Natural Language Generation

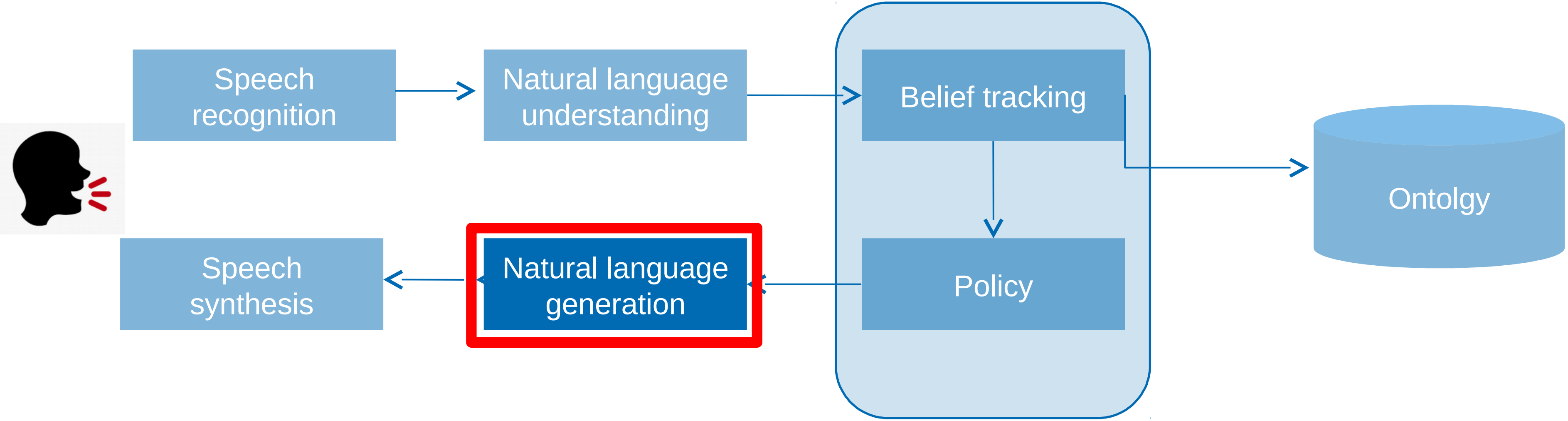
Marco Moresi

06.12.2019

Outline

1. What is NLG?
   - NLG in Spoken Dialogue Systems (SDS)
2. Evaluation Metrics
3. Traditional Approaches
   - Template-based
   - Tree-based
4. Language Modeling (LM) for NLG
   - Class based LM for NLG
   - Phrase based NLG
5. Neural Approaches
   - Semantically Conditioned
   - Domain adaptation

What is NLG?


Speech recognition
Natural language understanding
Belief tracking
Ontolgy
Policy
Natural language generation
Speech synthesis

What is NLG?

- A conversion of an abstract and formalized representation of a piece of information into natural language utterance.

What is NLG?

Heinrich Heine
Universität Düsseldorf

- A conversion of an abstract and formalized representation of a piece of information into natural language utterance.

- In SDS, NLG module is in charge of convert the abstract representation coming from the dialogue manager into a natural languange sentence.

What is NLG?

- A conversion of an abstract and formalized representation of a piece of information into natural language utterance.

- In SDS, NLG module is in charge of convert the abstract representation coming from the dialogue manager into a natural languange sentence.

Inform(restaurant=Sunset, food=Mexican) → Sunset is a Mexican restaurant.

What is NLG?

- A conversion of an abstract and formalized representation of a piece of information into natural language utterance.

- In SDS, NLG module is in charge of convert the abstract representation coming from the dialogue manager into a natural languange sentence.

Inform(restaurant=Sunset, food=Mexican) → Sunset is a Mexican restaurant.

inform(name='capital restaurant‘,phone='4153976269') → Capital restaurant phone number is 4153976269

Evaluation Metrics

What should we consider to evaluate NLG? [Stent et al, 2005]

- Adequacy
- Fluency
- Readability
- Variation

Evaluation Metrics

What should we consider to evaluate NLG? [Stent et al, 2005]

- Adequacy – Express all the input meaning (Correct meaning)
- Fluency – Syntactically correct
- Readability – Efficacy in the context.
- Variation – Different ways to express the same idea

Evaluation Metrics

What should we consider to evaluate NLG? [Stent et al, 2005]

- Adequacy – Express all the input meaning (Correct meaning)
- Fluency – Syntactically correct
- Readability – Efficacy in the context.
- Variation – Different ways to express the same idea

Bye( )

Goodbye, have a nice day!

Goodbye, thanks for use our system!

Thanks for use our system, Goodbye!

Bye!

Evaluation Metrics

Is BLEU score an option? [Papineni et al, 2002]

Evaluation Metrics

Heinrich Heine Universität Düsseldorf

Is BLEU score an option? [Papineni et al, 2002]

- BLEU is used to score similarity between paired sentence, matching n-gram.

