

Conversational Information Access

[DAT640] Information Retrieval and Text Mining

Ivica Kostric & Krisztian Balog
University of Stavanger

17.10.2022



CC BY 4.0

In this module

1. Conversational AI
2. Architecture
3. Examples

Conversational AI

Recap

- So far the focus was on single turn IR techniques
 - Represent documents (items) and queries
 - Rank documents (items) with respect to queries
 - Evaluate the goodness of rankings
- Many information needs are not answered by a single query

Question

What is conversational AI?

Conversational AI

- The branch of computer science concerned with making machines interact with people in a human-like way
- Can use speech, text or mixed interfaces
- Design and abilities vary
 - Task-oriented dialogue agents (e.g., Siri, Alexa, Google Now/Home, Cortana)
 - Chatbots (e.g., ELIZA, PARRY, Xiaolce, BlenderBot)

Conversational AI - Traditional distinction¹

Task-oriented (goal-driven)

- Aim to assist users to **solve a specific task** (as efficiently as possible)
- Dialogues follow a **clearly designed structure** (flow) that is developed for a particular task in a **closed domain**
- Well-defined measure of performance that is explicitly related to **task completion**

Non-task-oriented (non-goal-driven)

- Aim to **carry on an extended conversation** ("chit-chat") with the goal of mimicking human-human interactions
- Developed for **unstructured**, open domain conversations
- Objective is to be **human-like**, i.e., able to talk about different topics (breadth and depth) in an engaging and coherent manner

¹Chen et al. A Survey on Dialogue Systems: Recent Advances and New Frontiers. SIGKDD Explor. Newsl. 19(2), 2017.

Conversational AI - Contemporary distinction²

Task-oriented

- Aim to assist users to **solve a specific task** (as efficiently as possible)
- Dialogues follow a **clearly designed structure** (flow) that is developed for a particular task in a **closed domain**
- Well-defined measure of performance that is explicitly related to **task completion**

Social chat

- Aim to **carry on an extended conversation** ("chit-chat") with the goal of mimicking human-human interactions
- Developed for **unstructured, open domain** conversations
- Objective is to be **human-like**, i.e., able to talk about different topics (breadth and depth) in an engaging and coherent manner

Interactive QA

- Aim to **provide concise, direct answers** to user queries
- Dialogues are **unstructured**, but commonly follow a question-answer pattern; mostly **open domain** (dictated by the underlying data)
- Evaluated with respect to the **correctness of answers** (on the turn level)

²Gao et al. Neural Approaches to Conversational AI. Found. Trends Inf. Retr. 13(2-3), 2019.

Question

What is conversational information access?

Question

Where does conversational information access fit within conversational AI?

Conversational information access

Task-oriented

- Aim to assist users to **solve a specific task** (as efficiently as possible)
- Dialogues follow a **clearly designed structure** (flow) that is developed for a particular task in a **closed domain**
- Well-defined measure of performance that is explicitly related to **task completion**



Social chat

- Aim to **carry on an extended conversation** ("chit-chat") with the goal of mimicking human-human interactions
- Developed for **unstructured, open domain** conversations
- Objective is to be **human-like**, i.e., able to talk about different topics (breadth and depth) in an engaging and coherent manner

Interactive QA

- Aim to **provide concise, direct answers** to user queries
- Dialogues are **unstructured**, but commonly follow a question-answer pattern; mostly **open domain** (dictated by the underlying data)
- Evaluated with respect to the **correctness of answers** (on the turn level)



Conversational information access

- A subset of conversational AI systems that specifically aim at a task-oriented sequence of exchanges
- Supports multiple user goals, including search, recommendation and exploratory information gathering
- Requires multi-step interactions over possibly multiple modalities
- Combine elements from both task-oriented and interactive QA systems
- Consider both long-term and short-term information about the user when solving information seeking tasks

Question

Hasn't search always been conversational?

Search as a conversation

Where will ECIR take place in 2022?

G

The ECIR conference is the premium European forum for presentation of new research results in the area of Information Retrieval (IR). The 2022 conference will take place in **Stavanger**. Dec 21, 2020

<https://www.uis.no> › research › uis-will-host-the-ecir-conf... ⋮

UiS will host the ECIR conference 2022 | University of Stavanger

🔍 About featured snippets • 🗉 Feedback

<https://ecir2022.org> ⋮

ECIR 2022 – 44th European Conference on Information ...

The 44th edition of ECIR is planned to be held as a physical conference (with support for ...
Abstracts are due on September 30, papers on October 7 (AoE).

