Unsupervised NLU in dialogue

Vojta Hudeček   09/16/2019

UC Davis

- 3rd largest UC enrollment
- ~ 40k students, 30k bikes
- Originally an university farm, nowadays wide variety of research

Zhou Yu

- Director Davis NLP since 2017
- Alexa prize 2018 winner
- Specialized in interactive systems
- Multi-modal dialogues, low-resource methods

Task-oriented systems - Example

User	I want to have some chinese food.
System	Which neighborhood are you interested in?
User	Somewhere downtown.
System	Golden palace is a good chinese restaurant in the city centre.
User	Sounds good, what's their address?
System	The address of Golden palace is 10 Downing Street.
User	Thanks.

Task-oriented systems - slots

User	I want to have some chinese food.
System	Which neighborhood are you interested in?
User	Somewhere downtown.
System	Golden palace is a good chinese restaurant in the city centre.
User	Sounds good, what's their address?
System	The address of Golden palace is 10 Downing Street.
User	Thanks.

Task-oriented systems - intents

User	I want to have some chinese food.
System	Which neighborhood are you interested in?
User	Somewhere downtown.
System	Golden palace is a good chinese restaurant in the city centre.
User	Sounds good, what's their address?
System	The address of Golden palace is 10 Downing Street.
User	Thanks.

Inform

Request

Task-oriented systems - state

| User	I want to have some chinese food.	{food: chinese}
System	Which neighborhood are you interested in?	
User	Somewhere downtown.	{food: chinese, area: center}
System	Golden palace is a good chinese restaurant in the city centre.	
User	Sounds good, what's their address?	{food: chinese, area: center, requested_address: yes}
System	The address of Golden palace is 10 Downing Street.	
User	Thanks.	

Task-oriented systems - state

User	I want to have some chinese food.
System	Which neighborhood are you interested in?
User	Somewhere downtown.
System	Golden palace is a good chinese restaurant in the city centre.
User	Sounds good, what's their address?
System	The address of Golden palace is 10 Downing Street.
User	Thanks.

`{food: chinese}`

`{food: chinese,`
` area: center}`   **API call**

`{food: chinese,`
` area: center,`
` requested_address: yes}`

Task-oriented dialogue systems

- The purpose is to achieve some goal, accomplish a task.
- Recommend a restaurant, book a flight, etc.
- Task accomplishment is more important than fluency or engagement.

Task-oriented systems - traditional way

- Pipelined modular architecture
- State representation
  - user intention
  - {(slot, value)}
- Later end-to-end systems, similar state representation
- Complex annotations
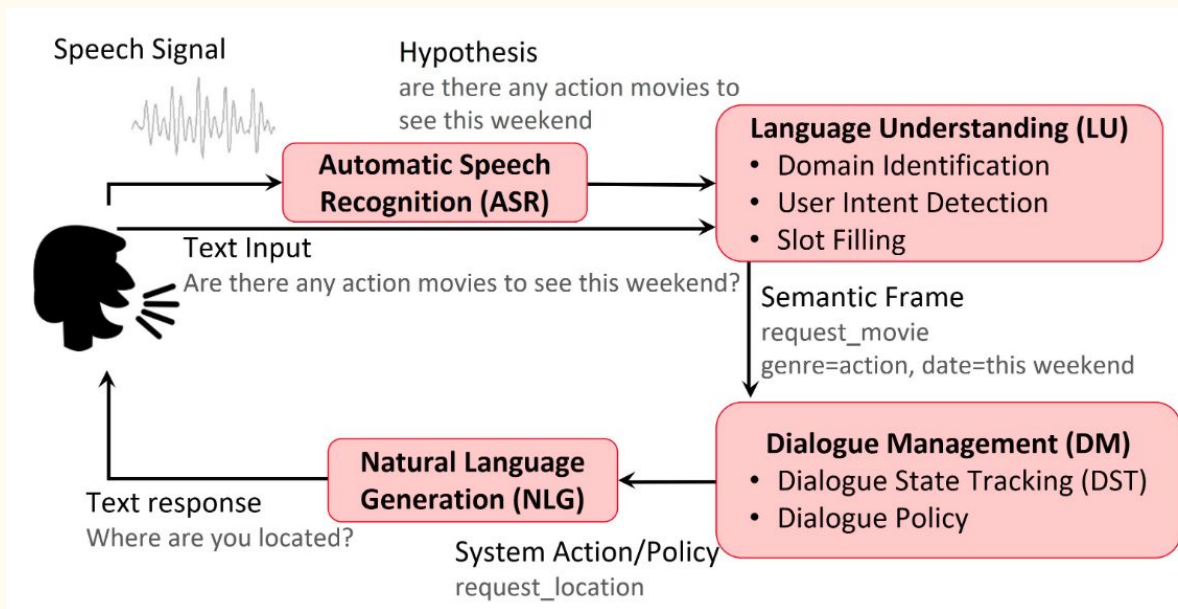
Task-oriented dialogue systems - recent work

