Questions?

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

snt regiocom

Thank you for the Attention!

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

snt regiocom