HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Towards Ontology-Independent Dialogue State Tracking

Michael Heck

Dialog Systems and Machine Learning

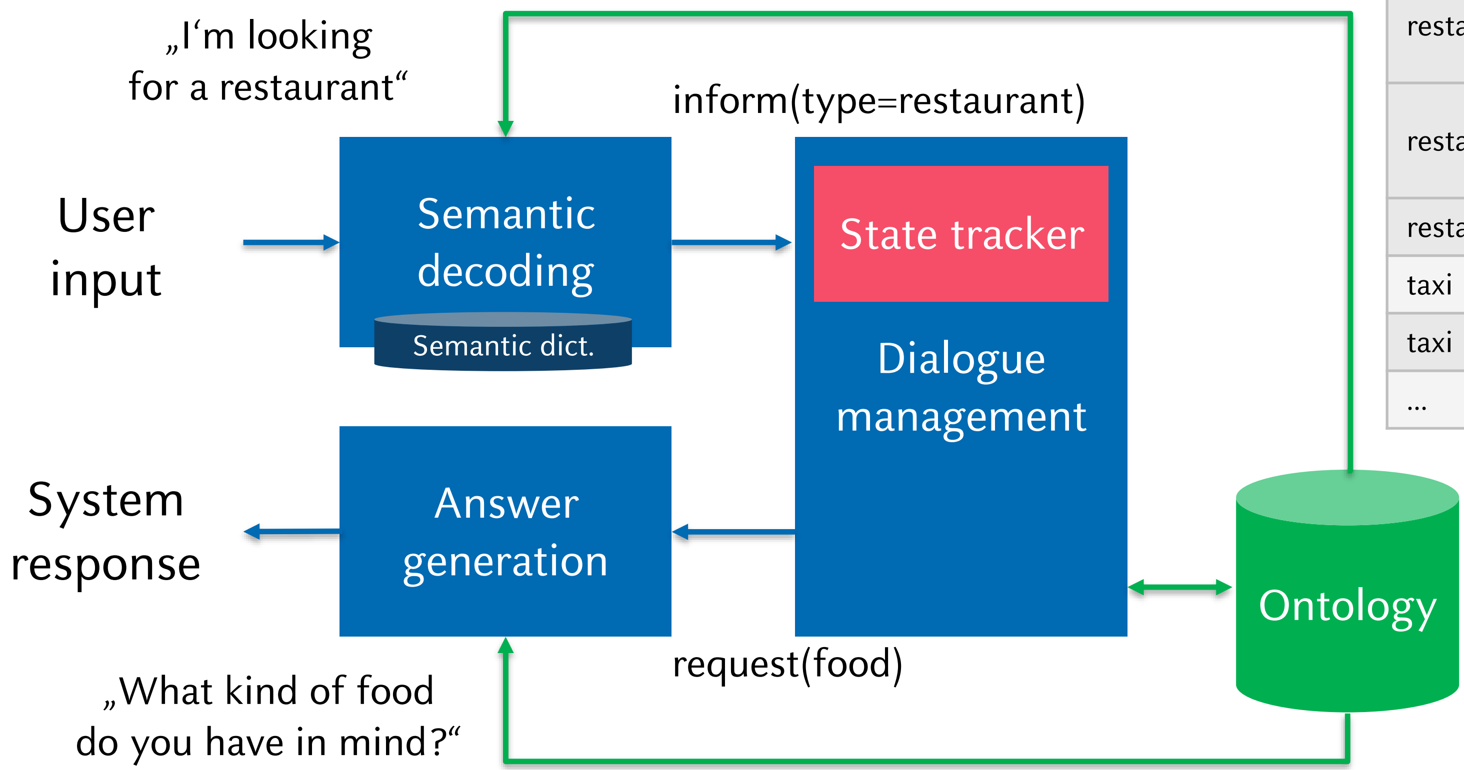
Düsseldorf, Germany,
24th April 2020

INTRODUCTION

Introduction

- Task-oriented dialogue systems (DS) – virtual assistants – gained increased popularity and **acceptance** over the years
  - Accomplish tasks such as bookings, searches, management, …
- DS need to support a wide variety of domains
  - Recent work focused on scalable multi-domain DS
- Data-driven deep learning based approaches improved system quality considerably
  - Shift from discrete to **continuous representations** of concepts

Statistical dialogue systems 101



Domains	Slots	Values
restaurant	food	Italian
		Chinese
restaurant	area	centre
		north
restaurant	price	cheap
taxi	depart	station
taxi	arrive	hotel
…	…	…

Dialogue state tracking

- Dialogue state: Summary of the conversation till current turn
  - Set of constraints, for example **slot-value pairs**
- Dialogue state tracking: Update dialogue state at each turn
  - Required to determine next system action

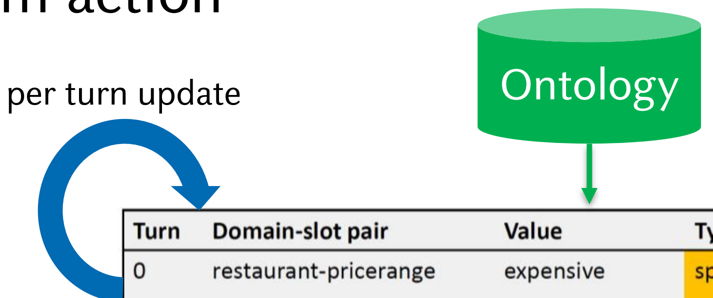
U: i'm looking for an **expensive** restaurant in the **center** of town.

S: there are 33 restaurants [...]. would you like to narrow your search by type of food?
U: i **do not care** about food. surprise me.

S: **fitzbillies restaurant** serves british food, [...]. would you like to hear about any others? [...].
U: that sounds fine. can you book it please and get me the reference number?

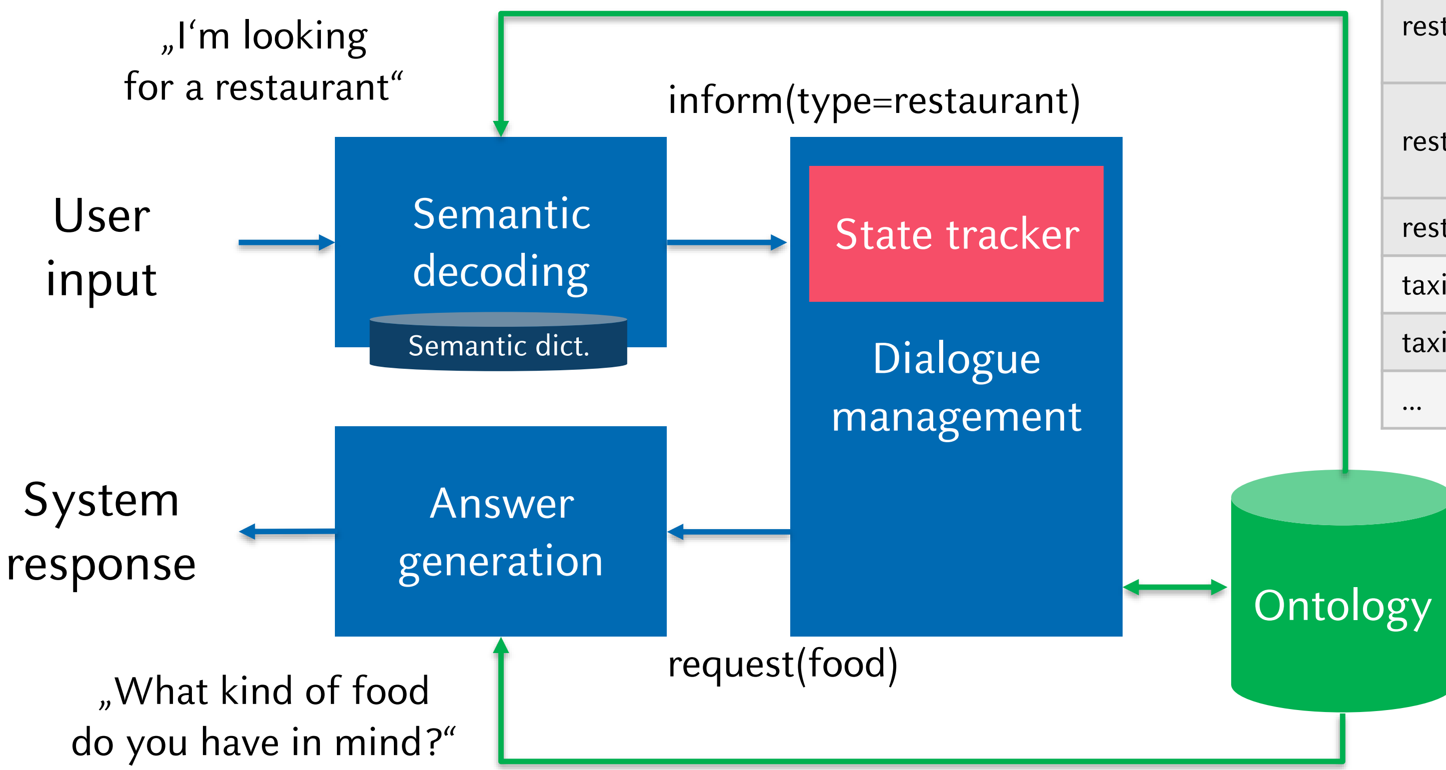
S: sure , what day and time would you like and how many people ?
U: i would like a table for **5** at **11:30** on **tuesday** [...]

S: okay, the booking was successful. [...]. is there anything else i can help you with?
U: i'm also looking for a place to stay. it needs [...] **free wifi** and [be] in the same area as the restaurant.