Evaluation Metrics

Is BLEU score an option? [Papineni et al, 2002]

- BLEU is used to score similarity between paired sentence, matching n-gram.
- It punishes differences in length between candidate and reference sentences.

Evaluation Metrics

Is BLEU score an option? [Papineni et al, 2002]

- BLEU is used to score similarity between paired sentence, matching n-gram.
- It punishes differences in length between candidate and reference sentences.
- It is sensitive to factors that do not change the semantic response [Liu et al, 2016]

Evaluation Metrics

Is BLEU score an option? [Papineni et al, 2002]

- BLEU is used to score similarity between paired sentence, matching n-gram.
- It punishes differences in length between candidate and reference sentences.
- It is sensitive to factors that do not change the semantic response [Liu et al, 2016]
  - Adequacy 0.388 Fluency -0.492 [Stent et al 2005]

Evaluation Metrics

Is BLEU score an option? [Papineni et al, 2002]

- BLEU is used to score similarity between paired sentence, matching n-gram.
- It punishes differences in length between candidate and reference sentences.
- It is sensitive to factors that do not change the semantic response [Liu et al, 2016]
  - Adequacy 0.388 Fluency -0.492 [Stent et al 2005]
- Don‘t use just one metric

Evaluation Metrics

Heinrich Heine
Universität Düsseldorf

Is BLEU score an option? [Papineni et al, 2002]

- BLEU is used to score similarity between paired sentence, matching n-gram.
- It punishes differences in length between candidate and reference sentences.
  - Adequacy 0.388 Fluency -0.492 [Stent et al 2005]
- It is sensitive to factors that do not change the semantic response [Liu et al, 2016]
  - Adequacy 0.388 Fluency -0.492 [Stent et al 2005]
- Don‘t use just one metric
- Real users is always the best way to evaluate NLG

Traditional Approaches

Templated-based NLG

Traditional Approaches

Templated-based NLG

Set of rules to map dialogue acts to utterances

(+) Simple to implement, easy-control

(-) Hard to maintain, not-scalable

Traditional Approaches

Templated-based NLG

Set of rules to map dialogue acts to utterances

(+) Simple to implement, easy-control

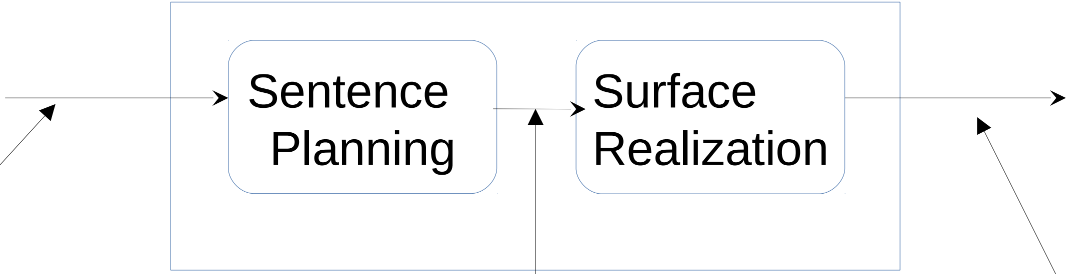
(-) Hard to maintain, not-scalable

hi()	Hello, how can I help you?
confirm(food=$Value)	Do you want a $Value restaurant?

Traditional Approaches

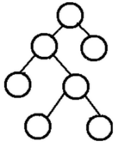
Tree-based NLG

Divide the problem into a pipeline



Inform(
name=Sunset,
food=Chinese,
)

DA



Sentence plan

RankBoost

Sunset is a Chinese restaurant

Utterance

[“Training a sentence planner for spoken dialogue using boosting”, Walker et al, 2002]

Traditional Approaches

Heinrich Heine Universität Düsseldorf

Tree-based NLG

(+) Can generate sentences with complex linguistic structure
(-) A lot of hand-craft job, feature engineering.

Language Modeling

Class-based LM

- Language model

$$P(W) = \prod_t P\left(w_t | w_0, w_1, \dots w_{t-1}\right)$$

- Class-based Language Modeling

$$P(W|\mathbf{u}) = \prod_t P\left(w_t | w_0, w_1, \dots w_{t-1}, \mathbf{u}\right)$$

- Decoding

$$W^* = \underset{W}{\operatorname{argmax}} P(W|\mathbf{u})$$

Utterance classes

- inform-hotel
- inform-airport
- query-confirm

.

.

.

- other

Markov assumption

["Stochastic Language Generation for Spoken Dialogue Systems", Oh and Rudnicky, 2000]

Language Modeling

(+) Easy to implement, easy to understand

(-) Does not scale, inefficient, error-prone, corpus-based

Neural Approaches

RNN as a solution, why?