- Defined as seq-to-seq
- State representation: **belief span**
- No explicit structure
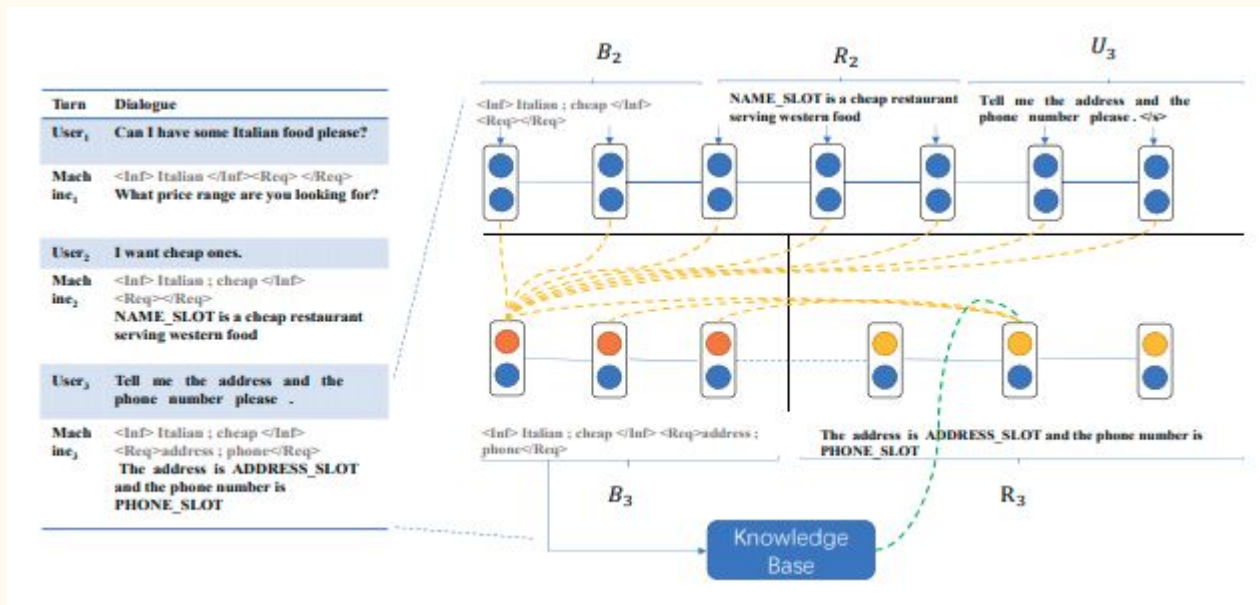
Lei et al. (2018, July). Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1437-1447).

Task-oriented dialogue systems - simplified state

- simpler structure
- easier to label
- lack of structure might be problematic (?)

```
● inform
    ○ price=cheap
    ○ food=chinese
    ○ area=UNK
    ○ ...
● request
    ○ address
    ○ ...
```

vs.

```
<inf>cheap;chinese</inf>
<req>address</req>
```

Task-oriented dialogue systems - no state?

- Can't we do without state then?
  - Latent state representations, end-to-end systems
  - Can somehow work in terms of dialogue generation
- State is not just for the agent's decision!
  - DB interaction
  - Interpretability
  - Hard rules (business policy)

Dialogue state tracking - summary

- Two important kind of information - pair (intent, slot).
- Need for expert annotation.
  - Design of structure and labeling itself.
- Can we automate the annotation?
- Are the noisy annotations useful?

Frame semantic parsing

- To capture semantic frames and respective frame elements
- A lot of research, error prone on out-of-domain data
- https://framenet.icsi.berkeley.edu/fndrupal/

can i have a cheap restaurant

Frame: expensiveness
FT LU: cheap

Frame: capability
FT LU: can FE Filler: i
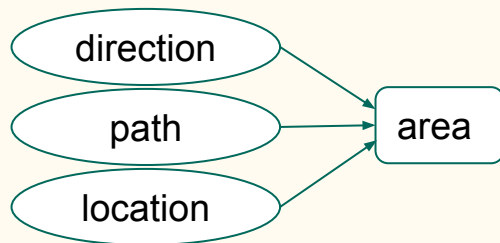
Frame: locale_by_use
FT/FE LU: restaurant