Turn	Domain-slot pair	Value	Type
0	restaurant-pricerange	expensive	span
0	restaurant-area	center	span
1	restaurant-food	*<dontcare>*	*(dontcare)*
2	restaurant-name	fitzbillies	informed
3	restaurant-people	5	span
3	restaurant-book_time	11:30	span
3	restaurant-book_day	tuesday	span
4	restaurant-internet	*<true>*	*(bool)*
5	hotel-area	center	coreference

Statistical dialogue systems 101

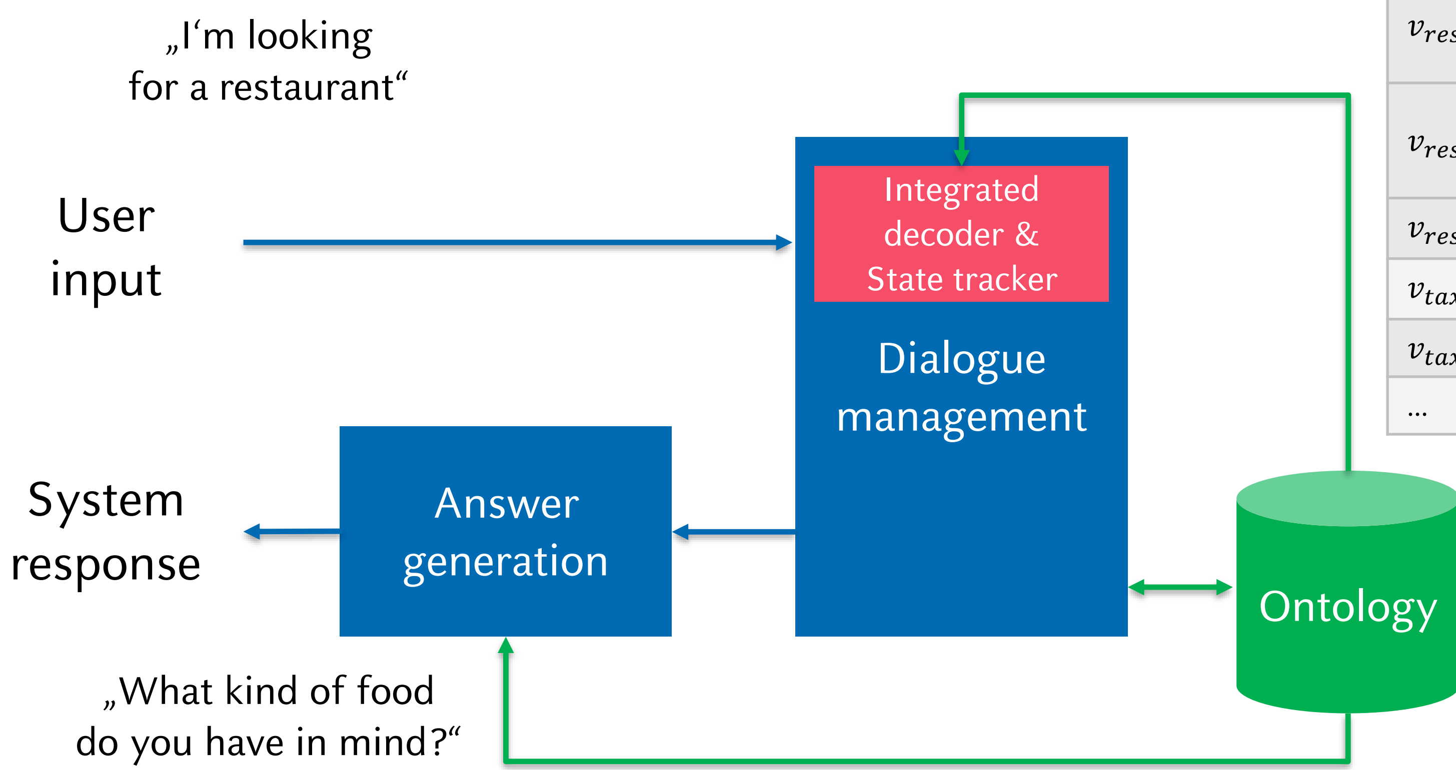


Domains	Slots	Values
restaurant	food	Italian
		Chinese
restaurant	area	centre
		north
restaurant	price	cheap
taxi	depart	station
taxi	arrive	hotel
…	…	…

- Discrete representation of concepts limits capacities

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Continuous representations in DS



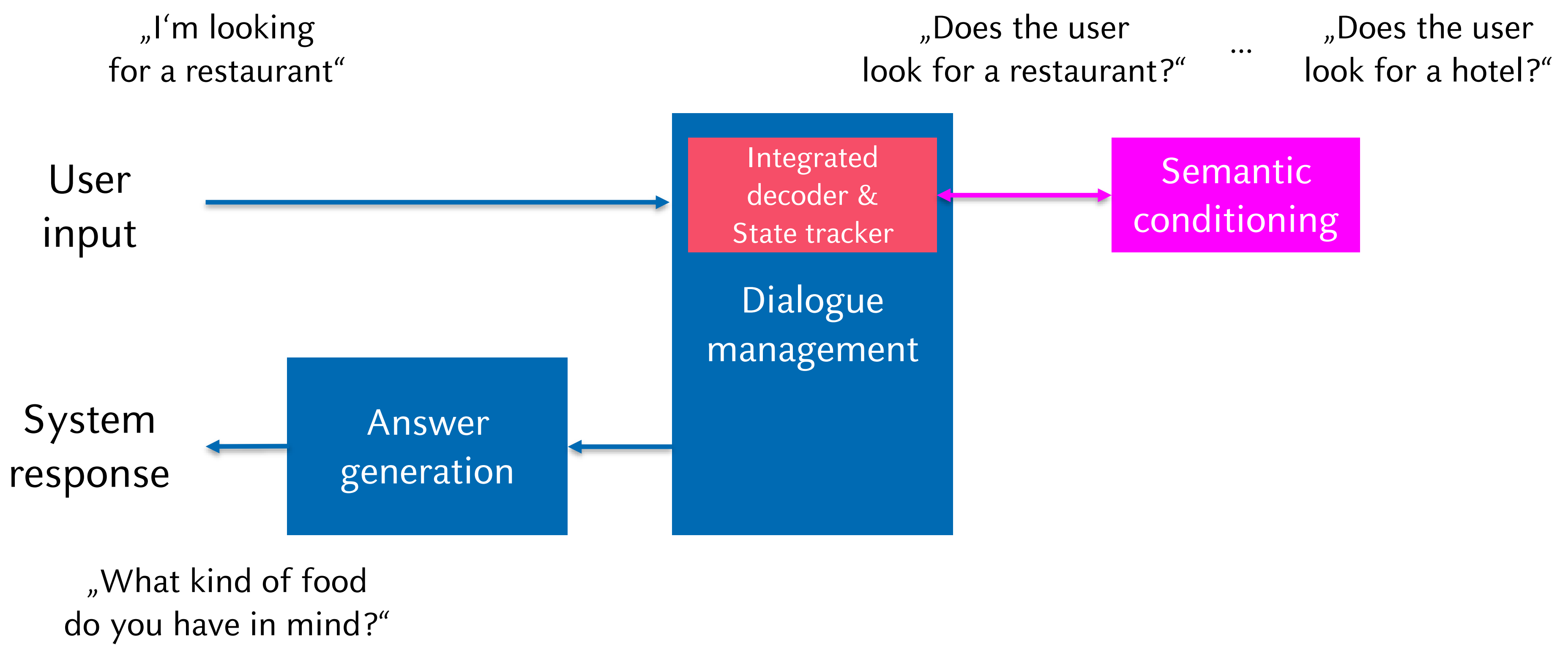
Domains	Slots	Values
$v_{restaurant}$	v_{food}	$v_{Italian}$
		$v_{Chinese}$
$v_{restaurant}$	v_{area}	v_{centre}
		v_{north}
$v_{restaurant}$	v_{price}	v_{cheap}
v_{taxi}	v_{depart}	$v_{station}$
v_{taxi}	v_{arrive}	v_{hotel}
…	…	…

- Vector representations mitigate semantic decoding problem
  - Similarity measures replace exact matching

Mrksic et al., 2017, Neural Belief Tracker - Data-Driven Dialogue State Tracking

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Ontology-independent DS



- Conditioning with natural language replaces fixed ontology
  - Measure semantic similarity between input and concepts

DEEP LEARNING BASED DST

Deep learning based DST

- Achieves state-of-the-art performance in DST evaluations
  - Utilization of **semantic representations** is driving force
    - Leverages semantic similarity of concepts (slots, values, etc.)
    - Representation of previously unseen concepts is possible
    - Tighter integration of DS components

- Picklist based
  - DS as distribution over all possible slot-values
  - Individual scoring of all slot-value pairs

Henderson et al., 2014, Word-based dialog state tracking with recurrent neural networks
Wen et al., 2017, A network-based end-to-end trainable task-oriented dialogue system
Mrksic et al., 2017, Neural Belief Tracker - Data-Driven Dialogue State Tracking
Ramadan et al., 2018, Large-Scale Multi-Domain Belief Tracking with Knowledge Sharing

