RUTGERS

The ACM Conference Series on Recommender Systems

UNIVERSITY OF CALIFORNIA SANTA CRUZ

Tutorial on Conversational Recommendation Systems

Zuohui Fu
Computer Science
Rutgers University
zuohui.fu@rutgers.edu

Yikun Xian
Computer Science
Rutgers University
yikun.xian@rutgers.edu

Yongfeng Zhang
Computer Science
Rutgers University
yongfeng.zhang@rutgers.edu

Yi Zhang
School of Engineering
UC Santa Cruz
yiz@soe.ucsc.edu

Outline

- Introduction and Background
- Problem Formalization
- Datasets and Evaluation
- Conversational Recommendation Methods
- Tool-kits and Real-world Systems

Introduction and Background

- **Introduction and Background**
  - Introduction to Conversational Recommendation
  - A Brief History of Conversational Recommendation Research
  - The New Boom in Conversational Recommendation Research
- Problem Formalization
- Datasets and Evaluation
- Conversational Recommendation Methods
- Tool-kits and Real-world Systems

Introduction to Conversational Recommendation

- Conversational Recommendation
  - Conversational recommendation provides personalized recommendations through natural language dialog with users

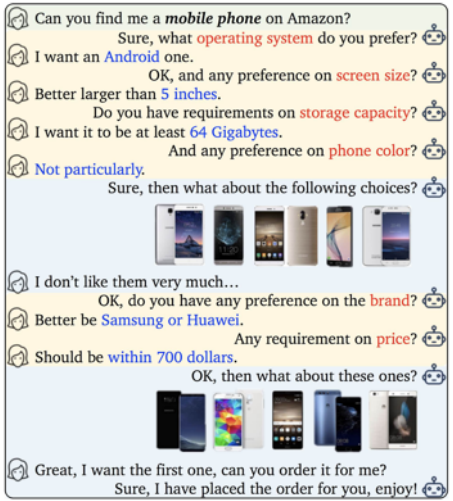


Image from Zhang et al. Towards conversational search and recommendation: System ask, user respond. CIKM 2018.

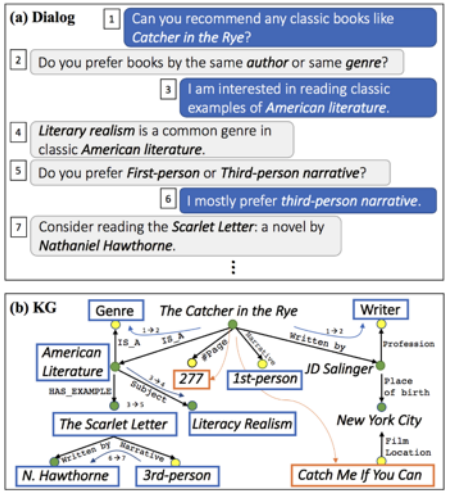
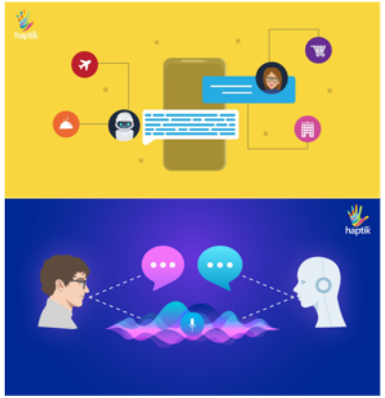


Image from Moon, Seungwhan, et al. Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs. ACL. 2019.

Image from Zhang et al. Liu, Zeming, et al. Towards Conversational Recommendation over Multi-Type Dialogs. ACL. 2020.

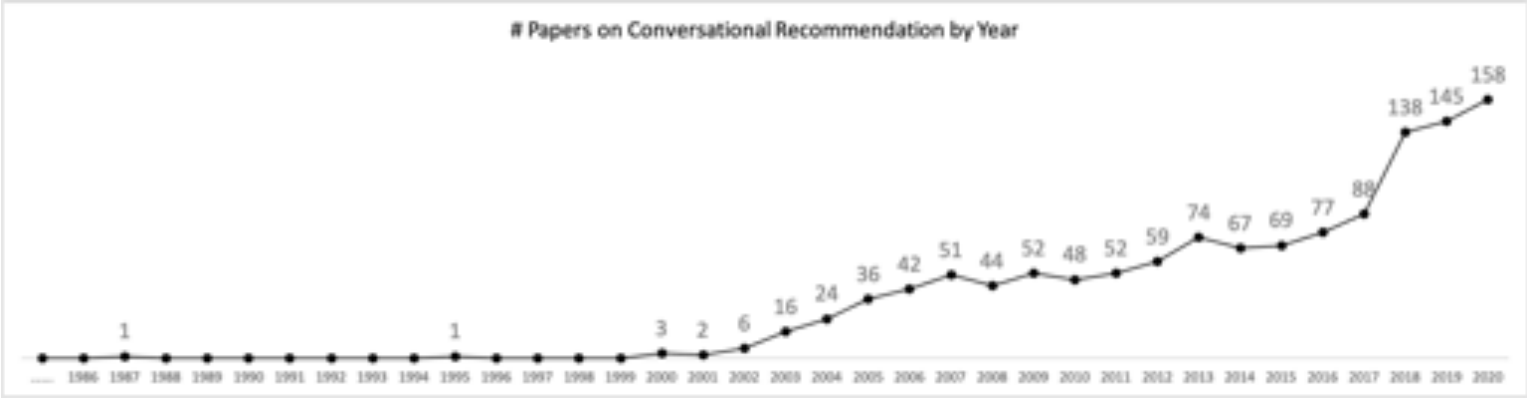


Dialog can be carried out using either visual or spoken interfaces.