[Important Dates](#) · [Calls](#) · [Call for Full Papers](#) · [Call for Demonstrations](#)

Search as a conversation

places to visit in Stavanger



G

Top sights in Stavanger

Most visited Kid-friendly Outdoors Museums



Pulpit Rock
4.8 ★★★★★ (2,695)
Landmark outcropping with fjord views



Norwegian Petroleum Museum
4.5 ★★★★★ (2,454)
Petroleum museum in landmark building



Sverd i fjell
4.5 ★★★★★ (3,895)
Monument with 3 cast swords by a fjord

→ More things to do

www.planetware.com - stavanger-norway

15 Top-Rated Tourist Attractions in Stavanger | PlanetWare

Nov 17, 2020 — Pulpit Rock (Preikestolen) Pulpit Rock (Preikestolen) Norwegian Petroleum Museum Norwegian Petroleum Museum Norwegian Gaming Museum Norwegian Gaming Museum | Chris Walsh / photo modified. Stavanger Cathedral The Vaberg Tower (Vabergstønen) Stavanger Museum, Gamle Stavanger (Old Stavanger) Lysefjord

sfullivan.com - stavanger best things to do

Best Things To Do in Stavanger Norway (These 8 Are a Must!)

Nov 13, 2020 — This is not a complete list of Stavanger attractions. This beautiful region has so much more to offer than that, but if you want to see the main ...

People also ask

Is Stavanger worth visiting?

What is Stavanger famous for?

What is there to do in Stavanger today?

How far is Oslo from Stavanger?

Feedback

www.tripadvisor.com - Stavanger

Things to Do in Stavanger, Norway - Stavanger Attractions

Top Attractions in Stavanger - 1. Lysefjord 1,555 reviews See 18 Experiences - 2. Old Stavanger 2,345 reviews - 3. Flor og Fjære 745 reviews - 4. Norwegian ...



Stavanger

City in Norway

Stavanger is a city in southwestern Norway, in the center of town, Stavanger Cathedral dates back to the city's 12th-century founding. Stavanger Museum chronicles the city's history and displays preserved whistles. The Norwegian Petroleum Museum illustrates the oil industry with submarines, a large drill bit and an escape chute. The shopping street Øvre Hødegata is known for its colorful houses. — Google

Area: 71.35 km²

Metropolitan area: 2,288 km²

Urban area: 17.96 km²

Weather: -3°C, Wind E at 2 m/s, 70% Humidity

Population: 130,754 (2014) and today

Mayor: Karl Nilsen Nordfjord

Local time: Friday 17:52

Plan a trip

Things to do

3-star hotel averaging NOK 1,147

Upcoming Events

Colleges and Universities: University of Stavanger, VID Specialized University

People also search for

View 15+ more



Feedback

Search as a conversation

places to visit in Stavanger

G

Top sights in Stavanger

Most visited | Airports | Outdoors | Museums

Pulpit Rock
4.2 ★★★★★ (1,000)
Landmark overlooking with best views

Norwegian Petroleum Museum
4.1 ★★★★★ (2,400)
Petroleum museum in landmark building

Breidt (Spill)
4.1 ★★★★★ (1,000)
Monument with 3 cast screws by a spill

[More things to do](#)

[www.planetware.com/stavanger-norway.html](#)

15 Top-Rated Tourist Attractions in Stavanger | PlanetWare
Nov 17, 2023 — Pulpit Rock (Preikestolen) Pulpit Rock (Preikestolen) Norwegian Petroleum Museum Norwegian Petroleum Museum Norwegian Gaming Museum Norwegian Gaming Museum Chris Walsh (photo modified) Stavanger Cathedral The Viking Tower (Vikingtårnet) Stavanger Museum. Gaute Stavanger (Stavanger) Lyndeland

Adventure alert: stavanger best things to do

Best Things To Do in Stavanger Norway (These 8 Are a Must!)
Nov 15, 2023 — This is not a complete list of Stavanger attractions. This beautiful region has so much more to offer than that, but if you want to see the main...

People also ask

Is Stavanger worth visiting? ▾

What is Stavanger famous for? ▾

What is there to do in Stavanger today? ▾

How far is Oslo from Stavanger? ▾

[Feedback](#)

Things to Do in Stavanger, Norway - Stavanger Attractions
Top Attractions in Stavanger: 1. Lyndeland 1,000 reviews See 18 Experiences 2. 050 Stavanger 2,340 reviews 3. For up (Fjell) 740 reviews 4. Norwegian ...