Neural Approaches

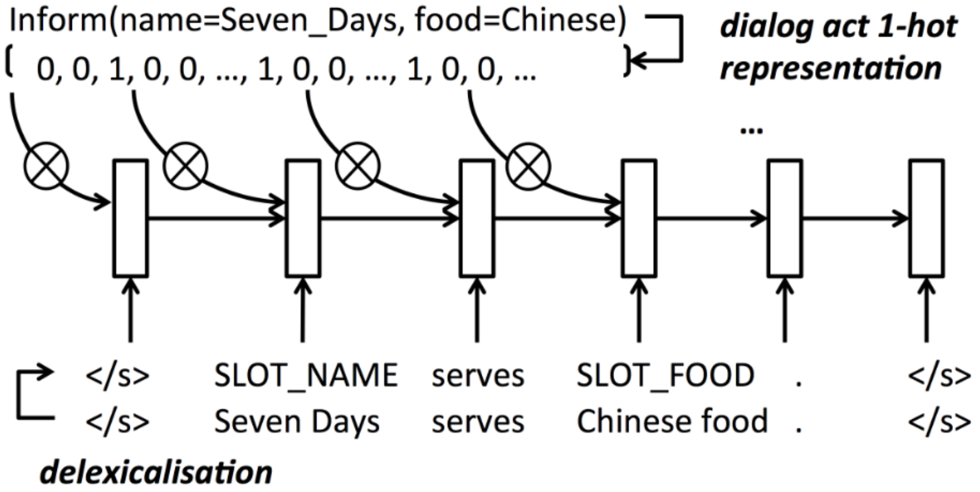
RNN as a solution, why?

- More control
- No hand-crafted rules
- No feature engineering
- No expert knowledge is required
- Long term dependencies

Neural Approaches

Heinrich Heine Universität Düsseldorf

Stochastic Language Generation in Dialogue



[Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking,Wen et al,2015a]

Neural Approaches

Stochastic Language Generation in Dialogue

(+) Provide control in the generated utterances, not semantic alignments are necessary.

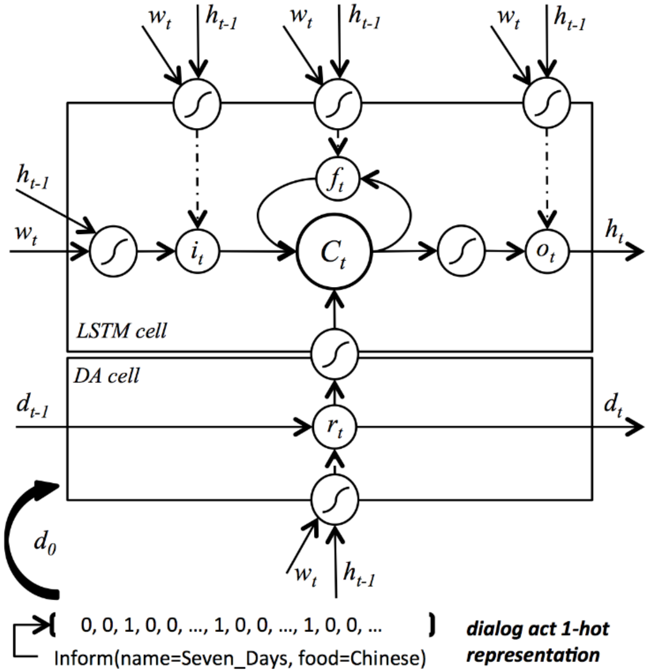
(-) Empirically, utterances with semantic repetitions are observed
e.g: “Sunset is a great Chinese restaurant that serves Chinese”

(-) RNN select words based only on the preceding history, not take in count backward context

[Wen et al,2015a]