Image credit: haptik.ai

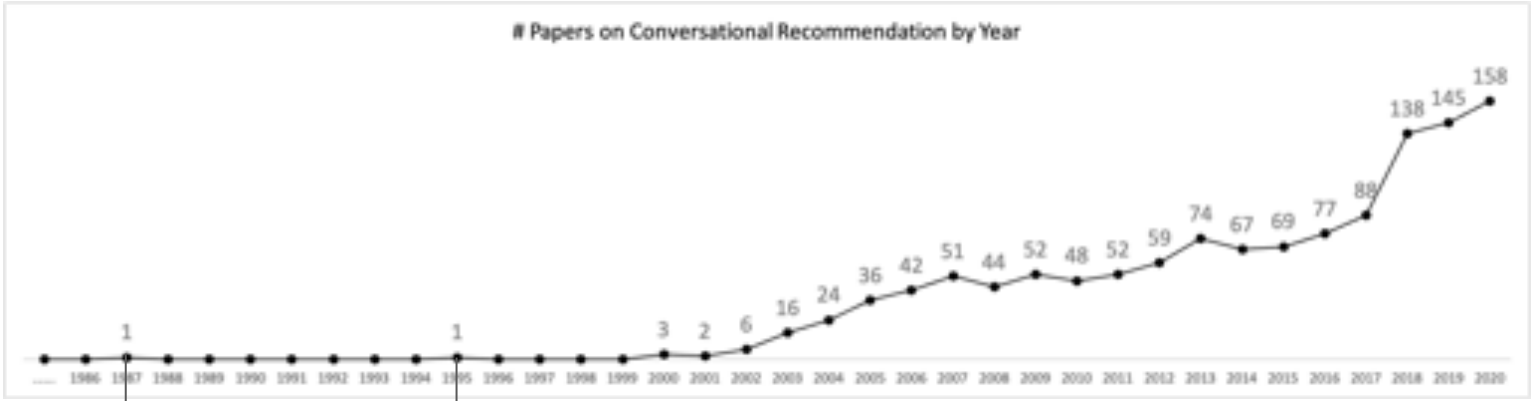
A Brief History of Conversational Recommendation

- Conversational Recommendation
  - Research on conversational recommendation has been emerging in recent years
  - The basic concept dates back to many early reseach in IR, RecSys, HCI, etc.



Papers in Google Scholar using query ("conversational recommendation" OR "conversational recommender").
May not represent all papers in this direction since many papers on the related topic may not include these exact words.

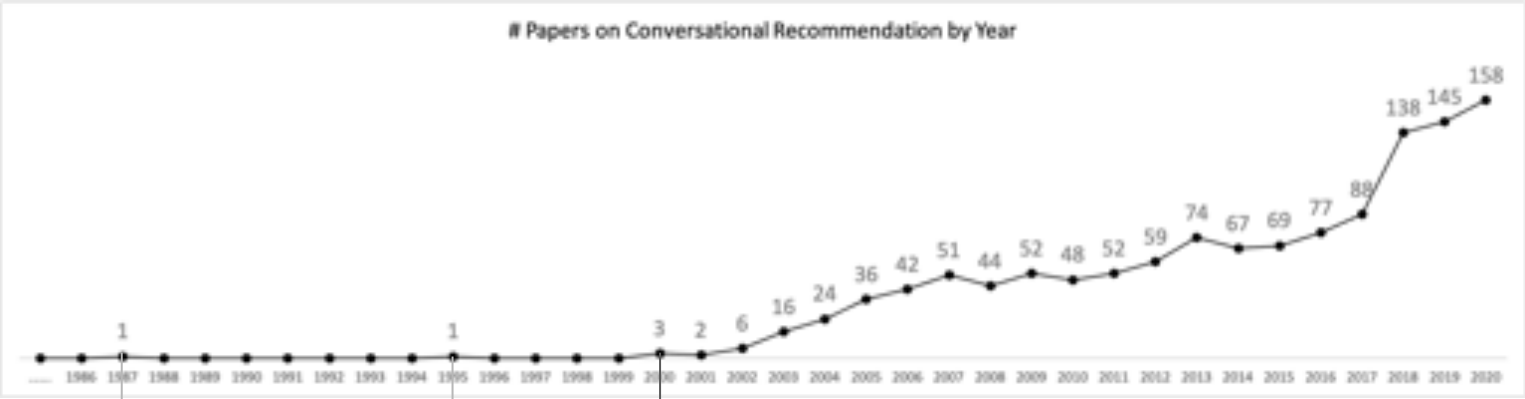
A Brief History of Conversational Recommendation



Croft and Thompson. I3R: A New Approach to the Design of Document Retrieval Systems. JASIST 1987.
User-system interaction through dialog for information seeking.

Belkin et al. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. 1995.
MERIT: An interactive information seeking system using script-based conversational interactions.

A Brief History of Conversational Recommendation



Croft and Thompson. 13R: A New Approach to the Design of Document Retrieval Systems. JASIST 1987.
User-system interaction through dialog for information seeking.

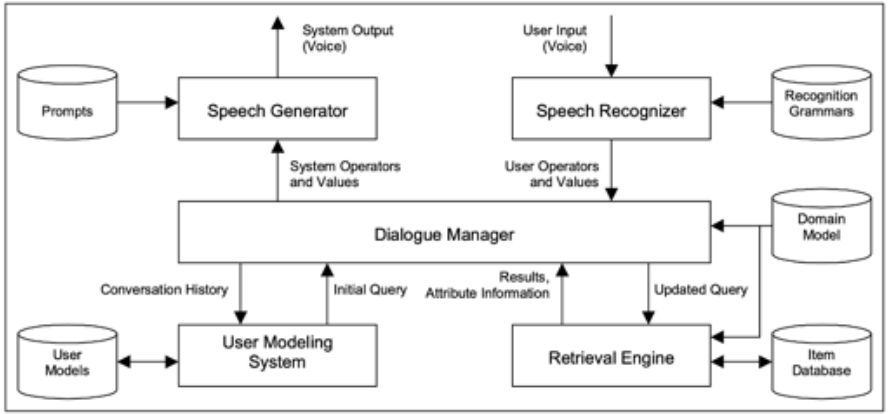
Belkin et al. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. 1995.
MERIT: An interactive information seeking system using script-based conversational interactions.

*Goker and Thompson. The adaptive place advisor: A **conversational recommendation system**. 2000.*
A conversational recommendation system for place recommendation.
The term “Conversational Recommendation” is formally introduced.

A Brief History of Conversational Recommendation

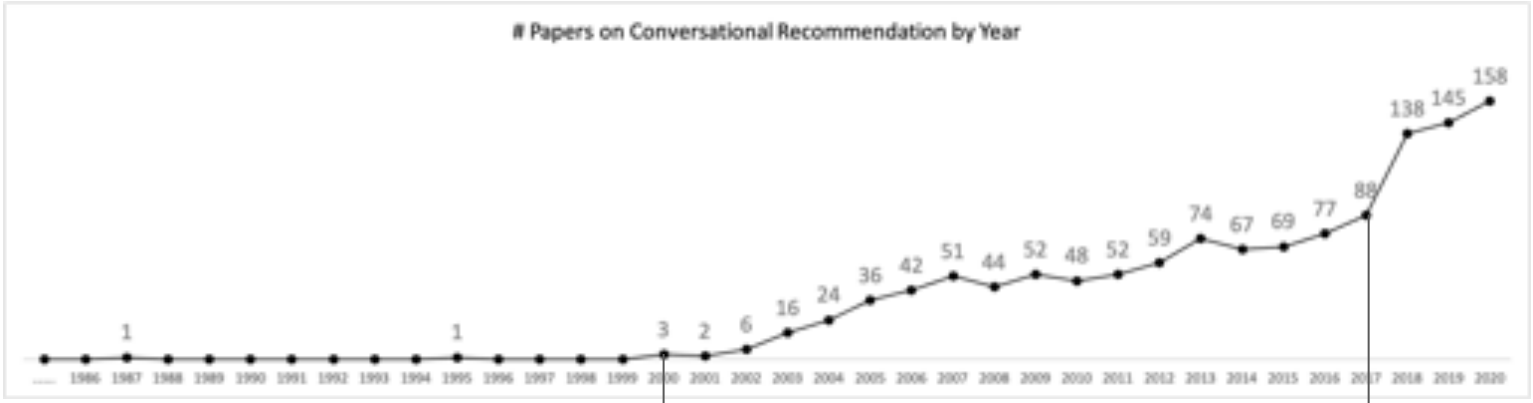
Goker and Thompson. The adaptive place advisor: A ***conversational recommendation system****. 2000.*
A conversational recommendation system for place recommendation.