What are the top attractions to visit in Stavanger? ▾

What are the best outdoor activities in Stavanger? ▾

What are the most popular things to do in Stavanger with kids? ▾

Stavanger
City in Norway

Stavanger is a city in southwestern Norway, in the center of west, Stavanger Cathedral dates back to the city's 10th-century founding. Stavanger Museum chronicles the city's history and displays preserved vessels. The Norwegian Petroleum Museum dominates the oil industry with submersibles, a large drill bit and an escape chute. The shopping street Øvre Havnegate is known for its colorful houses. — Google

Area: 71.33 km²
Metropolitan area: 2,088 km²
Urban area: 77.86 km²

Weather: -1°C, Wind E at 2 m/s, 70% Humidity
Population: 130,794 (2019) See more
Mayor: Håge Hovda Nordbø
Local time: Friday 17:02

Plan a trip

Things to do

30+ 3-star hotel averaging NOK 1,147

Upcoming Events

Colleges and Universities: University of Stavanger, VSD Specialist University

People also search for

[Bergen](#) [Sandnes](#) [Trondheim](#) [Oslo](#) [Norway](#)

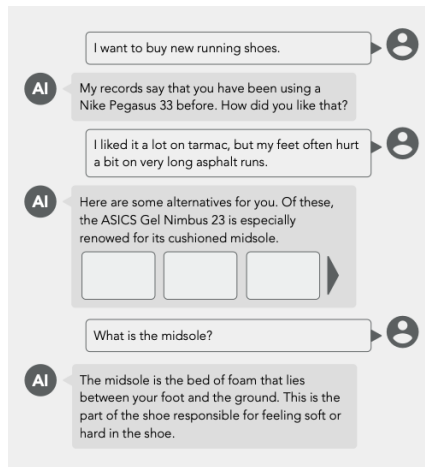
By Stavanger you meant the city in Norway, right?

Do you want to know more about Stavanger?

Would you be interested in other cities in Norway as well?

Conversational, in the “modern” sense

- Degree of personalization and long-term user state
- Support for complex (multi-step) tasks
- Answer generation vs. answer retrieval
- Dialogue setting where a screen or keyboard may not be present
- Mixed initiative



Architecture

Architecture

- Chatbots
- Task-oriented Dialogue Systems

Chatbots

- Goal is to mimic the unstructured conversations characteristic of informal human-human interaction
- The simplest kinds of dialogue systems
- Often have an entertainment value
- Can carry on extended conversation
- Several types
 - Rule-based (e.g., ELIZA, PARRY)
 - Corpus-based (e.g., Xiaolce)
 - Hybrid

Rule-based chatbot - ELIZA

- Most important chatbot dialogue system in the history of the field
- Shows we do not need to know anything about the real world to have a conversation
- Designed to simulate a Rogerian psychologist
 - Methods involve drawing the patient out by reflecting patient's statements back at them
- Most chatbots trying to pass the Turing test choose a domain with similar properties

Rule-based chatbot - ELIZA

Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

::

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

My father

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT AGGRESSIVE

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

Bullies.

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR
BOYFRIEND MADE YOU COME HERE

Corpus-based chatbots

- Mine conversations of human-human conversations
- Data-intensive, requiring hundreds of millions or even billions of words for training (e.g., excerpts of movie dialogues, use of crowdworkers to have conversations)
- Due to lack of large conversational corpora, many systems pre-train first on pseudo-conversations generated from Twitter, Reddit and other social media
- After deployment, chatbots can collect usage data
- Response can be done by
 - **retrieval** - using IR techniques to find a response in a corpus
 - **generation** - using a language model or encoder-decoder to generate the response

Response by retrieval

- The goal is to find a response $r \in C$ such that r is most similar to the query q (user's turn)
- Compute tf-idf models for C and q , choosing the r that has the highest tf-idf cosine with q

$$response(q, C) = \operatorname{argmax}_{r \in C} \frac{q \cdot r}{|q||r|}$$

- If the collection contains conversational turns, we find the most similar turn to the query and return the turn following it

Response by retrieval - cont'd

- Alternatively, we can use neural IR (e.g., Transformer-based encoder model)
- In the spirit of dense retrieval we train two separate encoders, one for the query and one for the responses and use the dot product as similarity

$$h_q = BERT_Q(q)[CLS]$$

$$h_r = BERT_R(r)[CLS]$$

$$response(q, C) = \operatorname{argmax}_{r \in C} h_q \cdot h_r$$

Response by generation

- It was proposed early on to think of response generation as a kind of translation
 - "Translate" an input query q into an response r
- Use of encoder-decoder model was natural
- The task is to generate each token r_t of the response by conditioning on the encoding of the entire query q and the response so far $r_1 \dots r_{t-1}$

$$r_t = \operatorname{argmax}_{w \in V} P(w|q, r_1 \dots r_{t-1})$$

- It is common practice to include a longer context, forming the query not just from the user's turn but from the entire conversation so far