Neural Approaches

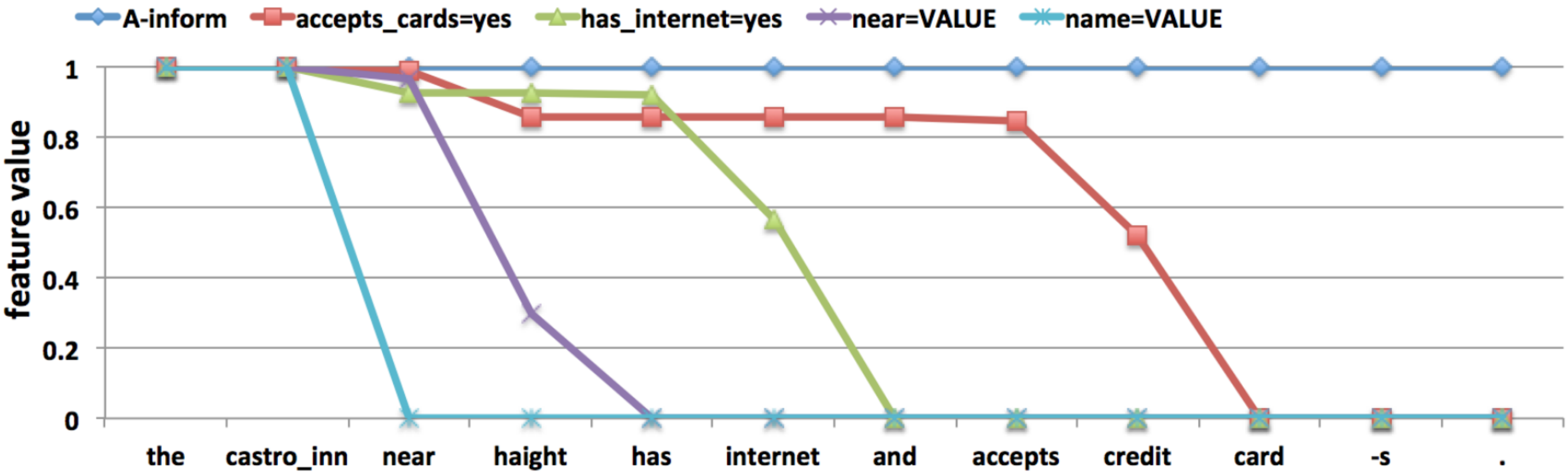
Semantically conditioned LSTM



["Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems", Wen et al,2015b]

Neural Approaches

Semantically conditioned LSTM



(b) An example realisation from SF hotel domain

[“Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems”, Wen et al,2015b]

Neural Approaches

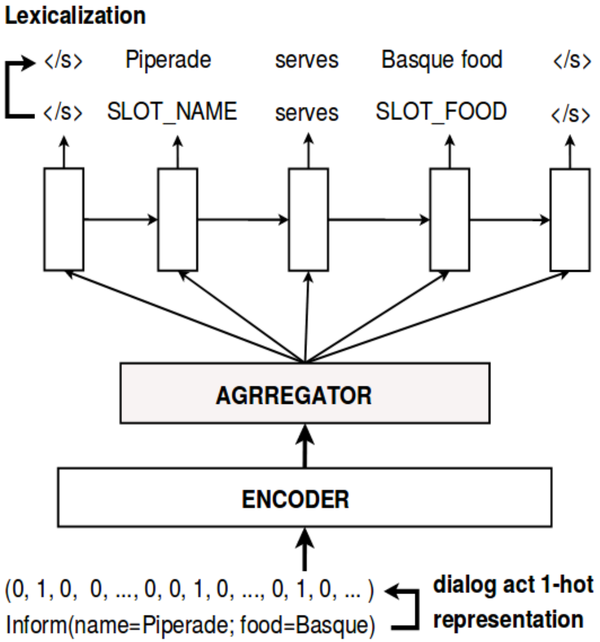
Heinrich Heine Universität Düsseldorf

Semantically conditioned LSTM

(+) Provide control of the generated utterances.
(+) Incorporate new gates to avoid semantic repetition
(-) Can not handle binary neither don’t_care slots

Neural Approaches

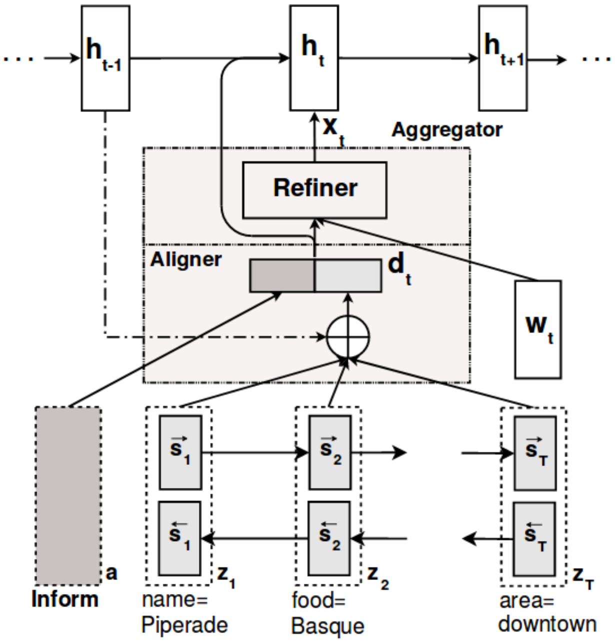
Encoder-Decoder with Semantic Aggregation



[“Neural-based natural language generation in dialogue using rnn encoder-decoder with semantic aggregation”,Van-Khanh and Le-Minh, 2017]

Neural Approaches

Encoder-Decoder with Semantic Aggregation



Attentional Refiner over Attention

$$\mathbf{x}_t = f_R(\mathbf{d}_t, \mathbf{w}_t)$$

$$\mathbf{d}_t = \mathbf{a} \oplus \sum_i \alpha_{t,i}\mathbf{s}_i$$

$$\mathbf{s}_i = \overrightarrow{s_i} + \overleftarrow{s_i}$$

$$\mathbf{z}_i = \mathbf{o}_i \oplus \mathbf{v}_i$$

[“Neural-based natural language generation in dialogue using rnn encoder-decoder with semantic aggregation”,Van-Khanh and Le-Minh, 2017]

Neural Approaches

Encoder-Decoder with Semantic Aggregation

- (+) Propose a new way to represent the DA and semantic meaning, Also propose a possible solution to binary slots.