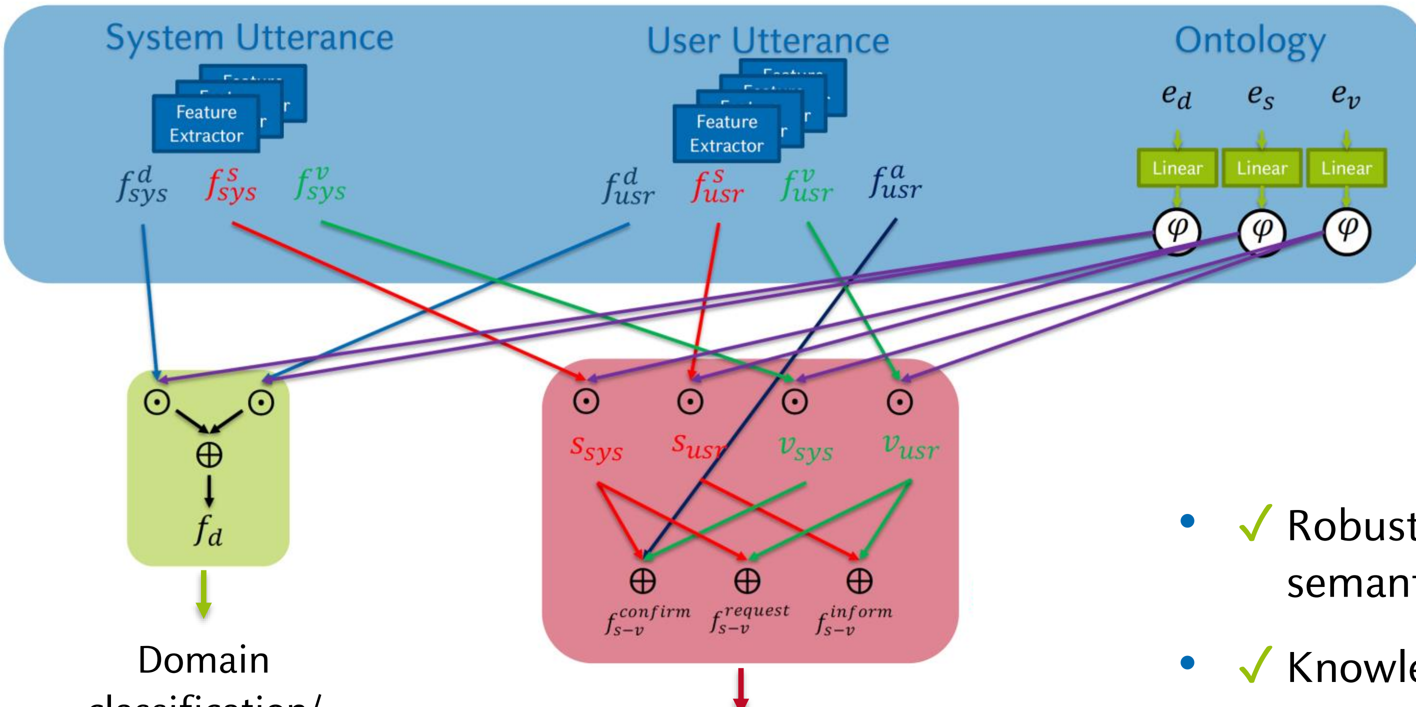
Deep learning based DST

Picklist-based DST

Domains	Slots	Values
restaurant	food	Italian
		Thai
restaurant	area	centre
		north
restaurant	price	cheap
taxi	depart	station
taxi	arrive	hotel
…	…	…

What food would you like?

I'd like Thai food



Domain classification/ tracking

Slot-value classification/ tracking

Input encoders produce vector representations

- ✓ Robustness due to semantic representations
- ✓ Knowledge sharing across domains
- ✗ Limited scalability

Ramadan et al., 2018, Large-Scale Multi-Domain Belief Tracking with Knowledge Sharing

Deep learning based DST

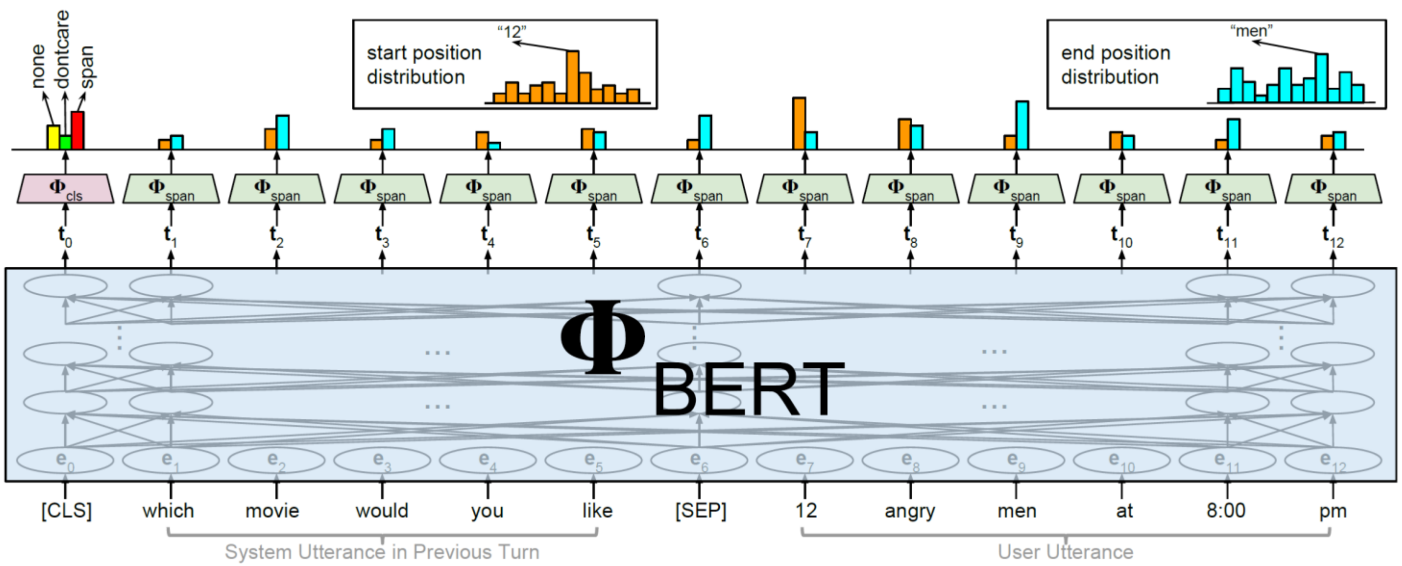
- Picklist based
  - DS as distribution over all possible slot-values
  - Individual scoring of all slot-value pairs

- Span based
  - Find values through span matching in dialogue context

Gao et al., 2019, Dialog state tracking: A neural reading comprehension approach
Chao and Lane, 2019, BERT-DST: Scalable end-to-end dialogue state tracking with bidirectional encoder representations from transformer
Kim et al., 2019, Efficient dialogue state tracking by selectively overwriting memory

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Span-based DST



- Transformer produces contextual representations of input
  - Sentence representation used to determine presence of value
  - Token representations used to determine value span
- ✘ Limited to extractive values

Chao and Lane, 2019, BERT-DST: Scalable end-to-end dialogue state tracking with bidirectional encoder representations from transformer

Deep learning based DST

- Picklist based
  - DS as distribution over all possible slot-values
  - Individual scoring of all slot-value pairs

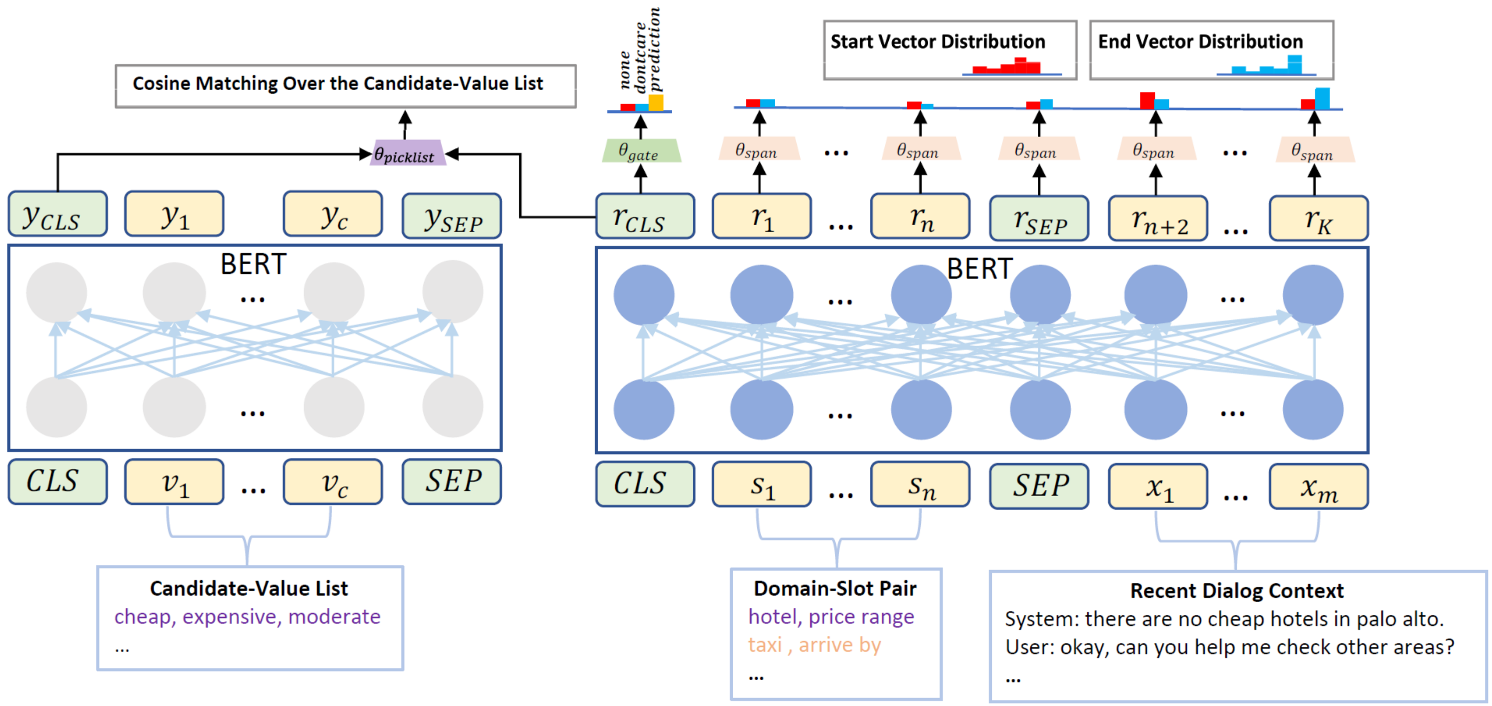
- Span based
  - Find values through span matching in dialogue context