Response by generation - cont'd

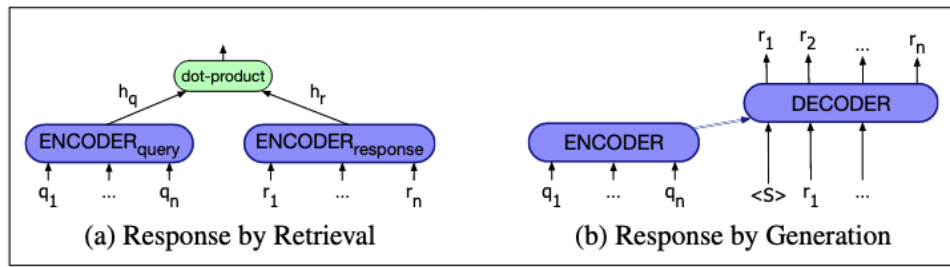


Figure: Two architectures for generating responses for a neural chatbot. In response by retrieval (a) we choose a response by finding the turn in the corpus whose encoding has the highest dot-product with the user's turn. In response by generation (b) we use an encoder- decoder to generate the response.

Response by retrieving and refining knowledge

- Responses are a combination of retrieval and generation
 - First use IR to retrieve potentially useful passages from Wikipedia (retrieval)
 - Create multiple candidates by concatenating each retrieved Wikipedia sentence to the dialogue context with a separator token
 - Each candidate can be given as the encoder context to the encoder-decoder model (generation)

Architecture

- Chatbots
- Task-oriented Dialogue Systems

Task-oriented Dialogue Systems

- Task-oriented dialogue agents use conversation with users to help complete task
- They can answer questions, give directions, control appliances, find restaurants, or make calls
- Two most common architectures are
 - Simple Frame-based Dialogue Systems
 - The Dialogue-State Architecture

Simple Frame-based Dialogue Systems

- Representative of the architecture is a very simple GUS system for travel planning introduced in 1977
- Even though its old, it underlies most modern commercial digital assistants
- A frame is a kind of knowledge frame structure representing the kinds of intentions the system can extract from user sentences
- A frame consists of a collection of **slots**, each of which can take a set of possible **values**
 - The set of slots in a frame specifies what the system needs to know
 - The filler of each slot is constrained to values of a particular semantic type

Simple Frame-based Dialogue Systems

GUS: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?

Client: I want to go to San Diego on May 28

GUS: What time do you want to leave?

Client: I must be in San Diego before 10 am

GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am ?

Client: What is the next flight ?

GUS: Air California flight 310 that leaves at 8.30am and arrives at 10.00am

Client: I'll take the first one

GUS: In what name should I make a reservation, please?

Client: Dan Bobrow

GUS: I have confirmed the following flight: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am.

GUS: What date do you want to return on ?

Client: On Friday in the evening.

GUS: Would you like the flight that leaves at 7.45 pm ?

Client: That's fine.

GUS: I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm Thank you for calling. Goodbye

Figure: The travel domain: A transcript of an actual dialogue with the GUS system.

Simple Frame-based Dialogue Systems

- Types can have hierarchical structure (e.g., *date* in GUS is a frame with slots like *month*, *year*, *day*, *weekday*)

Slot	Type	Question Template
ORIGIN CITY	city	"From what city are you leaving?"
DESTINATION CITY	city	"Where are you going?"
DEPARTURE TIME	time	"When would you like to leave?"
DEPARTURE DATE	date	"What day would you like to leave?"
ARRIVAL TIME	time	"When do you want to arrive?"
ARRIVAL DATE	date	"What day would you like to arrive?"

Figure: A frame in a frame-based dialogue system, showing the type of each slot and a question used to fill the slot.

Determining Domain, Intent, and Slot fillers

- In each turn of the conversation the system needs to extract 3 things from a user utterance
- The first task is domain classification
 - Unnecessary for single-domain systems
- The second is user intent determination
 - What general task or goal is the user trying to accomplish?
- The third is slot filling
 - Extract the particular slots and fillers that the user intends the system to understand from their utterance with respect to their intent
 - Method used in the original GUS system, and still quite common in industrial applications, is to use handwritten rules

Control structure for frame-based dialogue

- The system's goal is twofold
 - Fill the slots in the frame with the fillers the user intends
 - Perform the relevant action for the user (e.g., answering a question, or booking a flight)
- First goal is achieved with the system asking questions of the user (using pre-specified question templates)
- Once the system has enough information it performs the necessary action (e.g., querying a database of flights)

Other components of frame-based dialogue

- The ASR (automatic speech recognition) component takes audio input outputs a transcribed string of words
- The NLU (natural language generation) module of any dialogue system produces the utterances that the system says to the user
 - Generally, template-based generation (can be fixed or include variables)