[Van-Khanh and Le-Minh, 2017]

Neural Approaches

- Limited data of a particular domain.
- Add a new domain.

How can we tackle these situations?

Neural Approaches (Domain Adaptation)

Heinrich Heine
Universität Düsseldorf

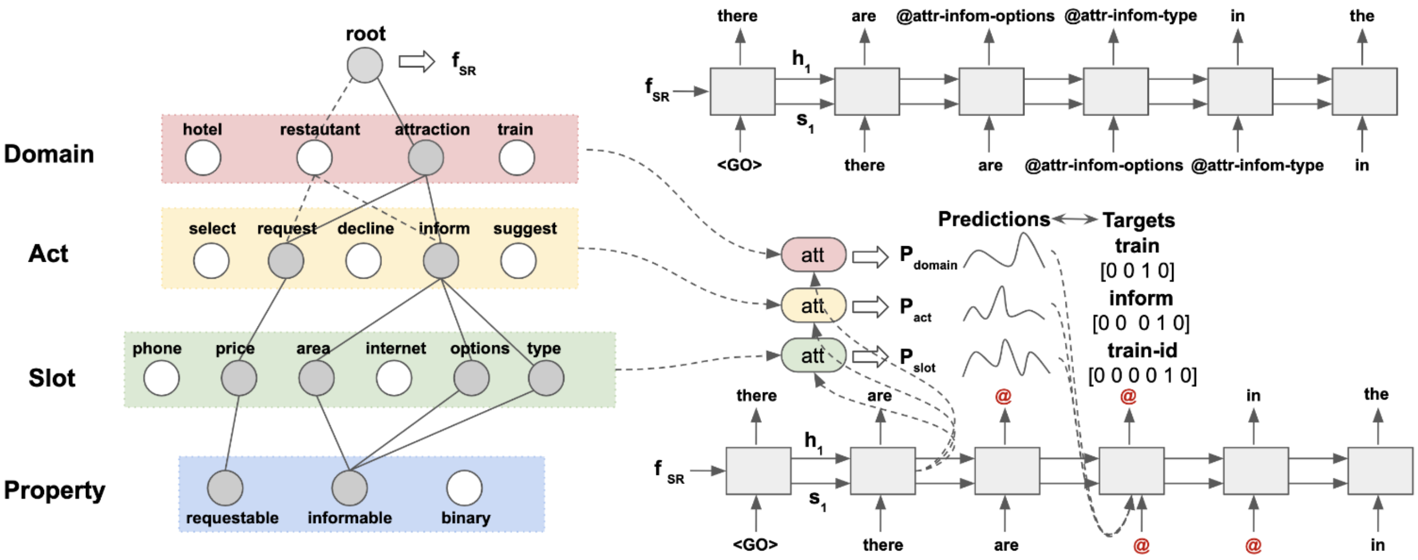
Domain Adaptation

- Moving from limited-domain NLG to open domain implies that the number of semantic input combinations grows exponentially.

- Allows us to share knowledge from one domain to another

Neural Approaches (Domain Adaptation)

Tree Structured semantic Encoder with Knowledge Sharing



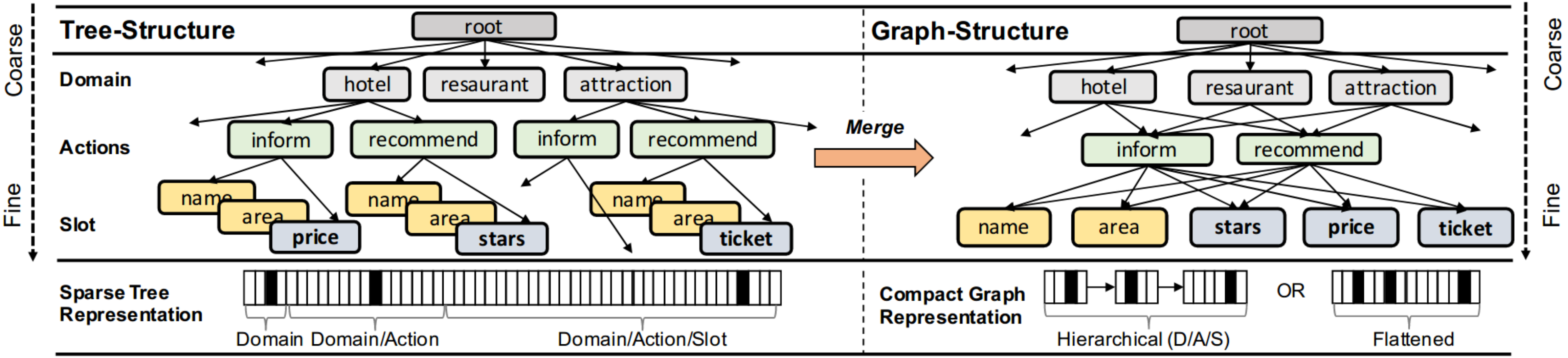
(a) Semantically Tree-structured Encoder