- Hybrid
  - Combine picklists with span prediction

Zhang et al., 2019, Find or classify? dual strategy for slot-value predictions on multi-domain dialog state tracking

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Hybrid approaches



- Similarity matching with candidates in picklist, or span pred.
- Slot name (and domain name) as part of input

Zhang et al., 2019, Find or classify? dual strategy for slot-value predictions on multi-domain dialog state tracking

SCHEMA-GUIDED PARADIGM

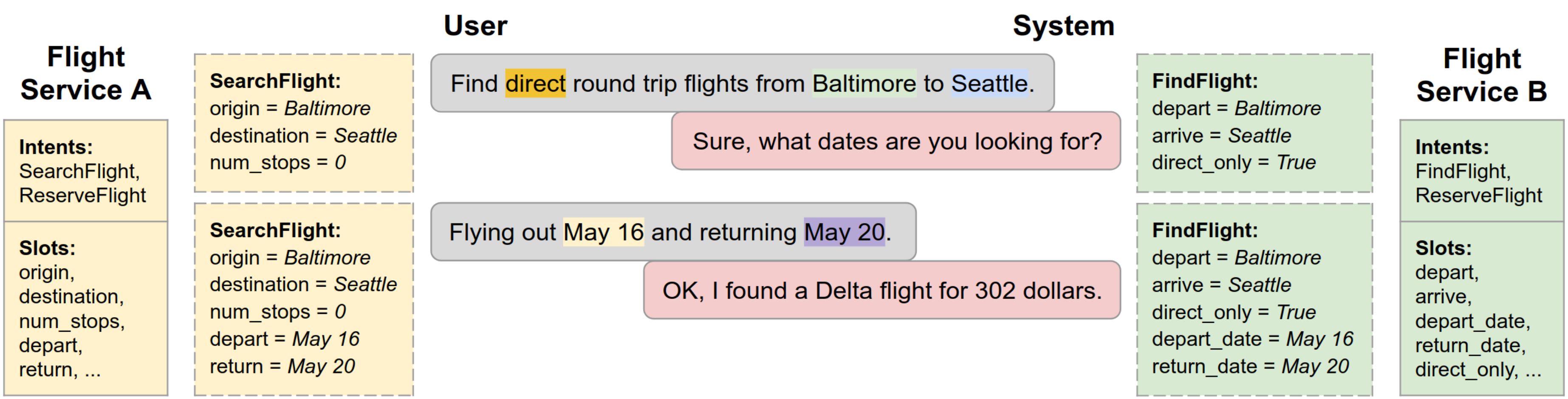
Schema-guided paradigm

Reality check

- Current evaluations don‘t fully capture reality of scenarios
  - Few domains, one service per domain, static ontologies

vs.

- Many domains, many services (defined by APIs)
  - Mismatch of training and testing conditions



Schema-guided paradigm

Shortcomings of recent systems

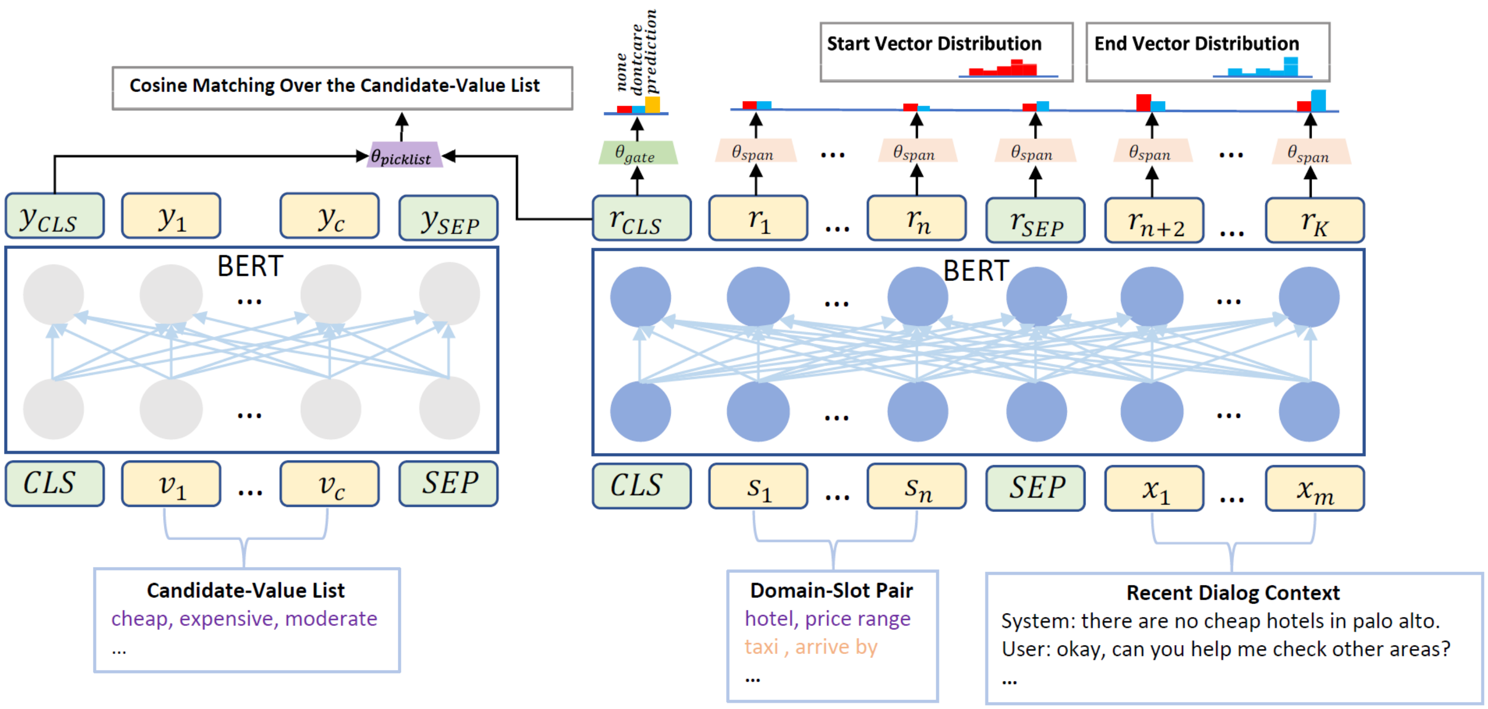
- Recent systems parse dialogues in terms of **fixed concepts**
  - Lack understanding of the **semantics** of concepts
- Example: “I want to buy tickets for a movie.”
  - Models predict “BuyMovieTickets” based on observed patterns
    - No association with real action of buying movie tickets
    - Similarity to action of buying concert tickets not captured
- Models not robust to changes
  - Need to be retrained as new slots or intents are added
  - Domain-specific parameters unsuitable for zero-shot application

Challenges of building large-scale systems

- Support of heterogenous services/APIs
  - Might overlap in functionality
- Robustness towards changes in API
  - Robustness towards new slots and intents
  - Generalization to new slot values (with little or no retraining)
- Generalization to new APIs
  - Joint modelling across APIs
  - Zero-shot generalization

Approaches to related problems

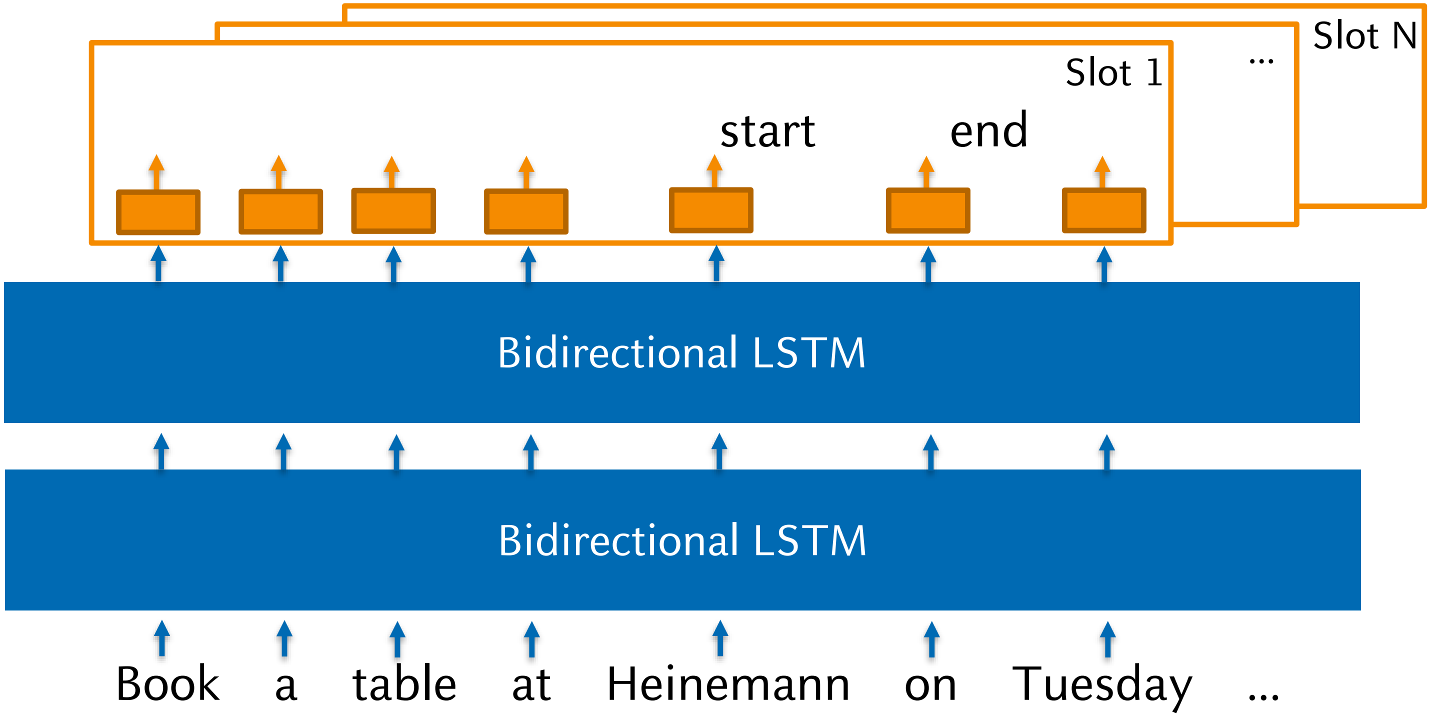
- Adaptation and transfer learning for Slot-filling for DST
  - Parameter sharing for domain adaptation and joint training