The Dialogue-State Architecture

- More sophisticated version of the frame-based architecture
- The four main components include
 - Natural language understanding (NLU)
 - Responsible for determining domain, intent, and slot fillers
 - Generally using machine learning rather than rules
 - Dialogue State Tracking
 - Maintains the current state of the dialogue including the most recent **dialogue act** and the entire set of slot-filler constraints the user has expressed so far
 - Dialogue Policy
 - decides what the system should do or say next
 - More advanced than in GUS where the policy was to ask questions until the frame is full and report back the results
 - Natural language generation (NLG)
 - Rather than using templates, condition on the exact context to produce turns that seem much more natural

The Dialogue-State Architecture

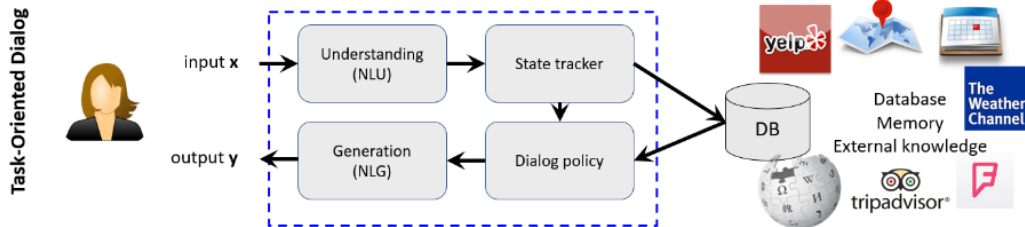


Figure: Architecture of a dialogue-state system for task-oriented dialogue³

³J. Gao, M. Galley, and L. Li, "Neural Approaches to Conversational AI." arXiv, Sep. 10, 2019.

Dialogue Acts

- Dialogue acts represent the interactive function of the turn or sentence
- Different types of dialogue systems require labeling different kinds of acts

Tag	Sys	User	Description
HELLO($a = x, b = y, \dots$)	✓	✓	Open a dialogue and give info $a = x, b = y, \dots$
INFORM($a = x, b = y, \dots$)	✓	✓	Give info $a = x, b = y, \dots$
REQUEST($a, b = x, \dots$)	✓	✓	Request value for a given $b = x, \dots$
REQALTS($a = x, \dots$)	✗	✓	Request alternative with $a = x, \dots$
CONFIRM($a = x, b = y, \dots$)	✓	✓	Explicitly confirm $a = x, b = y, \dots$
CONFREQ($a = x, \dots, d$)	✓	✗	Implicitly confirm $a = x, \dots$ and request value of d
SELECT($a = x, a = y$)	✓	✗	Implicitly confirm $a = x, \dots$ and request value of d
AFFIRM($a = x, b = y, \dots$)	✓	✓	Affirm and give further info $a = x, b = y, \dots$
NEGATE($a = x$)	✗	✓	Negate and give corrected value $a = x$
DENY($a = x$)	✗	✓	Deny that $a = x$
BYE()	✓	✓	Close a dialogue

Figure: Dialogue acts used by the HIS restaurant recommendation system.^a

^aYoung, S. J. et al, 2010. The Hidden Information State model: A practical framework for POMDP-based spoken dialogue management. Computer Speech & Language

Slot Filling

- Use neural model to jointly associate each sentence with the correct set of slots, domain, and intent
- Classify each token to a predefined BIO label (beginning (B), inside (I), and outside (O))
- There are $2n + 1$ tags, where n is the number of slots

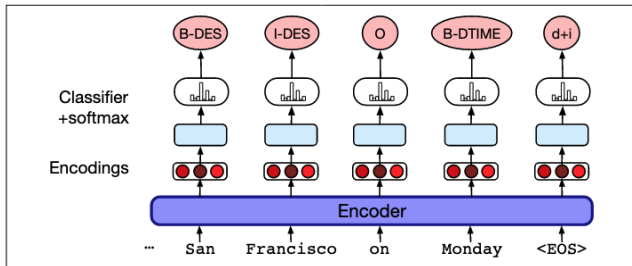


Figure: A simple architecture for slot filling

Dialogue State Tracking

- Job is to determine both the current state of the frame and the user's most recent dialogue act
- Contains the value of the frame since the beginning of the conversation
- Simple solution
 - take the output of a slot-filling sequence-model after each sentence (NLU)
- More complex model
 - Train a classifier for each slot to decide whether its value is being changed in the current sentence or should be carried over from the previous sentence

Dialogue Policy

- Decide what action the system should take next, i.e., what dialogue act to generate
- At turn i in the conversation predict which action A_i to take, based on the entire dialogue state. Given the sequence of dialogue acts from the system (A) and from the user (U) we want to compute

$$\hat{A}_i = \operatorname{argmax}_{A_i \in A} P(A_i | (A_1, U_1, \dots, A_{i-1}, U_{i-1}))$$