(b) LSTM-based Decoder

[“Tree-Structured Semantic Encoder with Knowledge Sharing for Domain Adaptation in Natural Language Generation”. Tseng et al 2019]

Neural Approaches (Domain Adaptation)

Hierarchical Disentangled Self-Attention

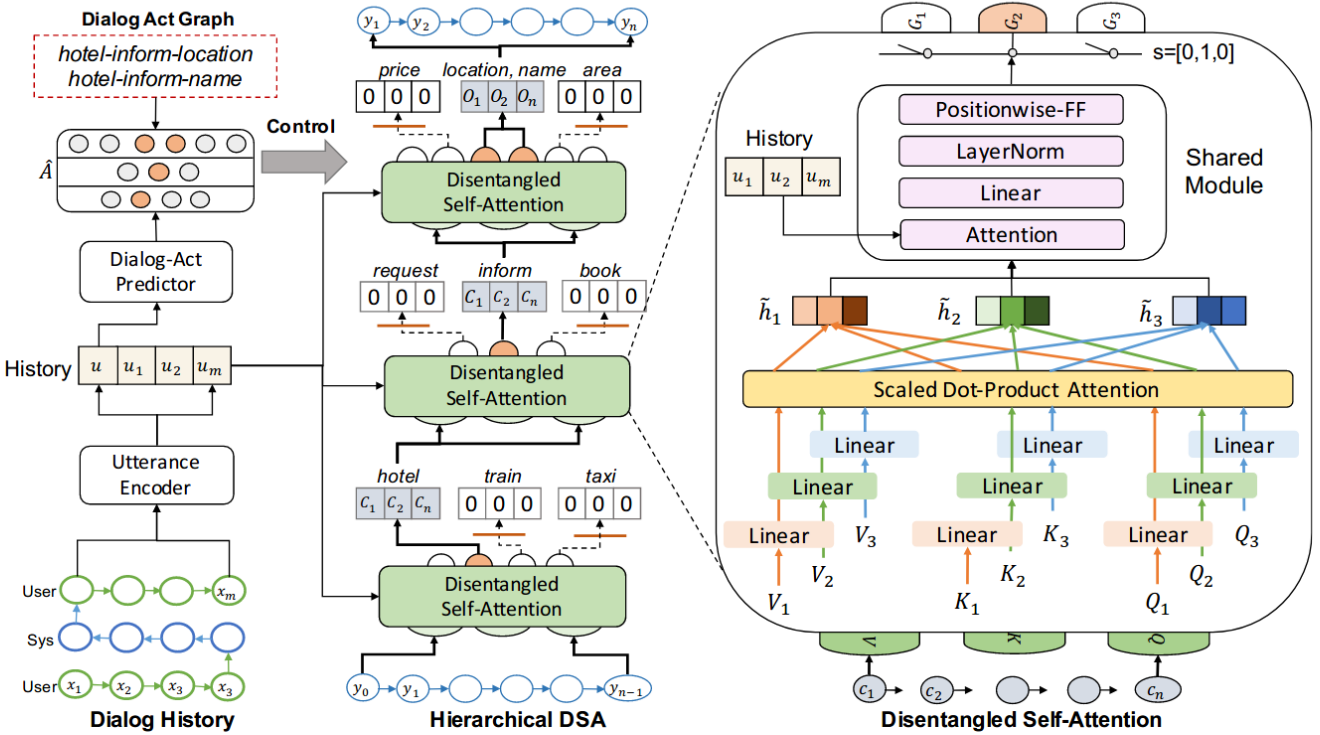


["Semantically Conditioned Dialog Response Generation via Hierarchical Disentangled Self-Attention" Wenhu et al 2019]

Neural Approaches (Domain Adaptation)