Das et al. (2010) SEMAFOR 1.0: A probabilistic frame-semantic parser. Language Technologies Institute, School of Computer Science, Carnegie Mellon University.
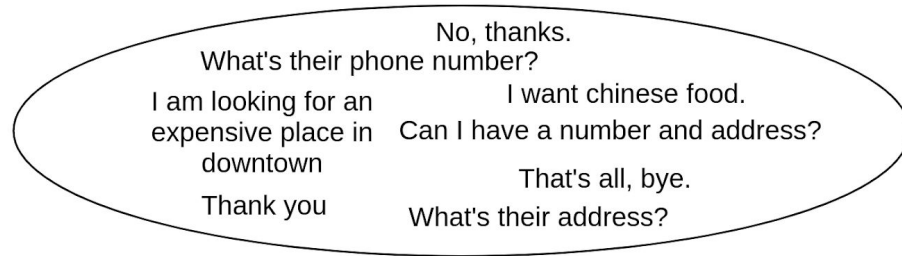
Frame semantic parsing - problems

- Many frames are irrelevant to the task.
- Some of them contain the same information.
    - In the respective domain context
- Parser performance is limited.
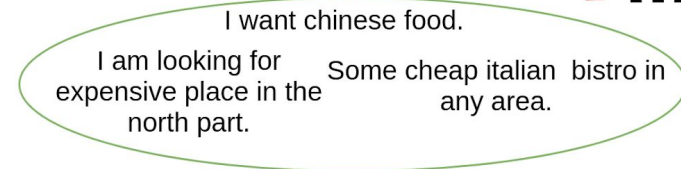- Frames themselves don't give us information about intent.

User utterances

No, thanks.

What's their phone number?

I am looking for an expensive place in downtown

I want chinese food.

Can I have a number and address?

Thank you

That's all, bye.

What's their address?

Clustering

Intent 1 Intent 2 Intent n ■ ■ ■

I want chinese food.

I am looking for expensive place in the north part.

Some cheap italian bistro in any area.

Intent detection

Slot induction

Frame semantic parser

{Origin: Chinese, Desire: want, ...}
{Expensiveness: cheap, Origin: italian, ...}
{Expensiveness: cheap, Direction: north, ...}

Ranking model

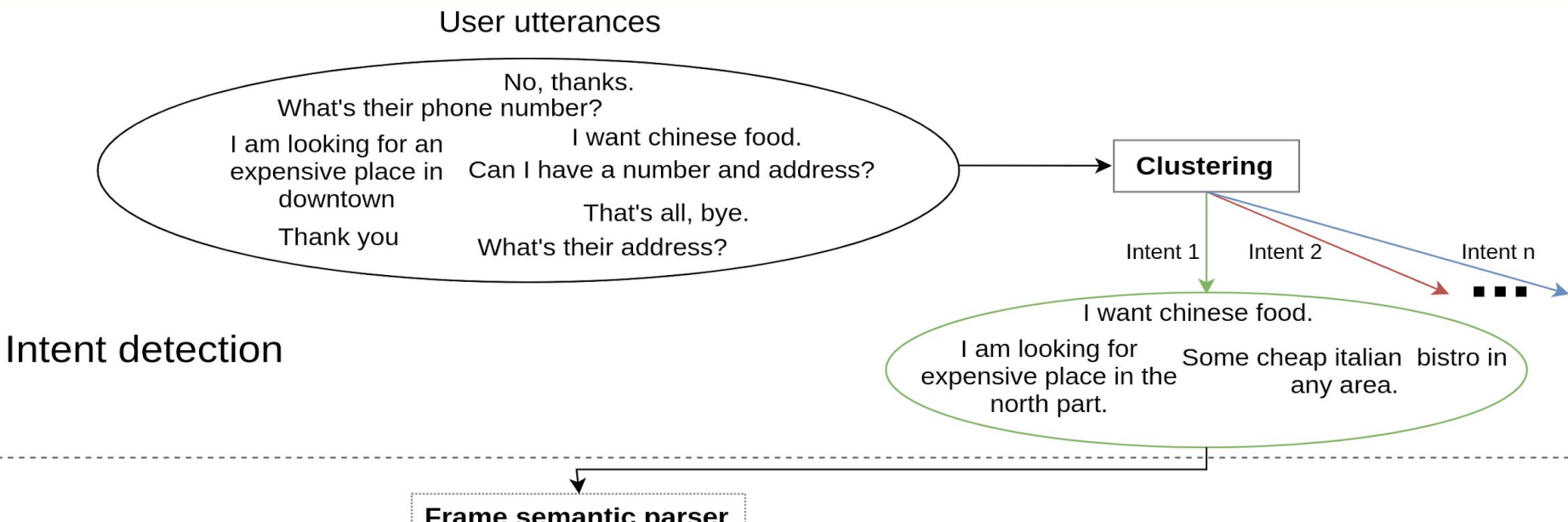| Frame	Freq.	Coh.	Instances
....			
Origin	0.16	0.88	italian, chinese, british, ...
Direction	0.11	0.89	north, west, straight, ...
Path	0.12	0.90	north, west, east, ...