- Or alternatively, collapsing across the many different conversational paths that could lead to the same set of filled slots

$$\hat{A}_i = \operatorname{argmax}_{A_i \in A} P(A_i | \text{Frame}_{i-1}, A_{i-1}, U_{i-1})$$

- Estimate the probabilities using neural models

Policy Example: Confirmation and Rejection

- Modern dialogue systems often make mistake
- We can use confirmation and rejection dialogue acts to make sure we have understood the user correctly
- Confirmation is used to confirm the understanding with the user
 - can be **explicit** or **implicit**
 - Explicit and implicit confirmation have complementary strengths
- rejection is when a system expresses the lack of understanding (e.g., *I'm sorry, I didn't understand that*)

Explicit Understanding

- A system asks the user a direct question to confirm the system's understanding
- Makes it easier for users to correct the system's misrecognitions
- It is awkward and increases the length of the conversation

S: Which city do you want to leave from?

U: Baltimore.

S: **Do you want to leave from Baltimore?**

U: Yes.

U: I'd like to fly from Denver Colorado to New York City on September twenty first in the morning on United Airlines

S: **Let's see then. I have you going from Denver Colorado to New York on September twenty first. Is that correct?**

U: Yes

Implicit Understanding

- System can demonstrate its understanding as a grounding strategy repeating back the system's understanding as part of asking the next question
- It is much more conversationally natural

U: I want to travel to Berlin

S: **When do you want to travel to Berlin?**

U2: Hi I'd like to fly to Seattle Tuesday Morning

A3: **Traveling to Seattle on Tuesday, August eleventh in the morning.**
Your full name?

Natural language generation (NLG)

- Often modeled in two stages
 - content planning (what to say)
 - sentence realization (how to say it)
- Assume that content planning was done by dialogue policy (By selecting a dialogue act and a set of slots and values)
- Training data is hard to come by, increase generality by **delexicalization**

<code>recommend(restaurant name= Au Midi, neighborhood = midtown, cuisine = french</code>
1 Au Midi is in Midtown and serves French food.
2 There is a French restaurant in Midtown called Au Midi.
<code>recommend(restaurant name= Loch Fyne, neighborhood = city centre, cuisine = seafood)</code>
3 Loch Fyne is in the City Center and serves seafood food.
4 There is a seafood restaurant in the City Centre called Loch Fyne.

Figure: Two examples of inputs to the sentence realization phase of NLG

Natural language generation (NLG) - Delexicalization

- The process of replacing specific words in the training set that represent slot values with a generic placeholder token representing the slot
- Mapping from frames to delexicalized sentences is generally done by encoder decoder model
- The encoder reads all the input slot/value representations, and the decoder outputs the following delexicalized English sentence
- Use the input frame from the content planner to relexicalize

```
recommend(restaurant name= Au Midi, neighborhood = midtown,  
cuisine = french
```

```
1 restaurant_name is in neighborhood and serves cuisine food.
```

```
2 There is a cuisine restaurant in neighborhood called restaurant_name.
```

Natural language generation (NLG) - Generating Clarification Questions

- Example of NLG algorithm specific to a particular dialogue act
- Instead of using dialogue act *reject("Please repeat")* it is better to use targeted clarification questions

User: What do you have going to UNKNOWN_WORD on the 5th?

System: Going where on the 5th?

- Targeted clarification questions can be created by rules or by building classifiers to guess which slots might have been misrecognized in the sentence

Examples

Conversational Recommender Systems

Definition

A recommendation system that can elicit the dynamic preferences of users and take actions based on their current needs through real-time multiturn interactions.

Main challenges in CRS

- Question-based user preference elicitation
 - What to ask?
 - How to adjust the recommendations based on user response?
- Multi-turn Conversational Recommendation Strategies
 - When to ask questions and when to make recommendations?
- Natural Language Understanding and Generation
 - Communicating like a human continues to be one of the hardest challenges
- Trade-offs between Exploration and Exploitation
 - Users only have limited time and energy to interact with the system, a failed exploration will waste time and lose the opportunity to make accurate recommendations.
- Evaluation and User Simulation
 - Emphasis on the user experience during dynamic interactions makes online evaluation preferred but very expensive

Examples

- Soliciting User Preferences in Conversational Recommender Systems via Usage-related Questions