Inquirer:	Where do you think I should eat tonight?
Advisor:	Well, what type of cuisine would you like?
Inquirer:	What types are there?
Advisor:	Some examples are Chinese, Indian, and Mediterranean.
Inquirer:	Oh, maybe Indian.
Advisor:	What quality rating would you like?
Inquirer:	I don't care, as long as it's cheap.
Advisor:	How do you want to pay?
Inquirer:	No, I think I'd like Chinese instead.
Advisor:	Okay, we'll switch to Chinese food. What city do you prefer?
Inquirer:	How about Berkeley?
Advisor:	I know three cheap Chinese restaurants in Berkeley. One is the Long Life Vegi House on 2129 University Avenue. Does that sound alright?
Inquirer:	Sure, that sounds fine.



The expected conversation and model architecture have been very close to "modern" conversational recommendation systems. Imagine 20 years ago with no deep learning and even shallow factorization model is yet to become mature.

A Brief History of Conversational Recommendation

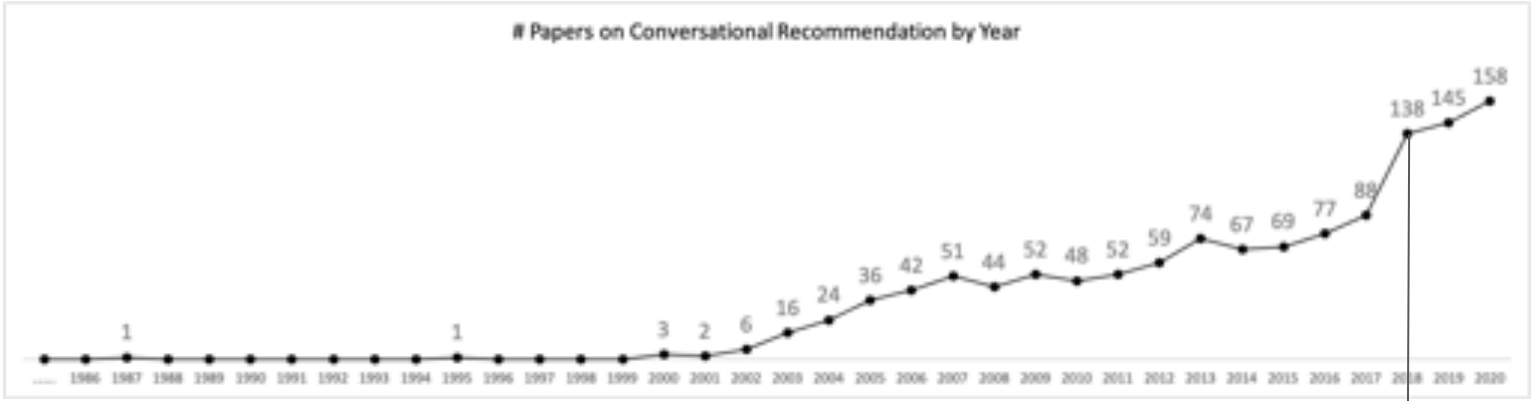


Stable but slow growth in the domain

Some (not all) representative works in the era:

- Bridge, D.G. *Towards Conversational Recommender Systems: A Dialogue Grammar Approach*. ECCBR workshops. 2002.
  - Begin to consider the **natural language** problems in conversational recommendation.
- McGinty and Smyth. *On the role of diversity in conversational recommender systems*. Inter. Conf. on Case-Based Reasoning. 2003
  - Considers the **diversity** in conversational recommendation
- Smyth, McGinty, Reilly and McCarthy. *Compound critiques for conversational recommender systems*. In Web Intelligence. 2004.
  - **Critique-based** conversational recommendation
- Wärnestål. *User evaluation of a conversational recommender system*. Knowledge and Reasoning in Practical Dialogue Systems. 2005.
  - Considers the **evaluation** of conversational recommendation systems.
- Mahmood, Mujtaba and Venturini. *Dynamic personalization in conversational recommender systems*. Info. Sys. and e-Business. 2014.
  - Considers **dynamics** of user preference.
- Christakopoulou, Radlinski and Hofmann. *Towards conversational recommender systems*. KDD. 2016.
  - Considers conversational recommendation for **cold-start** scenarios.

A Brief History of Conversational Recommendation



A new boom in this direction appeared in 2018, with deep learning and reinforcement learning as the key methods.

Some (not all) representative works in the era:

- Sun and Zhang. *Conversational recommender system*. SIGIR. 2018.
  - Adopted **reinforcemenat learning** for dialog state management.
- Zhang, Chen, Ai, Yang and Croft. *Towards conversational search and recommendation: System ask, user respond*. CIKM 2018.
  - Adopted **deep learning** for dialog understanding and response generation.
  - Provided a “System Ask - User Respond” (SAUR) paradigm for conversational recommendation

The New Boom in Conversational Recommendation

- Why?
- Technically:
  - Conversational AI is one of the **closest task to** Turing Test and Strong AI
  - Conversational Recommendation is a **well-defined** Conversational AI task
  - A good scenario to test **how far have we gone** on the road of AI with current techniques
  - Also help to **refine** the current techniques and to **develop** new techniques
- Commercially:
  - Recent advancement on deep learning and reinforcement learning has made conversational recommendation **usable** (though not perfect) for end users
  - Thus more and more commercial conversational recommender systems are **being used**
  - This helps to accumulate more and more **conversational data** and **usage pattern**
  - Which further helps to **refine** the conversational recommendation models
- A **virtuous cycle** between user and system, both commercially and technically

Problem Formalization

- Introduction and Background
- **Problem Formalization**
  - Basic Problem Formalization
  - Paradigm 1: System is Active, User is Passive (SAUP)
  - Paradigm 2: System is Active, User Engages (SAUE)
  - Paradigm 3: System is Active, User is Active (SAUA)
  - Challenges
- Datasets and Evaluation
- Conversational Recommendation Methods
- Tool-kits and Real-world Systems