Origin Type ... **Direction Path ...** **Expensiveness ...**

Topic detection

Intent clustering

- Sentences transformed into feature vectors
- Spectral clustering
- How many intents?

User utterances

No, thanks.
What's their phone number?
I am looking for an expensive place in downtown
I want chinese food.
Can I have a number and address?
That's all, bye.
Thank you
What's their address?

Clustering

Intent 1   Intent 2   Intent n

Intent detection

I want chinese food.
I am looking for expensive place in the north part.
Some cheap italian bistro in any area.

Frame semantic parser

Slot Induction

- Each intent cluster processed in isolation
- Semantic frames are ranked based on several criterions (unsupervised)
- Best candidates are selected and grouped based on topics

Slot induction

Frame semantic parser

{Origin: Chinese, Desire: want, ...}
{Expensiveness: cheap, Origin: italian, ...}
{Expensiveness: cheap, Direction: north, ...}

Ranking model

Frame	Freq.	Coh.	Instances
....			
Origin	0.16	0.88	italian, chinese, british, ...
Direction	0.11	0.89	north, west, straight, ...
Path	0.12	0.90	north, west, east, ...

Origin
Type
...

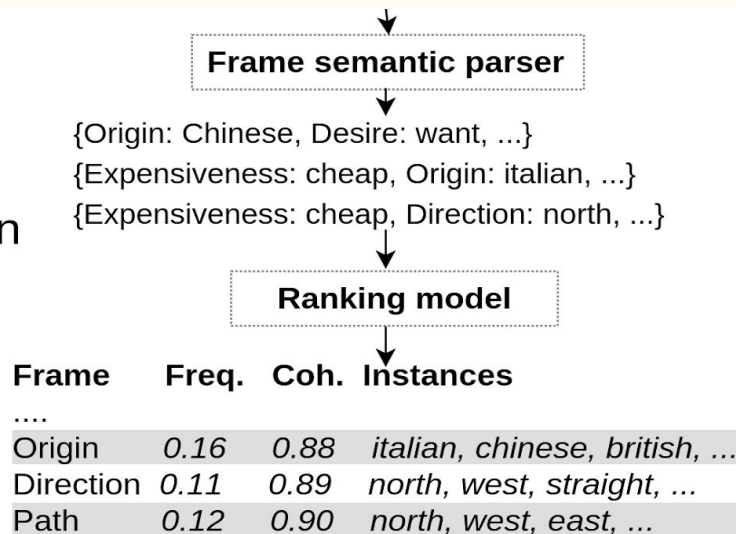Direction
Path
...

Expensiveness
...

Topic detection
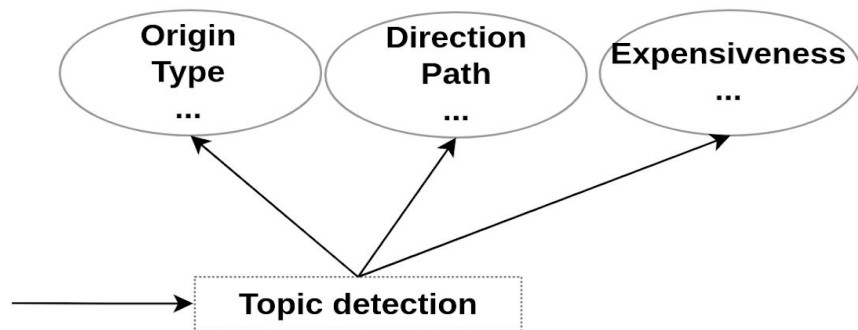
Slot Induction - results

- F1 on the slot filling task
  - We map frame groups to reference slots for evaluation

CamRest676

price	area	food	average
.353	.426	.584	.454

MultiWOZ-hotel

price	area	people	day	type	average
.059	.181	.652	.866	.000	.352

Examples

I am interested in a restaurant located in the north part of town
that specializes in Indonesian food, please.
{0: "north", 1: "Indonesian"}
what about italian food?
{0: "north", 1: "italian"}
Yes, may I have the address and phone number please?
{0: "north", 1: "~~Yes~~"}
thank you, goodbye.

cheap restaurant, please
{2: "cheap"}
do you have any serves jamaican food ?
{1: "jamaican food", 2: "cheap"}
How about mediterranean?
{1: "~~jamaican food~~", 2: "cheap"}
Yes, and the postcode please.
{1: "~~jamaican food~~", 2: "cheap"}
No, thank you, goodbye,

Does it help to train a system?

- Joint goal accuracy
  - percentage of correctly labeled turns

% supervised data	0	25
original	0,6543	0,8383
noisy state	**0,6721**	**0,8952**

Future work

- Get rid of semantic parser dependency
- Improve the precision.
- Use the labels better in the downstream task
- Joint detection of intents and slots (some experiments already)

Thank you!
(questions?)