Zhang et al., 2019, Find or classify? dual strategy for slot-value predictions on multi-domain dialog state tracking

Approaches to related problems

- Zero-shot learning for Slot-filling for DST
  - Infusing semantic slot representations into unified model

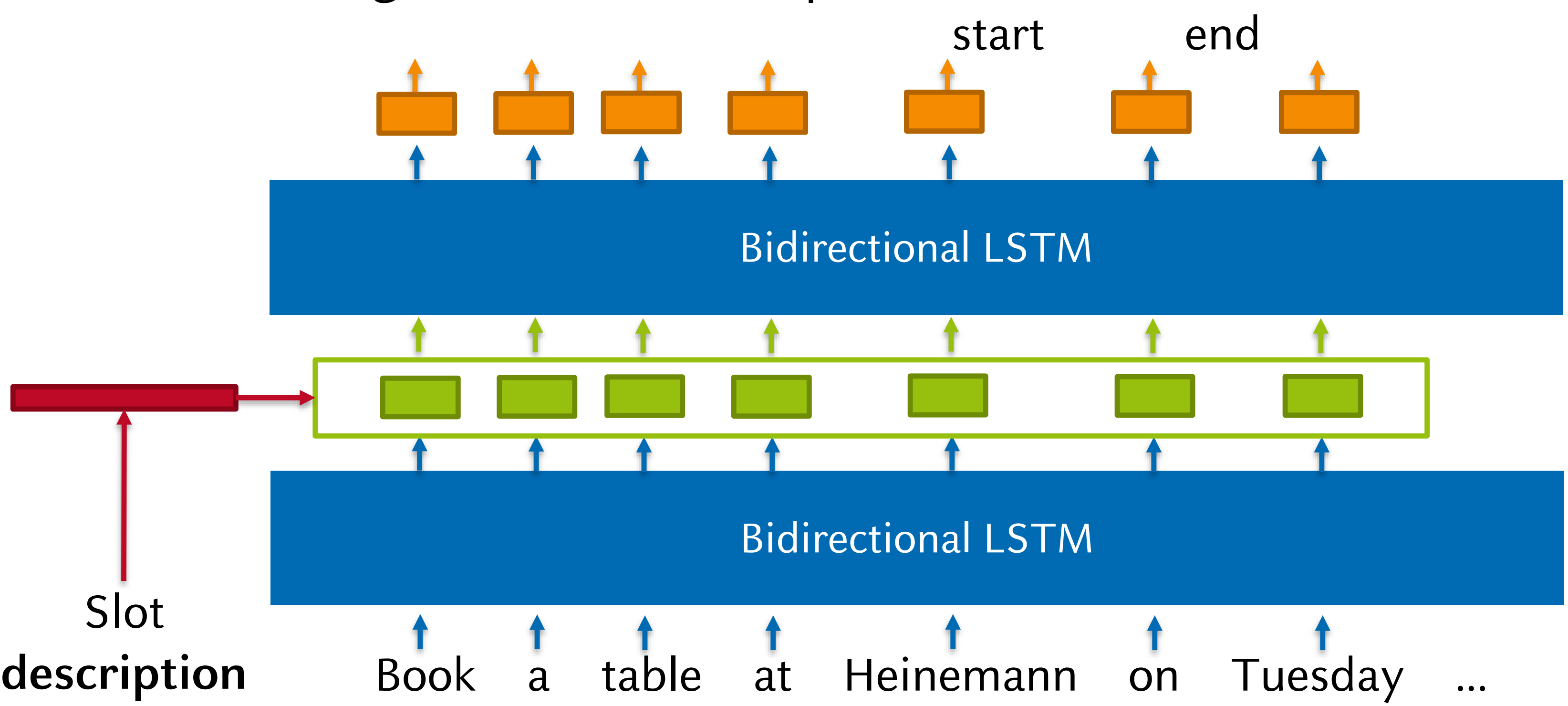


Bapna et al., 2017, Towards Zero-Shot Frame Semantic Parsing for Domain Scaling
Shah et al., 2019, Robust Zero-Shot Cross-Domain Slot Filling with Example Values

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Approaches to related problems

- Zero-shot learning for Slot-filling for DST
  - Infusing semantic slot representations into unified model



Bapna et al., 2017, Towards Zero-Shot Frame Semantic Parsing for Domain Scaling
Shah et al., 2019, Robust Zero-Shot Cross-Domain Slot Filling with Example Values

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Schema-guided paradigm for dialogue modeling

- Each **service** provides a **schema**
  - Lists supported slots and intents
  - Provides natural language descriptions for schema elements

Service
service_name: *"Payment"*
description: *"Digital wallet to make and request payments"*

Slots
name: *"account_type"* categorical: True
description: *"Source of money to make payment"*
possible_values: [*"in-app balance", "debit card", "bank"*]

name: *"amount"* categorical: False
description: *"Amount of money to transfer or request"*

name: *"contact_name"* categorical: False
description: *"Name of contact for transaction"*

Intents
name: *"MakePayment"*
description: *"Send money to your contact"*
required_slots: [*"amount", "contact_name"*]
optional_slots: [*"account_type" = "in-app balance"*]

name: *"RequestPayment"*
description: *"Request money from a contact"*
required_slots: [*"amount", "contact_name"*]

Figure: Example schema for a service called „payment".

Rastogi et al., 2020, Towards Scalable Multi-Domain Conversational Agents: The Schema-Guided Dialogue Dataset

Schema-guided paradigm for dialogue modeling

- Advocates building a single **unified** dialogue model for all services and APIs using **semantic conditioning**
  - A model should not contain service specific components
  - A service's schema serves as input to the model
    - Uses descriptions to obtain **semantic representations** of schema elements
    - Predictions are **conditioned** on semantics of schema
    - Predictions over dynamic sets of intents and slots
  - A model should generalize to unseen services, APIs, concepts

Schema-guided paradigm for dialogue modeling

- Zero-shot learning by using semantic modeling
- Knowledge sharing by ...
  - ... relating semantically similar concepts
  - ... using single unified model
- Handling of unseen services and API changes by using
  - natural language input
  - semantic representations

  to condition the model

SCHEMA-GUIDED DST

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Schema-guided DST track at DSTC8

- SGD Dataset
  - Benchmark highlighting challenges for large-scale systems

	DSTC2	WOZ2.0	FRAMES	M2M	MultiWOZ	SGD
Domains	1	1	3	2	7	**16**
Slots	8	4	61	13	30	**214**
Values	212	99	3,871	138	4,510	14,139
Dialogues	1,612	600	1,369	1,500	8,438	16,142
Avg. turns per dialogue	14.49	7.45	14.60	9.86	13.46	20.44

Table: Statistics of training portions of datasets

	train	dev	test	TOTAL
Dialogs	16,142	2,482	4201	22825
Domains	16	16	18	20
Services	26	17	21	45
Dialogs w/ unseen APIs	-	42%	70%	-
Turns w/ unseen APIs	-	45%	77%	-

Table: Split of SGD dataset

Domain	Services	Domain	Services
Alarm	1	Movies	3
Banks	2	Music	3
Buses	3	Payment	1
Calendar	1	RentalCars	3
Events	3	Restaurants	2
Flights	4	RideSharing	2
Homes	1	Services	4
Hotels	4	Train	1
Media	3	Travel	1
Messaging	1	Weather	1

Table: Domains and services in SGD dataset

- Slot types:
  - **Non-catecorical**: set of possible values is unrestricted
    - Eval sets contains unseen values
  - **Categorical**: possible values are pre-defined and fixed