Overview of Conversational AI

Conversational AI System


Conversational RecSys


Conversational Search

Conversational QA


Social Chatbot


Voice Commanding

- CRS vs Conventional Recommendation
  - Common: Help users find relevant information
  - Difference: Multiple-round vs single-round interaction
- CRS vs Conversational Search
  - Common: Rank relevant items via multi-turn dialogue
  - Difference: User modeling vs query understanding
- CRS vs Conversational Question Answering
  - Common: Multiple rounds of user-system interaction thorugh questions
  - Difference: Recommendation as the goal vs answering question as the goal
- CRS vs Social Chatbot
  - Common: Multiple rounds of conversational interactions
  - Difference: Task-oriented (short dialogue) vs chitchat (prolong dialogue)
- CRS vs Voice Commanding
  - Common: Both are goal-oriented conversation
  - Difference: Recommendation as the goal vs implementing user command as the goal (e.g., turn on the light, play a song)

Basic Problem Formalization

Conversational Recommendation System (CRS)

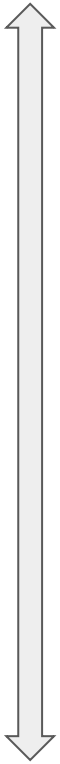
Input:

- Dialogue **history**: last *n* utterances
- (optional) User preferences
- (optional) External knowledge of items

Output:

- Next **utterance** to interact with user (in each turn)
- Item(s) recommended to user (once or multiple times)
- (optional) Explanations

Different Types of Utterances in Conversations



System Initiative

- System is Active, User is Passive (SAUP)
  - System leads the conversation by asking questions to user
  - User can only respond to the questions directly

Mixed Initiative

- System is Active, User Engages (SAUE)
  - System asks questions and user responds to the questions
  - Both system and user also chit-chat. User may not directly answer questions
- System is Active, User is Active (SAUA)
  - Both system and user can lead the conversation by asking questions
  - Both system and user also chit-chat

User Initiative

- User is Active, System is Passive (Voice command, QA)
  - User drives the conversation by asking questions to system

Focuses of this tutorial

Paradigm 1: System is Active, User is Passive (SAUP)

Typical Form: System Ask User Respond (SAUR)

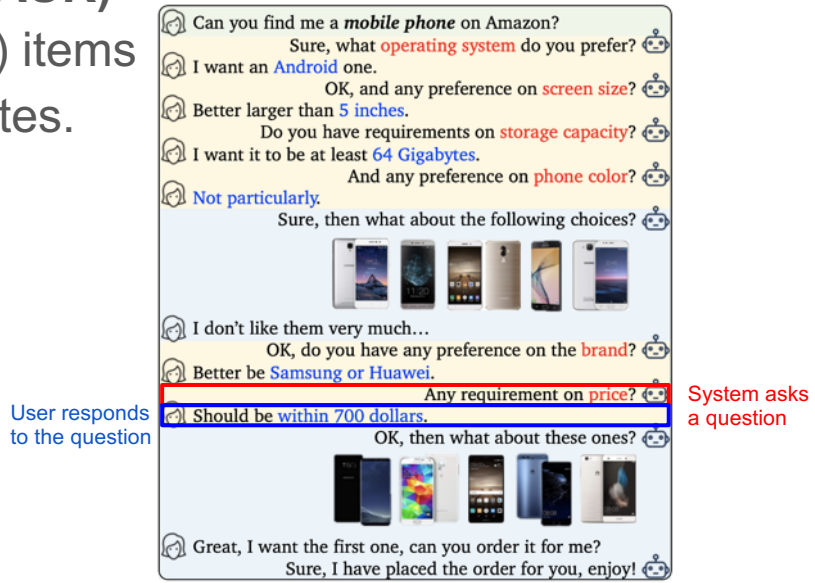
- System asks questions about (attributes of) items to narrow down the recommended candidates.

Input:

- Dialogue history

Output:

- Next utterance **= question to ask user**
- Item(s) recommended to user



Zhang, Yongfeng, et al. "Towards conversational search and recommendation: System ask, user respond." CIKM. 2018.

Paradigm 2: System is Active, User Engages (SAUE)

Typical Form: SAUR + Chitchat

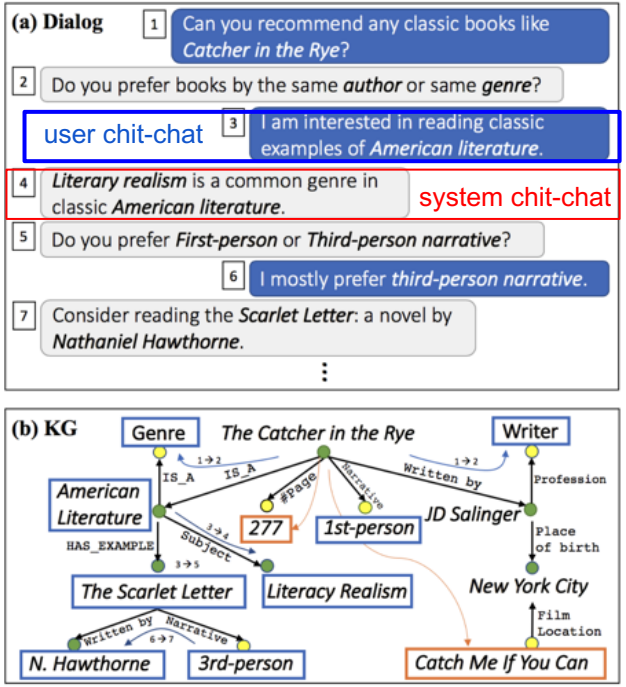
- System asks questions about (attributes of) items and also chats with user.

Input:

- Dialogue history

Output:

- Next utterance **= question + chitchat**
- Item(s) recommended to user



Moon, Seungwhan, et al. "Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs." ACL. 2019.