Heinrich Heine Universität Düsseldorf

Hierarchical Disentangled Self-Attention

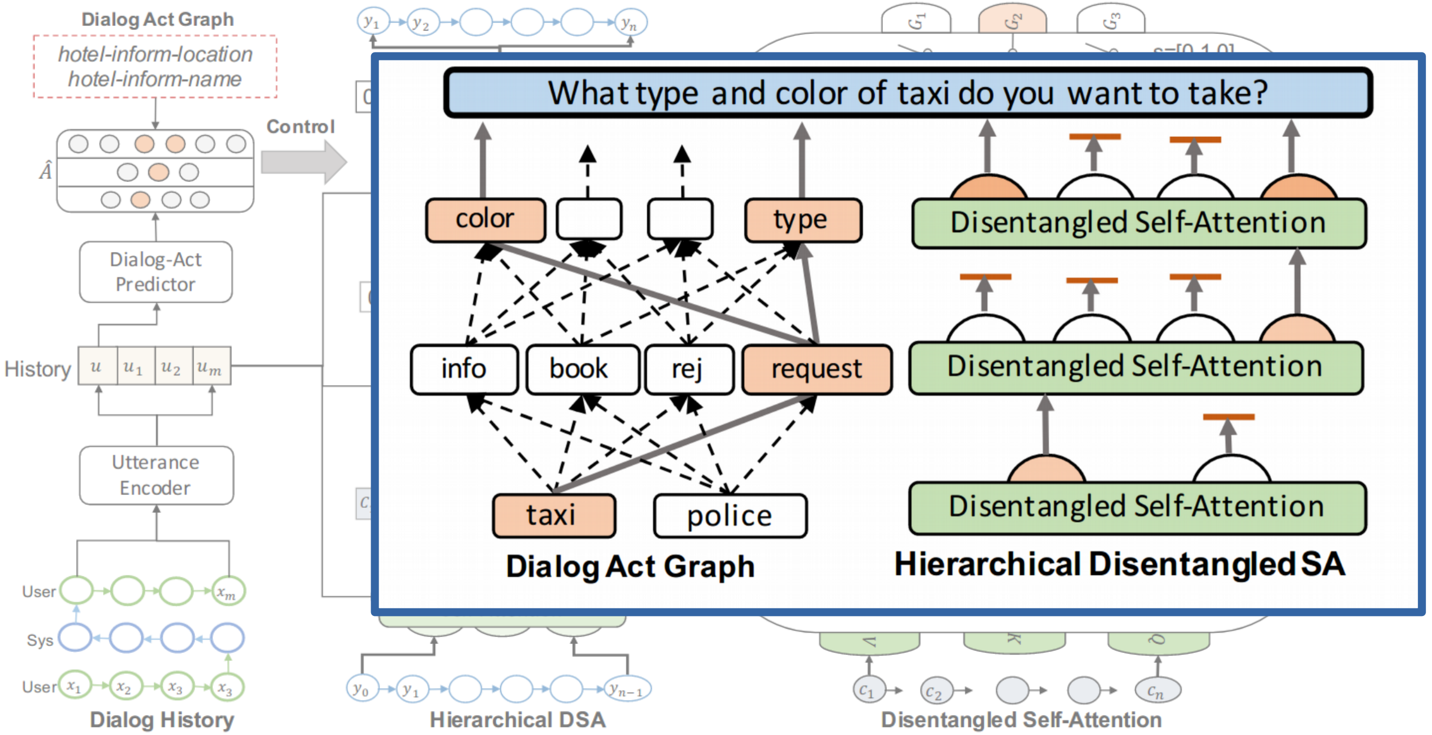


["Semantically Conditioned Dialog Response Generation via Hierarchical Disentangled Self-Attention" Wenhu et al 2019]

Neural Approaches (Domain Adaptation)

Heinrich Heine Universität Düsseldorf

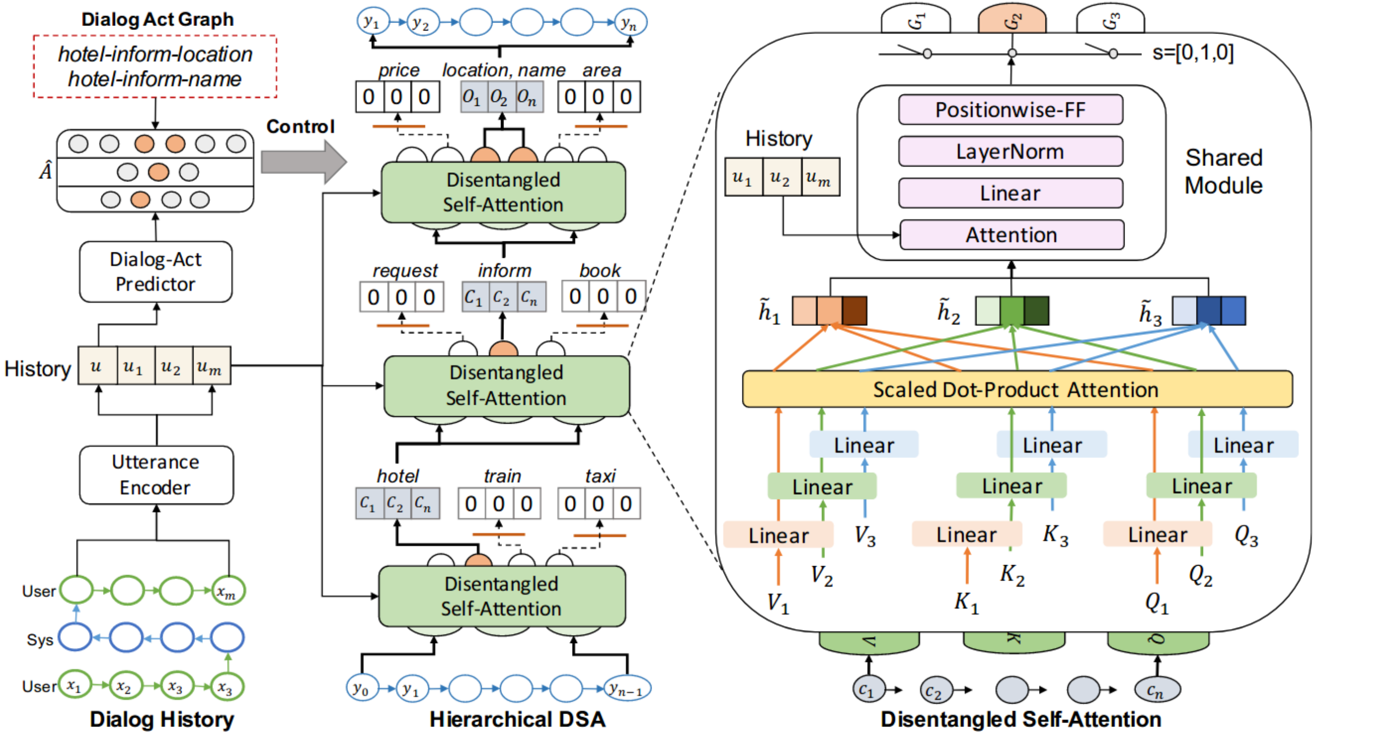
Hierarchical Disentangled Self-Attention