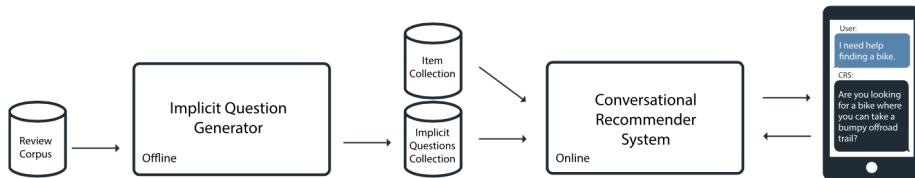
- IAI MovieBot

- DAGFiNN

Usage-related Questions - Motivation

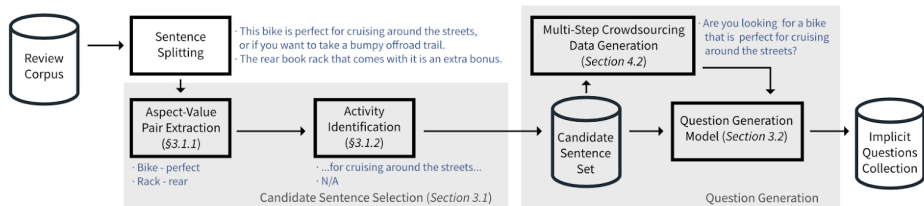
- Examples of traditional preference elicitation:
 - How thick should the tires be?
 - How many gears should the bike have?

- Usage based preference elicitation:
 - Will you be using this bike mostly for commuting?



Usage-related Questions - Approach

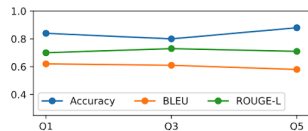
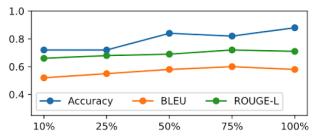
- Extract usage related sentences from a review corpus.
- Use crowdsourcing to generate labeled data.
- Train a text-to-text model (T5) to generate questions.



Usage-related Questions - Results

Model	#Parameters	Accuracy	BLEU-4	ROUGE-L
T5-small	60.5 M	0.76	0.56	0.69
T5-base	222 M	0.84	0.53	0.66
T5-large	737 M	0.88	0.58	0.71

Fig. 3. Model performance (T5-large) with sentence-based (Left) or question-based (Right) training data reduction.



Examples

- Soliciting User Preferences in Conversational Recommender Systems via Usage-related Questions

- IAI MovieBot

- DAGFiNN

IAI MovieBot

- Open-source conversational movie recommender system
- Models users' preferences dynamically
- Supports multi-turn recommendations
- Typical dialogue-state architecture
- Multi-modal chat interface
- Student project

Discussion

Question

If you had to design a movie recommender system, what intents would you choose?

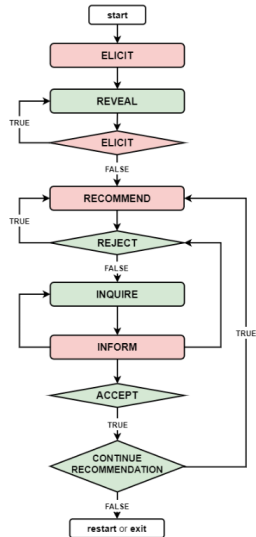
IAI MovieBot - Overview of intents

	Intent	Description
	User Revealment [7] / Query Formulation [9] / Query Reformulation [9]	
User	Reveal	The user wants to reveal a preference. <i>"Do you have any sports movies?"</i>
	Remove preference	The user wants to remove any previously stated preference. <i>"I don't want to see any sports movies anymore."</i>
Agent	Elicit	Ask the user to describe their preferences. <i>"Which genres do you prefer?"</i>
	System Revealment [7] / Result Exploration [9]	
User	Inquire	Once the agent has recommended an item, the user can ask further details about it. <i>"Please tell me more about this movie."</i>
	Accept/Reject	The user can decide if they like the recommendation or not.
	Accept	The user accepts (likes) the recommendation. This will determine the success of the system as being able to find a recommendation the user liked. <i>"I like this recommendation."</i>
	Reject	The user either has already seen/consumed the recommended item or does not like it. <i>"I have already seen this one."</i>
	Continue recommendation	If the user likes a recommendation, they can either restart, quit or continue the process to get a similar recommendation. <i>"I would like a similar recommendation."</i>

IAI MovieBot - Overview of intents

Agent	Reveal	Reveal the results or the number of matching results to the user.
	Too many results	The number of items matching the user's preferences is larger than a maximum limit. This will be followed by an <i>elicit</i> intent. <i>"There are almost 1100 action movies."</i>
	Recommend	Based on the user's preferences, make a recommendation. <i>"I would like to recommend a fairy tale film, named Shrek."</i>
	No Results	The database does not contain any items matching the user's preferences. <i>"Sorry. I couldn't find any romantic Korean movies."</i>
	Inform	If the user <i>inquires</i> about the recommended item, the agent provides the relevant information. <i>"The director of this movie is XYZ."</i>
Miscellaneous Intents		
User	Hi	When the user initiates the conversation, they start with a formal <i>hi/hello</i> or <i>reveal</i> preferences.
	Acknowledge	Acknowledge the agent's question where required.
	Deny	Negate the agent's question where required.
	Bye	End the conversation by sending a bye message or an exit command.
Agent	Welcome	Start the conversation by giving a short introduction.
	Acknowledge	Acknowledge the user's query where required.
	Cant Help	The agent does not understand the user's query or is not able to respond properly based on the current dialogue state.
	Bye	End the conversation.