Paradigm 3: System is Active, User is Active (SAUA)

Typical Form: SAUR + Chitchat + User Ask System Respond

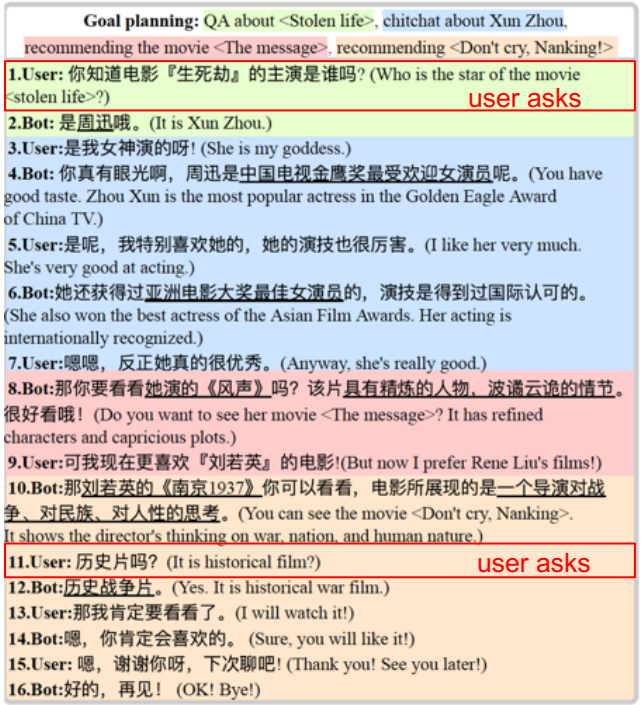
- Users actively ask questions and participate in the recommendation process.

Input:

- Dialogue history

Output:

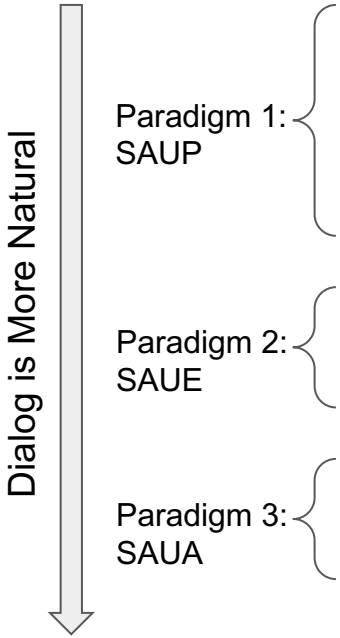
- Next utterance **= question or response to user's question**
- Item(s) recommended to user



Xu, Hu, et al. "User Memory Reasoning for Conversational Recommendation." arXiv preprint arXiv:2006.00184 (2020).

Related Work

Representative works categorized by 3 paradigms:



Paradigm 1: SAUP

- Sun, Yueming, et al. “Conversational Recommender System.” SIGIR. 2018.
- Zhang, Yongfeng, et al. "Towards conversational search and recommendation: System ask, user respond." CIKM. 2018.
- Christakopoulou, Konstantina, et al. "Q&R: A two-stage approach toward interactive recommendation." KDD. 2018.
- Zou, Jie, et al. "Towards Question-based Recommender Systems." SIGIR. 2020.
- Zhang, Jian-Guo, et al. "Find or classify? dual strategy for slot-value predictions on multi-domain dialog state tracking." SEM. 2020.
- Lei, Wenqiang, et al. "Estimation-action-reflection: Towards deep interaction between conversational and recommender systems." WSDM. 2020.
- Lei, Wenqiang, et al. "Interactive Path Reasoning on Graph for Conversational Recommendation." KDD. 2020.

Paradigm 2: SAUE

- Li, Raymond, et al. "Towards deep conversational recommendations." NeuIPS. 2018.
- Moon, Seungwhan, et al. "Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs." ACL. 2019.
- Chen, Qibin, et al. "Towards knowledge-based recommender dialog system." EMNLP. 2019.
- Chen, Zhongxia, et al. "Towards Explainable Conversational Recommendation." IJCAI, 2020.

Paradigm 3: SAUA

- Liu, Zeming, et al. "Towards Conversational Recommendation over Multi-Type Dialogs." ACL. 2020.
- Zhou, Kun, et al. "Improving Conversational Recommender Systems via Knowledge Graph based Semantic Fusion." KDD. 2020.
- Xu, Hu, et al. "User Memory Reasoning for Conversational Recommendation." arXiv. 2020.

Challenges

- Challenges in Conversation Recommendation
  - **How** to represent dialogue state?
  - **How** to represent dialogue action, both user action and system action?
  - **How** to understand user preference from dialogue and history behavior?
  - **When** to respond or recommend?
  - **What** to respond to user? E.g. What question to ask user?
  - **Which** item(s) to recommend?

Datasets and Evaluation

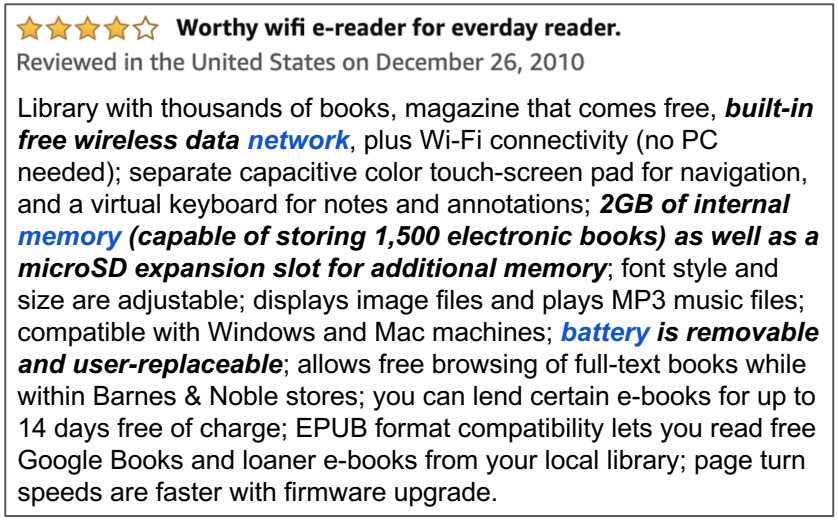
- Introduction and Background
- Problem Formulation
- **Datasets and Evaluation**
  - Frequently Used Datasets
  - Evaluation Protocol
- Conversational Recommendation Methods
- Tool-kits and Real-world Systems

Frequently Used Datasets

Dataset	Dialog Types	Domains	External Knowledge	Paradigm
ConvRec [10]	Rec	Restaurant	N	Paradigm 1: SAUP
SAUR [11]	Rec	E-commerce	N	Paradigm 1: SAUP
Cookie [24]	Rec	E-commerce	Y	Paradigm 1: SAUP
ReDail [13]	Rec, Chitchat	Movie	N	Paradigm 2: SAUE
OpendialKG [14]	Rec	Music, Sports	Y	Paradigm 2: SAUE
KBRD [15]	Rec	Movie	Y	Paradigm 2: SAUE
DuRecDial [21]	Rec, Chitchat, QA	Movie, Music, Restaurant, News, Weather	Y	Paradigm 3: SAUA
MGConvRex [23]	Rec, Chitchat, QA	Restaurant	Y	Paradigm 3: SAUA

Paradigm 1 (SAUP): SAUR Dataset

- Assumption: Each user review is converted to a conversation. Aspects that appear earlier in the review should appear earlier in the conversation.