["Semantically Conditioned Dialog Response Generation via Hierarchical Disentangled Self-Attention" Wenhu et al 2019]

Neural Approaches (Domain Adaptation)

Hierarchical Disentangled Self-Attention



["Semantically Conditioned Dialog Response Generation via Hierarchical Disentangled Self-Attention" Wenhu et al 2019]

Neural Approaches (Domain Adaptation)

(+) Sharing knowledge
(-) Define the tree/graph structure in advance

Conclusion

RNN in NLG

- Conditioned RNN, provides better utterances.
- More human-like and less structured.
- No more expert domain is required.
- Quicker development cycle, no more hand-crafted rules.

Domain Adaptation

- Sharing knowledge is a proper way to tackle lack of data problem in certain domains

Challenges

Heinrich Heine
Universität Düsseldorf

- Generate longer and complex sentences.

- How to incorporate non-labeled data into generation process to enrich the utterances?

- Scale systems across different domains

- Delexicalization

- Semantic representation

- Metric for automatic evaluation

Questions?

Thanks! :)

Bibliography

“Evaluating evaluation methods for generation in the presence of variation”,

- Amanda Stent, Mawhew Marge, Mohit Singhai, CICLing 2005
- "BLEU: a method for automatic evaluation of machine translation."
- Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu , ACL, 2002.
- "How not to evaluate your dialogue system: An empirical study of unsupervised
- evaluation metrics for dialogue response generation." Liu, Chia-Wei, Ryan Lowe,
- Iulian V. Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau.
- arXiv preprint arXiv:1603.08023 (2016).
- “Training a sentence planner for spoken dialogue using boosting”,
- Marilyn Walker, Owen Rambow, Monica Roga, Computer Speech and Language 2002
- “Stochastic Language Generation for Spoken Dialogue Systems”,
- Alice Oh and Alexander Rudnicky, NAACL workshop on Conversational Systems 2000

Bibliography

“Stochastic language generation in dialogue using recurrent neural networks
with convolutional sentence reranking”, Tsung-Hsien Wen, Milica Gašić, Nikola Mrksic,

- Pei-Hao Su, David Vandyke, and Steve Young, SigDial 2015a.
- “Semantically Conditioned LSTM-based Natural Language Generation for Spoken
- Dialogue Systems”, Tsung-Hsien Wen, Milica Gašić, Nikola Mrksic, Pei-Hao Su,
- David Vandyke, and Steve Young, EMNLP 2015b.
- "Neural-based natural language generation in dialogue using rnn encoder-decoder
- with semantic aggregation." Tran, Van-Khanh, and Le-Minh Nguyen.
- arXiv preprint arXiv:1706.06714 (2017)
- "Tree-Structured Semantic Encoder with Knowledge Sharing for Domain Adaptation
- in Natural Language Generation." Tseng, Bo-Hsiang, Paweł Budzianowski,
- Yen-Chen Wu, and Milica Gašić arXiv preprint arXiv:1910.06719 (2019).
- "Semantically Conditioned Dialog Response Generation via Hierarchical Disentangled
- Self-Attention." Chen, Wenhu, Jianshu Chen, Pengda Qin, Xifeng Yan, and William
- Yang Wang. arXiv preprint arXiv:1905.12866 (2019).

Neural Approaches (Domain Adaptation)

emantically Conditioned Dialog Response Generation via Hierarchical Disentangled Self-Attention” Wenhu et al 2019]