IAI MovieBot - Dialog Flow



Examples

- Soliciting User Preferences in Conversational Recommender Systems via Usage-related Questions
- IAI MovieBot
- DAGFiNN

DAGFiNN - Motivation

- Create an open-source, multi-domain, multi-modal conversational recommender system
- Provide a personalized and engaging user experience
- Allow users to ask a broad range of questions that naturally arise before and during the conference.
 - There are currently only a few open-source CRS prototypes available
 - The vast majority of research uses text as the sole modality and targets a single domain

DAGFiNN - Conversational Conference Assistant



Figure: Furhat robot.

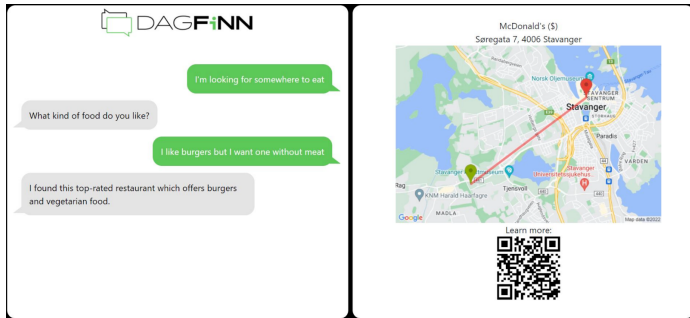


Figure: External screen with a rich response example.

DAGFiNN - Conversational Conference Assistant

- Provide recommendations across multiple domains, have basic conversational capabilities, and personalization integrated into a single system.
 - Currently, recommenders are built for point-of-interest (POI) and conference content
- Support multiple user interfaces (UI)
 - A Web chat and a Furhat robot with an optional external screen
 - Use of rich responses customized to each user interface to effectively convey information.
- Utilize multiple input modalities
 - Furhat is a social robot with a microphone, speaker, and camera, which makes it possible to recognize users that interacted with the robot previously.

DAGFiNN - Architecture

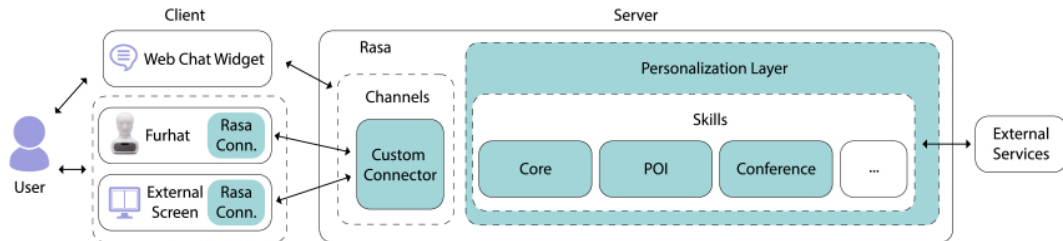


Figure: DAGFiNN architecture. Left: Clients with different interfaces; Web chat and Furhat robot with external screen. Right: Server.

DAGFiNN - Example questions

- What is there to see in Stavanger?
- Can you recommend me some bars?
- Do you know of any good Italian restaurants?
- Is there somewhere cheaper?
- Where can I go running?
- Where can I go hiking?
- How do I get there?
- How do I get to the airport?

Exercise

DAGFiNN Demo

Summary

- Conversational AI
 - Conversational information access
- Architecture
 - Chatbots
 - Rule-based chatbot
 - Corpus-based chatbots
 - Task-oriented Dialogue Systems
 - Simple Frame-based Dialogue Systems
 - The Dialogue-State Architecture
 - Evaluating Dialogue Systems
 - Evaluating Chatbots
 - Evaluating Task-Based Dialogue
 - Dialogue System Design
- Examples
 - Soliciting User Preferences in Conversational Recommender Systems via Usage-related Questions
 - IAI MovieBot
 - DAGFiNN

Reading

- *Chatbots and Dialogue Systems*, Daniel Jurafsky and James H. Martin ⁴

⁴<https://web.stanford.edu/~jurafsky/slp3/>