Rastogi et al., 2020, Schema-Guided Dialogue State Tracking Task at DSTC8

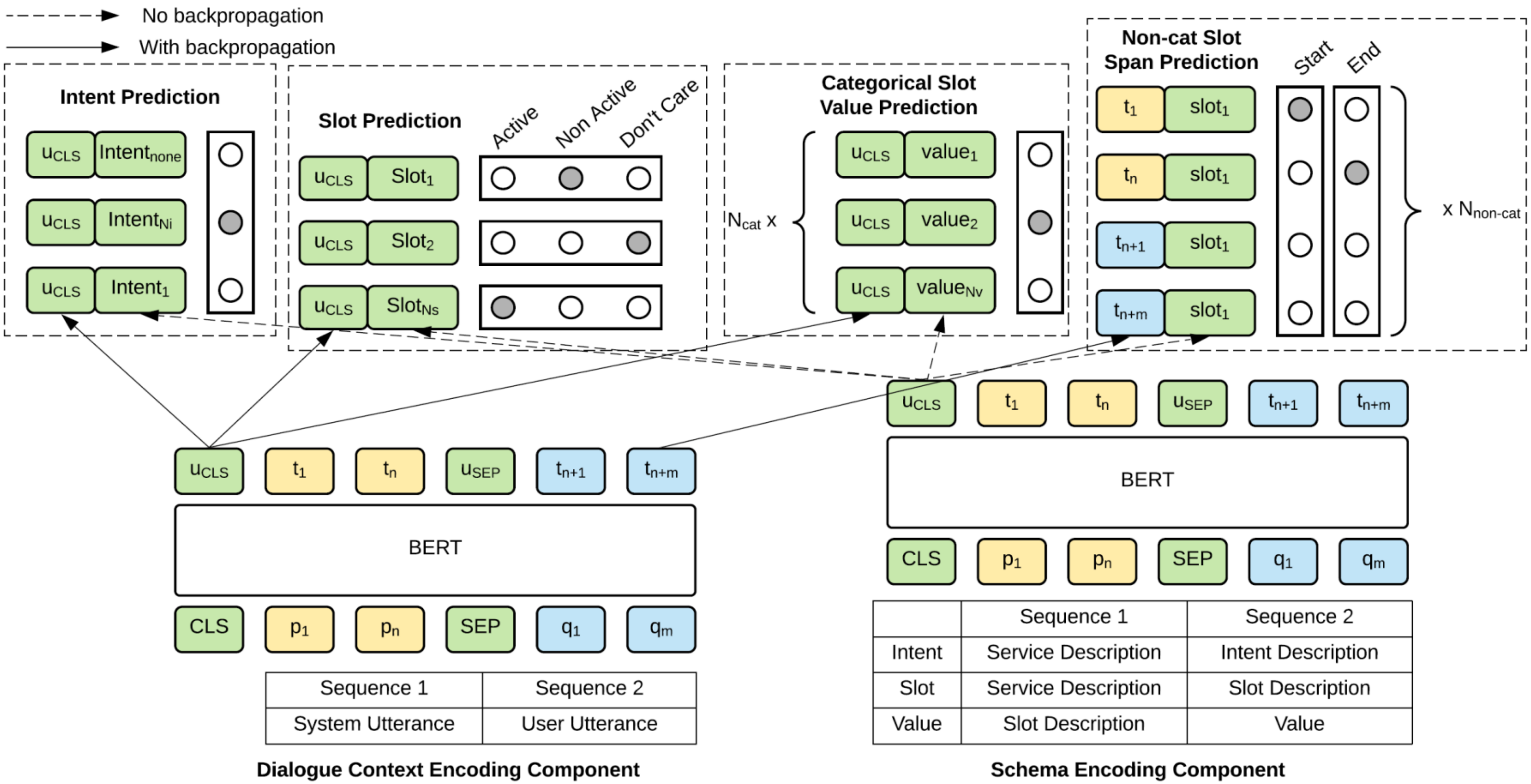
Schema-guided DST

Baseline: Zero-shot dialogue state tracking

- Model is shared among all services and domains
- Uses 2 contextual encoders:
  - Finetuned BERT encodes context
  - Fixed pre-trained BERT encoding schema element descriptions
    - Intents, slots, categorical slot values
- Schema element-wise classification
  - Concat. context representation and schema element represent
  - Do for each turn and for each schema element

Rastogi et al., 2020, Towards Scalable Multi-Domain Conversational Agents: The Schema-Guided Dialogue Dataset

Baseline: Zero-shot dialogue state tracking



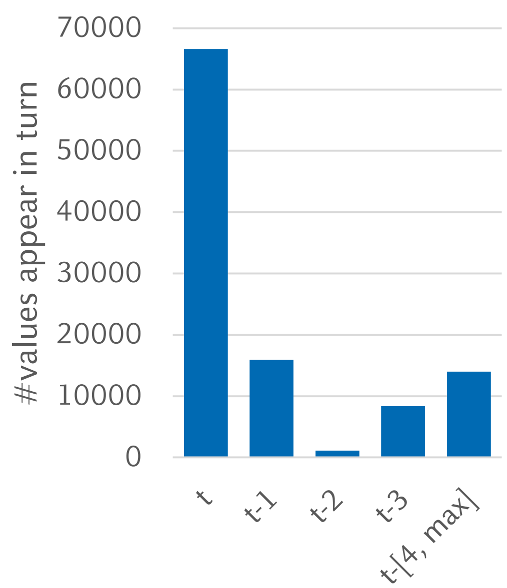
Evaluation metrics

- **Joint goal accuracy**
  - Average accuracy of predicting all slot assignments correctly
- Average goal accuracy
  - Average accuracy of predicting a slot value correctly
- Active intent accuracy
  - Fraction of user turns for which intent was predicted correctly
- Requested slot F1
  - Average F1 score for requested slots

Evaluation results

	All services				Seen services				Unseen services			
	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1
Baseline	0,25	0,56	0,91	0,97	0,41	0,68	0,95	1	0,2	0,52	0,89	0,95

- ✘ Drawbacks
  - No history (only single turn) in context
    - Many slot values appear multiple turns earlier
  - Separate models for context and schema
    - Interaction only after encoding
    - No finetuning of schema encoder



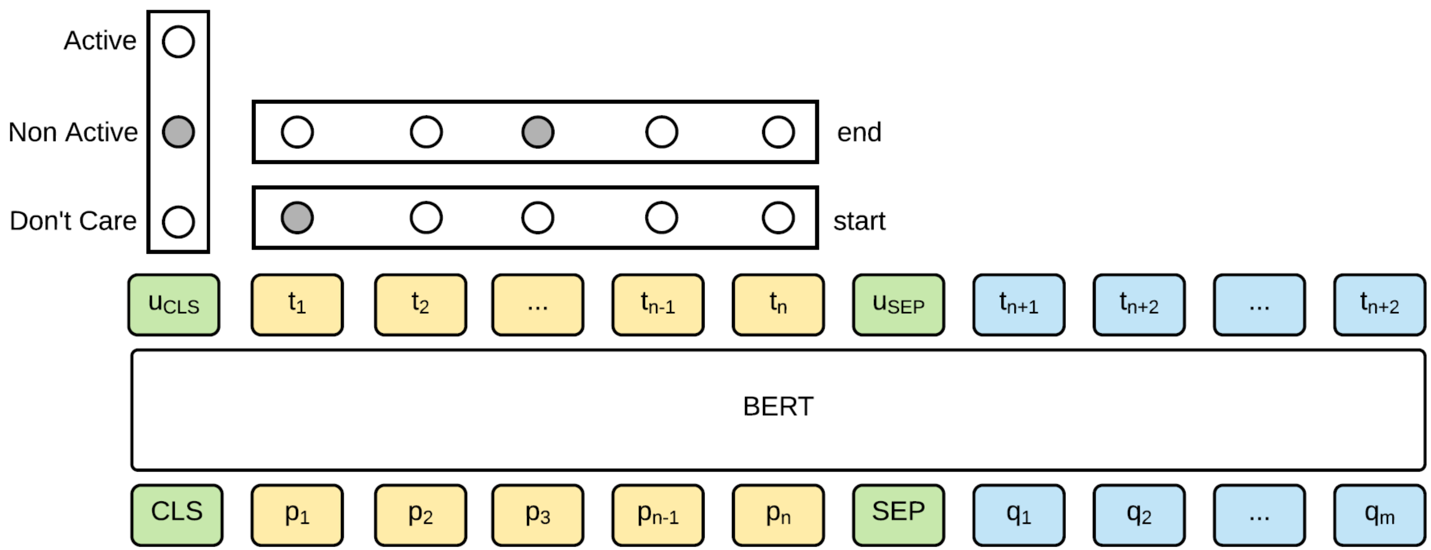
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Unified span detection framework for SG-DST

- **Single** BERT to encode context and schema elements
  - To facilitate more interaction and utilize attention mechanism
  - Multiple passes per turn and slot, one for each prediction task
    - Intent, categorical slot, non-categorical slot
- Adds (truncated) dialogue history to input
- Render all predictions a span prediction problem
  - To utilize same model architecture for multitask learning effect

Shi et al., 2020, A BERT-based Unified Span Detection Framework for Schema-Guided Dialogue State Tracking

Unified span detection framework for SG-DST



	Sequence 1	Sequence 2
Intent	$\text{turn}_{t\text{-}k}$ [SEP1] ... turn_t	[INTENT] Service Name [SEP2] Intent Name Intent Description
Categorical Slot	$\text{turn}_{t\text{-}k}$ [SEP1] ... turn_t [CSEP] value_1 [CSEP] value_2 ... [CSEP] value_n	[C_SLOT] Slot Name [SEP2] Slot Name Slot Description
Non Categorical Slot	$\text{turn}_{t\text{-}k}$ [SEP1] ... turn_t	[NC_SLOT] Slot Name [SEP2] Slot Name Slot Description

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Evaluation results

	All services				Seen services				Unseen services			
	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1
Baseline	0,25	0,56	0,91	0,97	0,41	0,68	0,95	1	0,2	0,52	0,89	0,95
Shi	0,54	0,8	0,91	0,87	0,53	0,75	0,96	0,85	0,55	0,82	0,9	0,88

- Important details
  - Uses BERT-large instead of BERT-base
  - Post-submission tests showed advantage of even longer history
- Observations
  - ✓ Very good generalization to new services
    - Authors attribute this to joint encoding of context and schema
  - ✗ Req. slot F1 significantly lower, reason unclear (not discussed)