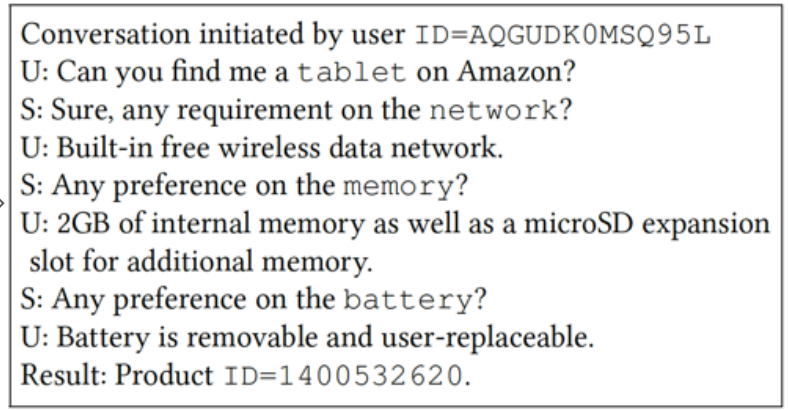


★★★★☆ **Worthy wifi e-reader for everday reader.**

Reviewed in the United States on December 26, 2010

Library with thousands of books, magazine that comes free, ***built-in free wireless data network***, plus Wi-Fi connectivity (no PC needed); separate capacitive color touch-screen pad for navigation, and a virtual keyboard for notes and annotations; ***2GB of internal memory (capable of storing 1,500 electronic books) as well as a microSD expansion slot for additional memory***; font style and size are adjustable; displays image files and plays MP3 music files; compatible with Windows and Mac machines; ***battery is removable and user-replaceable***; allows free browsing of full-text books while within Barnes & Noble stores; you can lend certain e-books for up to 14 days free of charge; EPUB format compatibility lets you read free Google Books and loaner e-books from your local library; page turn speeds are faster with firmware upgrade.

URL:https://www.amazon.com/gp/customer-reviews/R3RPNEBXXEXS7S/ref=cm_cr_getr_d_rvw_ttl?ie=UTF8&ASIN=1400532620



Conversation initiated by user `ID=AQGUDK0MSQ95L`
U: Can you find me a `tablet` on Amazon?
S: Sure, any requirement on the `network`?
U: Built-in free wireless data network.
S: Any preference on the `memory`?
U: 2GB of internal memory as well as a microSD expansion slot for additional memory.
S: Any preference on the `battery`?
U: Battery is removable and user-replaceable.
Result: Product `ID=1400532620`.

Zhang, Yongfeng, et al. "Towards conversational search and recommendation: System ask, user respond." CIKM. 2018.

Paradigm 1 (SAUP): SAUR Dataset

- Domain: E-commerce (Amazon)
- Statistics:

Dataset	#Users	#Items	#Reviews	#Aspect	#Value	#AV pairs	#Request	ℓ(Request)	Training/Testing	
									#Conversations	#Relevant Items per Conv
Electronics	142,421	53,278	365,341	479	500	475,020	989	6.40	255,739/109,602	1.21 ± 0.62/1.13 ± 0.26
CDs & Vinyl	64,847	60,405	427,031	514	747	659,737	694	5.71	298,922/128,109	2.82 ± 5.88/1.46 ± 1.26
Kindle Store	56,847	53,907	285,104	164	359	367,159	4,603	7.07	199,573/85,531	1.94 ± 3.63/1.62 ± 2.21
Cell Phones	21,615	9,292	52,178	325	402	68,709	165	5.93	36,525/15,653	1.66 ± 1.32/1.24 ± 0.16

Paradigm 1 (SAUP): ConvRec Dataset

- Assumption: Each user review is converted to a conversation. 5 item attributes are available as candidate facets to construct agent's questions.
- Domain: Restaurant (Yelp)
- Statistics:

	Number of Values
Users	62047
Items	21350
User-item pairs	875721
Category	191
State	13
City	189
Price	4
Rating Range	9

The target restaurant has the following facets.
{category: Mexican, state: AZ, city: Glendale,
price range: cheap, rating range: >=3.5}

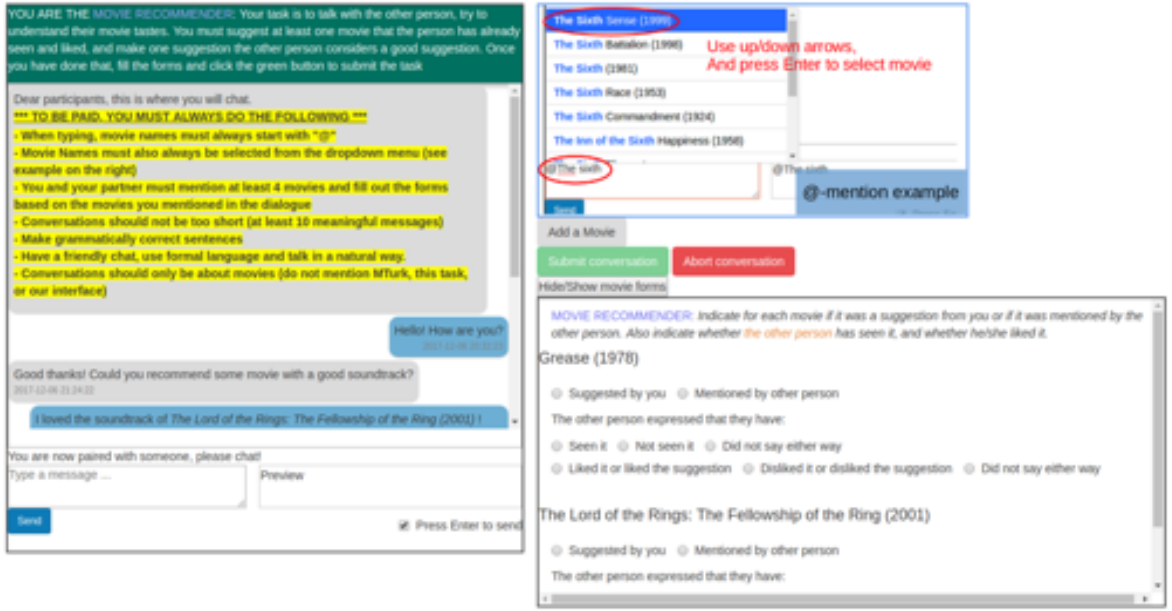
User: inform(city="Glendale", category="Mexican")
User Write: **I'm looking for Mexican food in Glendale.**
Agent: Which state are you in?
User: inform(state="AZ")
User Write: **I'm in Arizona.**
Agent: Which price range do you like?
User: inform(price_range="cheap")
User Write: **Low price.**
Agent: What rating range do you want?
User: inform(rating_range>="3.5")
User Write: **3.5 or higher**
Agent: <make recommendations>
User: thanks()
User Write: **thank you**

Sun, Yueming, et al. "Conversational Recommender System." SIGIR. 2018.

Paradigm 2 (SAUE): ReDail Dataset

- Assumption: Each dialogue contains at least 10 messages and 4 different movies, and the content is only about movies.
- Domain: Movie
- Statistics:

# conversations	10006
# utterances	182150
# users	956
# movie mentions	51699
seeker mentioned	16278
recommender suggested	35421
not seen	16516
seen	31694
did not say	3489
disliked (4.9%)	2556
liked (81%)	41998
did not say (14%)	7145



Li, Raymond, et al. "Towards deep conversational recommendations." NeuIPS. 2018.

Paradigm 3 (SAUA): DuRecDial Dataset

- Assumption: Seekers are assigned with predefined profiles, which should be consistent with seekers' utterances. Each dialogue follows the predefined task templates.
- Domain: Movie, Music, Restaurant, News, ...
- Statistics:

Knowledge graph	#Domains	7
	#Entities	21,837
	#Attributes	454
	#Triples	222,198
DuRecDial	#Dialogs	10,190
	#Sub-dialogs for QA/Rec/task/chitchat	6,722/8,756/3,234/10,190
	#Utterances	155,477
	#Seekers	1362
	#Entities recommended/accepted/rejected	11,162/8,692/2,470/

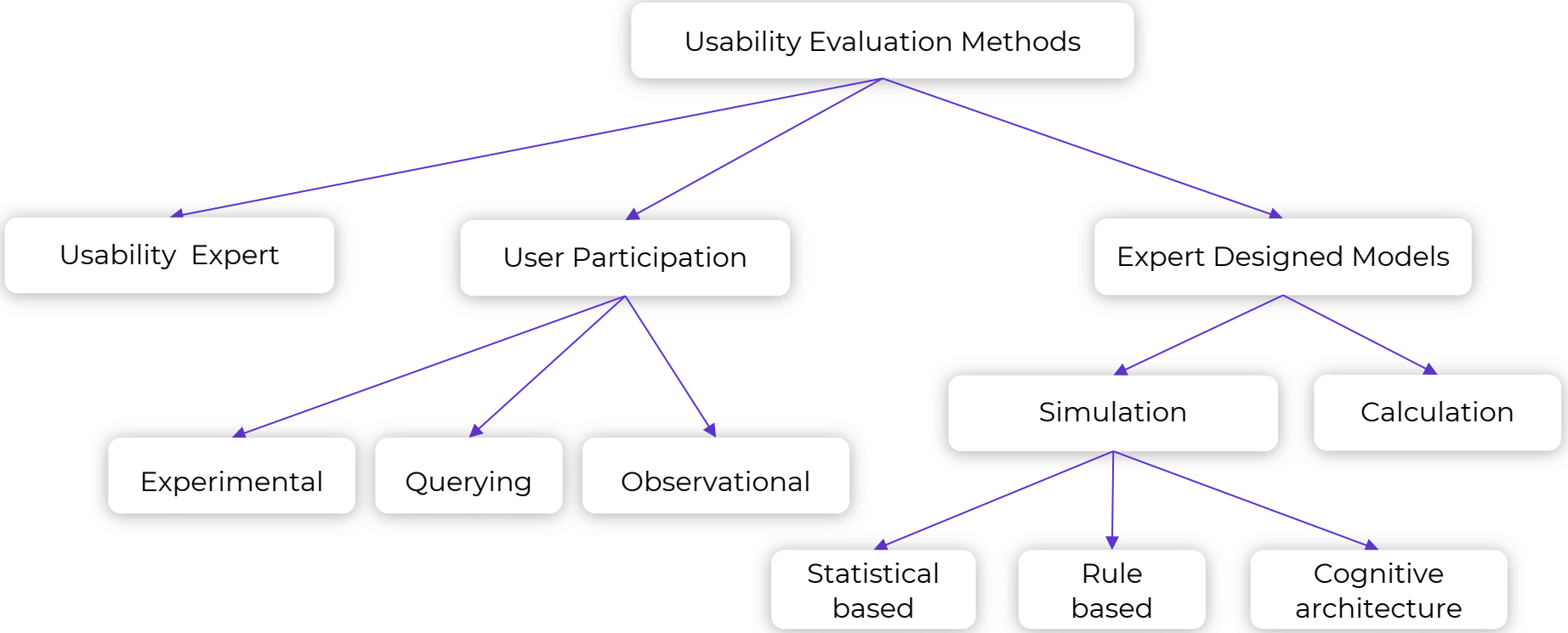
Goals	Goal description
Goal1: QA (dialog type) about the movie <Stolen life> (dialog topic)	The seeker takes the initiative, and asks for the information about the movie <Stolen life>; the recommender replies according to the given knowledge graph; finally the seeker provides feedback.
Goal2: chitchat about the movie star Xun Zhou	The recommender proactively changes the topic to movie star Xun Zhou as a short-term goal, and conducts an in-depth conversation;
Goal3: Recommendation of the movie <The message>	The recommender proactively changes the topic from movie star to related movie <The message>, and recommend it with movie comments, and the seeker changes the topic to Rene Liu's movies;
Goal4: Recommendation of the movie <Don't cry, Nanking!>	The recommender proactively recommends Rene Liu's movie <Don't cry, Nanking!> with movie comments. The seeker tries to ask questions about this movie, and the recommender should reply with related knowledge. Finally the user accepts the recommended movie.

Liu, Zeming, et al. "Towards Conversational Recommendation over Multi-Type Dialogs." ACL. 2020.

Evaluation Motivation

- Research Perspective: Understand and improve the virtual assistant **design** and AI **training**
  - Ensure AI solutions are as effective as it can be
  - Identify areas for improvement
- Business Perspective: Demonstrate the **impact** of success or **progress** of the project
  - Communicate the impact of the work to others
  - Improve team morale
  - Attract and retain supports from different stakeholders and decisions makers

Taxonomy of Approaches for Usability Evaluation of Interactive Systems



Common Evaluation Protocol

- Offline Evaluation
  - Evaluate on benchmark datasets
- Online Evaluation
  - Evaluate based on feedback from real users of the systems, e.g., A/B test
- User Study
  - Evaluate based on users' feedback on questionaires
- Simulation-based Evaluation
  - Evaluate based on simulated environments

Evaluation Metrics

- A complete evaluation of CRS should include at least two parts

Evaluation of Conversation Quality

Turn-level, Dialog-level, Business-level

Evaluation of Recommendation Quality

Turn-level, Dialog-level, Business-level

Joint Conversation-Recommendation Evaluation

Evaluation Metrics for Conversation Quality

- Turn-level Metrics
  - Quality of the system-generated sentences (e.g., BLEU, ROUGE, Readability)
  - Relevance of system-generated questions and/or responses (e.g., Accuracy, Coverage)
  - Frequencies and distributions of dialogue acts (e.g., recommend, ask question, respond)
  - User cooperativeness (e.g., provide responses when system asks a question)
  - Limitation: cannot measure the consistency of dialogue and conversion rate
- Dialogue-level Metrics
  - Dialogue length
  - Ratio of dialogue success and task completion
- Business-level Metrics
  - Conversion rate per session
  - Sales revenue
  - User satisfaction rating, user retention, customer loyalty, etc.