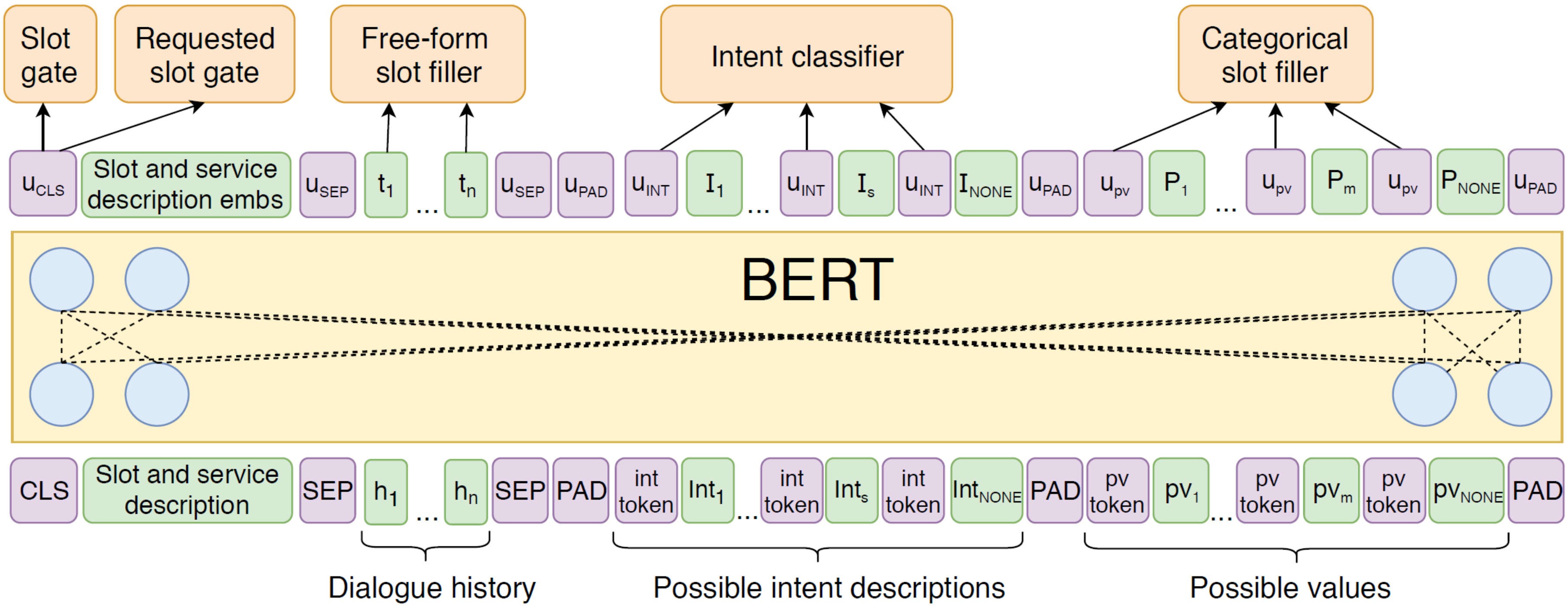
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Goal-oriented multi-task BERT-based DST

- **Single** BERT to encode context and schema elements
  - **Single pass** per turn and slot, all predictions are done at once
    - Intent + Slot (request, categorical, non-categorical)
  - Special classification heads work in parallel
- Adds (truncated) dialogue history to input
- Strict input format
  - Special tokens and padding for partitioning

Gulyaev et al., 2020, Goal-Oriented Multi-Task BERT-Based Dialogue State Tracker

Goal-oriented multi-task BERT-based DST



	Input sequence
Question	Slot and service description
Context	Dialogue history
Possible intents	Descriptions of intents supported by the service
Possible values	Possible slot values (for categorical slots only)

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Evaluation results

	All services				Seen services				Unseen services			
	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1
Baseline	0,25	0,56	0,91	0,97	0,41	0,68	0,95	1	0,2	0,52	0,89	0,95
Gulyaev	0,46	0,75	0,75	0,97	0,53	0,74	0,87	0,97	0,44	0,75	0,71	0,97
Shi	0,54	0,8	0,91	0,87	0,53	0,75	0,96	0,85	0,55	0,82	0,9	0,88

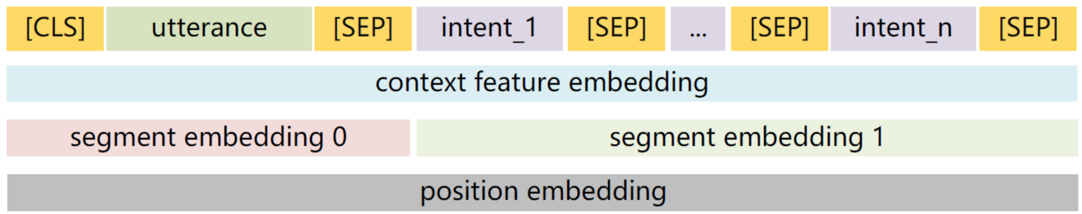
- Uses BERT-large (finetuned on Squad) instead of BERT-base
- Observations
  - ✓ Categorical slots as span prediction task boosts performance
    - Similarly, intent classification as span prediction boosts performance
  - ✘ Similar performance to (Shi), but lacks behind for intent acc.
    - Relies on token representations and span prediction
  - ✘ Struggles with domain switches, slot value transfers

Fine-tuning BERT for schema-guided zero-shot DST

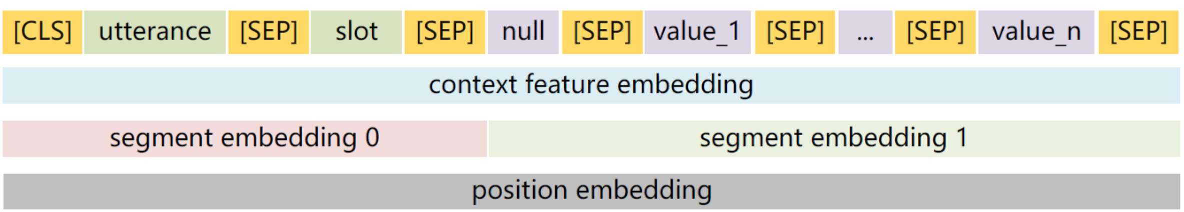
- **6** BERT fine-tuned models for prediction
  - Intent prediction
  - Slot prediction (Categorical, Free-form, Requested)
  - Slot transfer prediction (In-domain, Cross-domain)
  - Multiple passes: First Intent & Slot, then transfer prediction
- Adds (truncated) dialogue history to input
- Adds auxiliary context features to BERT input
  - Indicate if a value/intent was predicted in turn t-1
  - Indicate if a value was mentioned by the system

Ruan et al., 2020, Fine-Tuning BERT for Schema-Guided Zero-Shot Dialogue State Tracking

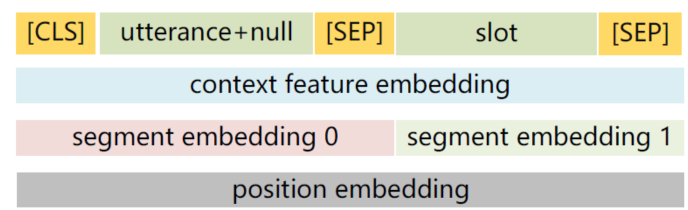
Fine-tuning BERT for schema-guided zero-shot DST



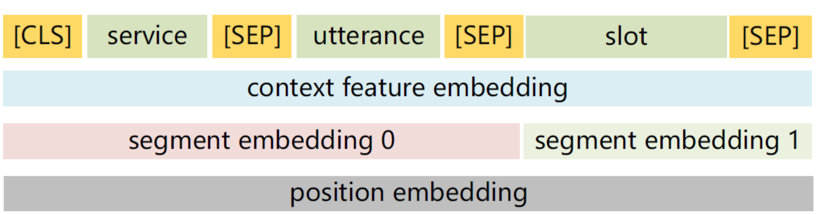
Input for intent prediction model



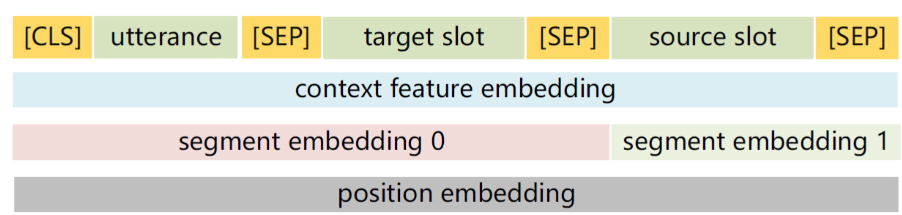
Input for categorical slot prediction model



Input for free-form slot prediction model
Input for requested slot prediction model



Input for in-domain slot transfer model



Input for cross-domain slot transfer model

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Evaluation results

	All services				Seen services				Unseen services			
	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1
Baseline	0,25	0,56	0,91	0,97	0,41	0,68	0,95	1	0,2	0,52	0,89	0,95
Gulyaev	0,46	0,75	0,75	0,97	0,53	0,74	0,87	0,97	0,44	0,75	0,71	0,97
Shi	0,54	0,8	0,91	0,87	0,53	0,75	0,96	0,85	0,55	0,82	0,9	0,88
Ruan	0,74	0,92	0,92	0,99	0,88	0,96	0,96	**1**	**0,69**	0,91	0,91	0,99

- Observations
  - ✓ Slot transfers significantly improve performance
    - In-domain transfers constitute value references across multiple turns
    - Cross-domain transfers rely on reference resolution mechanism
  - ✗ Joint GA drops considerably for unseen services
  - ✓ / ✗ Adding dev set data to training has some positive effect

Schema-guided DST

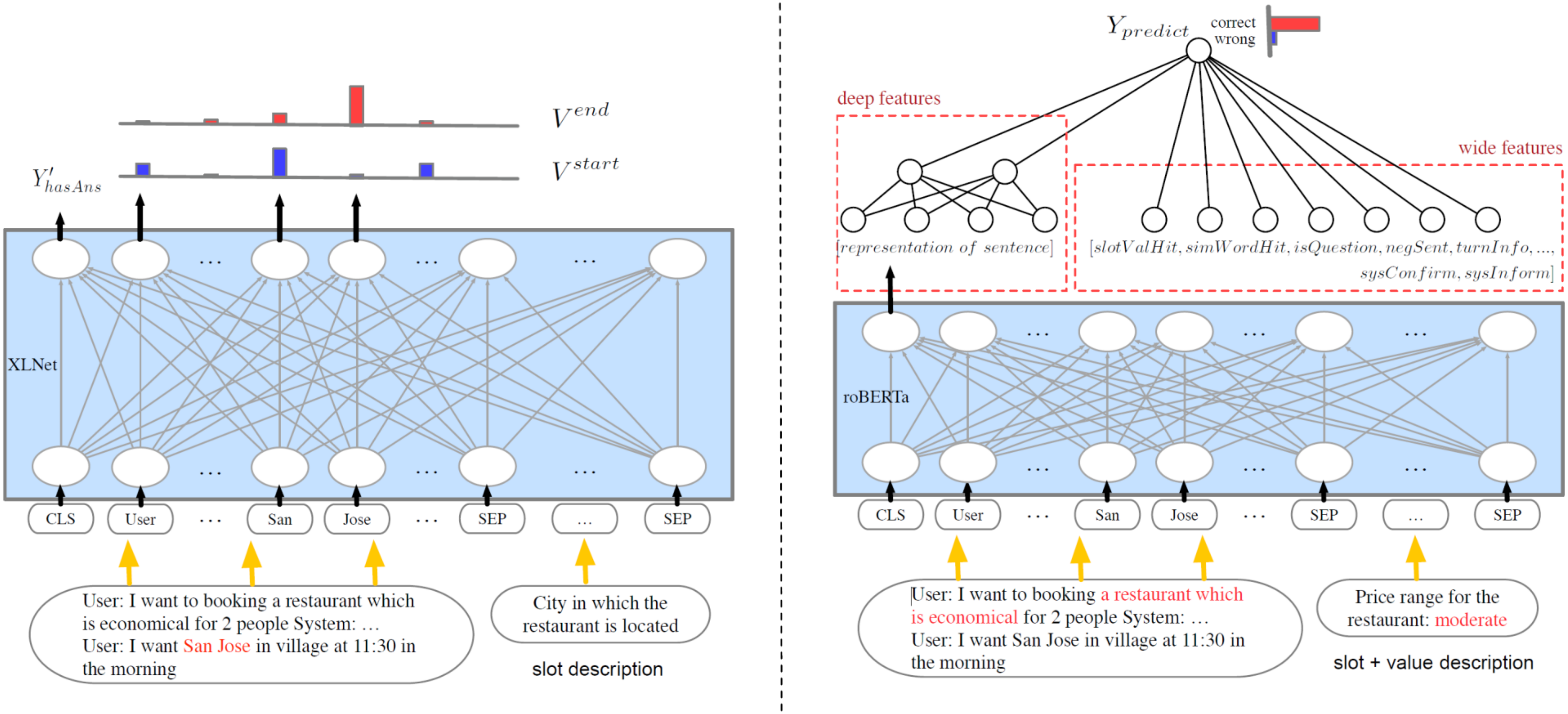
Reading comprehension and wide & deep DST

- Reading comprehension model for non-categorical slots
  - Unrestricted input size
  - Adds entire dialogue history to input
- Wide & deep model for categorical slots
  - Transformer model output + hand-crafted features
- Data augmentation to vary schema element descriptions
  - Automatic generation via back-translations
- Joint model for intent and requested slot prediction
  - Classify dialogue context + intent/slot description

Ma et al., 2020, An End-to-End Dialogue State Tracking System with Machine Reading Comprehension and Wide & Deep Classification

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Reading comprehension and wide & deep DST



a. MRC model for span-based slot and numerical slot

b. Wide & Deep model for boolean and text-based slot

Schema-guided DST

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Evaluation results

	All services				Seen services				Unseen services			
	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1
Baseline	0,25	0,56	0,91	0,97	0,41	0,68	0,95	1	0,2	0,52	0,89	0,95
Gulyaev	0,46	0,75	0,75	0,97	0,53	0,74	0,87	0,97	0,44	0,75	0,71	0,97
Shi	0,54	0,8	0,91	0,87	0,53	0,75	**0,96**	0,85	0,55	0,82	0,9	0,88
Ruan	0,74	0,92	0,92	**0,99**	0,88	0,96	**0,96**	**1**	0,69	0,91	0,91	**0,99**
Ma	**0,87**	**0,97**	**0,95**	0,98	**0,92**	**0,98**	**0,96**	0,99	**0,85**	**0,97**	**0,95**	0,98

- Important details:
  - Hand-crafted features are rule and heuristic based (+10% JGA)
  - Data augmentation by back-translation from Chinese (+6% JGA)
  - Numerical slots are rendered non-categorical
  - Partial delexicalization (phone numbers)
  - Dev set used as additional training data

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Summary & Analysis

	All services				Seen services				Unseen services			
	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1
Baseline	0,25	0,56	0,91	0,97	0,41	0,68	0,95	1	0,2	0,52	0,89	0,95
Gulyaev	0,46	0,75	0,75	0,97	0,53	0,74	0,87	0,97	0,44	0,75	0,71	0,97
Shi	0,54	0,8	0,91	0,87	0,53	0,75	**0,96**	0,85	0,55	0,82	0,9	0,88
Ruan	0,74	0,92	0,92	**0,99**	0,88	0,96	**0,96**	**1**	0,69	0,91	0,91	**0,99**
Ma	**0,87**	**0,97**	**0,95**	0,98	**0,92**	**0,98**	**0,96**	0,99	**0,85**	**0,97**	**0,95**	0,98

- What worked?
  - Approach: Reading comprehension + classification
    - Few submissions use a Baseline-style approach using similarity scoring
  - Most systems exploit synergy effects from multitasking
  - Maximizing context
    - Slot value reference resolution necessary across multiple turns
  - Using hand-crafted features and additional data

Summary & Analysis

	All services				Seen services				Unseen services			
	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1	Joint GA	Avg GA	Intent Acc	Req Slot F1
Baseline	0,25	**0,56**	0,91	0,97	0,41	0,68	0,95	1	0,2	0,52	0,89	0,95
Gulyaev	0,46	**0,75**	**0,75**	0,97	0,53	0,74	0,87	0,97	0,44	**0,75**	**0,71**	0,97
Shi	0,54	**0,8**	0,91	**0,87**	0,53	0,75	**0,96**	0,85	**0,55**	**0,82**	0,9	**0,88**
Ruan	0,74	0,92	0,92	**0,99**	0,88	0,96	**0,96**	**1**	**0,69**	0,91	0,91	**0,99**
Ma	**0,87**	**0,97**	**0,95**	0,98	**0,92**	**0,98**	**0,96**	0,99	**0,85**	**0,97**	**0,95**	0,98

- What worked maybe?
  - Specialized tags, input formatting, input processing
    - Benefits not investigated enough
  - Compartmentalizing: Specialized models for sub-tasks
    - Best systems employ multiple specialized encoders
    - Unified models are among most robust

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

DISCUSSION & CONCLUSION

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Mission accomplished?

- Multiple specialized models vs. unified models
  - What is the best use of semantic encoding?
    - Specialized representations for subtasks vs. generalized representations
    - Impact on architectures‘ generalization capacities? Trade-off observable
- Engineering, heuristics, augmentation
  - Potence of auxiliary features demonstrates insufficiencies in semantic encoding. How to overcome limitations of encoders?
- Role of similarity measures
  - No exploration of spaces of contextual representations
  - Post-encoding similarity scoring not sufficiently explored

Conclusion

- Semantic conditioning of complex models is promising
  - Huge performance gain within single challenge iteration:
    25% Joint GA -> **87%** Joint GA!
  - Seemingly a convergence towards a „universal“ approach
- What next?
  - Zero-shot performance still not satisfactory
    - Reliance on tweaks to minimize gap
  - What if information about active service is not provided?
  - What if user does out-of-service requests?
    - DSTC9: Incorporating external non-dialogue knowledge sources

Thank you!

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Select references

- Mrksic et al., 2017, Neural Belief Tracker - Data-Driven Dialogue State Tracking
- Henderson et al., 2014, Word-based dialog state tracking with recurrent neural networks
- Wen et al., 2017, A network-based end-to-end trainable task-oriented dialogue system
- Ramadan et al., 2018, Large-Scale Multi-Domain Belief Tracking with Knowledge Sharing
- Gao et al., 2019, Dialog state tracking: A neural reading comprehension approach
- Chao and Lane, 2019, BERT-DST: Scalable end-to-end dialogue state tracking with bidirectional encoder representations from transformer
- Kim et al., 2019, Efficient dialogue state tracking by selectively overwriting memory
- Zhang et al., 2019, Find or classify? dual strategy for slot-value predictions on multi-domain dialog state tracking
- Bapna et al., 2017, Towards Zero-Shot Frame Semantic Parsing for Domain Scaling
- Shah et al., 2019, Robust Zero-Shot Cross-Domain Slot Filling with Example Values
- Rastogi et al., 2017, Scalable Multi-Domain Dialogue State Tracking
- Rastogi et al., 2020, Towards Scalable Multi-Domain Conversational Agents: The Schema-Guided Dialogue Dataset
- Rastogi et al., 2020, Schema-Guided Dialogue State Tracking Task at DSTC8
- Shi et al., 2020, A BERT-based Unified Span Detection Framework for Schema-Guided Dialogue State Tracking
- Gulyaev et al., 2020, Goal-Oriented Multi-Task BERT-Based Dialogue State Tracker
- Ruan et al., 2020, Fine-Tuning BERT for Schema-Guided Zero-Shot Dialogue State Tracking
- Ma et al., 2020, An End-to-End Dialogue State Tracking System with Machine Reading Comprehension andWide & Deep Classification