Evaluation Metrics for Recommendation Quality

- Turn-level Metrics
  - Recommendation accuracy per turn (e.g., Precision, Recall, NDCG)
  - Frequencies and distributions of recommendation acts
  - Limitation: cannot measure the overall recommendation performance of the whole dialog
- Dialogue-level Metrics
  - Recommendation accuracy at round k (e.g., Precision@k, Recall@k, NDCG@k)
  - Dialogue success rate (e.g., SuccessRate@ k)
- Business-level Metrics
  - Conversion rate per dialog
  - Sales revenue
  - User satisfaction rating, user retention, customer loyalty

An Example of CRS Evaluation

Can you find me a ***mobile phone*** on Amazon?

Sure, what operating system do you prefer?

I want an Android one.

OK, and any preference on screen size?

Better larger than 5 inches.

Do you have requirements on storage capacity?

I want it to be at least 64 Gigabytes.

And any preference on phone color?

Not particularly.

Sure, then what about the following choices?

I don't like them very much…

OK, do you have any preference on the brand?

Better be Samsung or Huawei.

Any requirement on price?

Should be within 700 dollars.

OK, then what about these ones?

Great, I want the first one, can you order it for me?

Sure, I have placed the order for you, enjoy!

Conversation Evaluation: Evaluate the accuracy of the system-generated **questions**, i.e., percentage of questions that ask about the aspects that the user cares about.

Hit-Rate@(n, k) = #Rounds until round k where the correct aspect is ranked to top-n / k
For each round the system predicts n aspects, HR@(n,k) measures the percentage of rounds where the correct aspect is ranked to top n.

Recommendation Evaluation: Evaluate the recommendation quality at a **particular** recommendation round.

NDCG@n = NDCG of the top-n recommendation list at a particular round

Joint Conversation-Recommendation Evaluation: Evaluate the recommendation quality at the **final** recommendation round.

NDCG@(n, k) = NDCG of the top-n recommendation list until round k / k
Quality of the final recommendation list penalized by the dialog length.

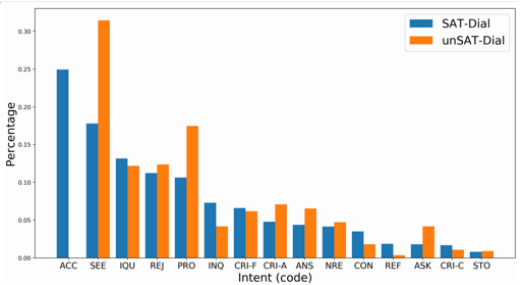
Evaluate by User Study

- Commonly used User Study Questionnaires
  - The bot was easy to interact with. (Usability)
  - The bot helped me complete my task(s). (Usability)
  - The bot was intelligent. (Bot intelligence)
  - I am confident about the quality of service/answers/recommendations (Trust)
  - The bot understood what I said. (Bot intelligence)
  - I feel comfortable interacting with the bot. (Friendliness / Trust)
  - The bot was friendly. (Friendliness)
  - I would be willing to use this bot (or a bot like it) again. (Loyalty)
  - I would recommend using this bot to a friend. (Loyalty)

Evaluate User Intent and Satisfaction

Seeker:	...
Recommender:	Another good one is Spaceballs.
Seeker:	**I did see that one, but I didn't really like it. I do love 80s movies though.**
Recommender:	Ok Well how about Planes, Trains and Automobiles.
Seeker:	**I may have seen that a long time ago but I can't remember. who is in that again?**
Recommender:	Steve Martin and John Candy. It is very funny.
Seeker:	**I love them both. I will try that one. Thanks so much!**

An example conversation in ReDial



User Satisfaction per Intent

Intent (Code)	Description	Example	Percentage
Ask for Recommendation			**18.26%**
Initial Query (IQU)	Seeker asks for a recommendation in the first query.	*"I like comedy do you know of any good ones?"*	12.91%
Continue (CON)	Seeker asks for more recommendations in the subsequent query.	*"Do you have any other suggestions?"*	3.10 %
Reformulate (REF)	Seeker restates her/his query with or without clarification/further constraints.	*"Maybe I am not being clear. I want something that is in the theater now."*	1.50%
Start Over (STO)	Seeker starts a new query to ask for recommendations.	*"Anything that I can watch with my kids under 10."*	0.84%
Add Details			**18.58%**
Provide Preference (PRO)	Seeker provides specific preference for the item s/he is looking for.	*"I usually enjoy movies with Seth Rogen and Jonah Hill."*	12.30%
Answer (ANS)	Seeker answers the question issued by the recommender.	*"Maybe something with more action."* (Q: *"What kind of fun movie you look for?"*)	4.91%
Ask Opinion (ASK)	Seeker asks the recommender's personal opinions.	*"I really like Reese Witherspoon. How about you?"*	2.39%
Give Feedback			**61.92%**
Seen (SEE)	Seeker has seen the recommended item before.	*"I have seen that one and enjoyed it."*	21.14%
Accept (ACC)	Seeker likes the recommended item.	*"Awesome, I will check it out."*	18.89%
Reject (REJ)	Seeker dislikes the recommended item.	*"I hated that movie. I did not even crack a smile once."*	11.50%
Inquire (INQ)	Seeker wants to know more about the recommended item.	*"I haven't seen that one yet. What's it about?"*	6.55%
Critique-Feature (CRI-F)	Seeker makes critiques on specific features of the current recommendation.	*"That's a bit too scary for me."*	6.50%
Critique-Add (CRI-A)	Seeker adds further constraints on top of the current recommendation.	*"I would like something more recent."*	5.35%
Neutral Response (NRE)	Seeker does not indicate her/his preference for the current recommendation.	*"I have actually never seen that one."*	4.29%
Critique-Compare (CRI-C)	Seeker requests sth similar to the current recommendation in order to compare.	*"Den of Thieves (2018) sounds amazing. Any others like that?"*	1.55%
Others	Greetings, gratitude expression, or chit-chat utterances.	*"Sorry about the weird typing."*	14.55%

[26] Cai, Wanling, and Li Chen. "Predicting User Intents and Satisfaction with Dialogue-based Conversational Recommendations." UMAP. 2020.

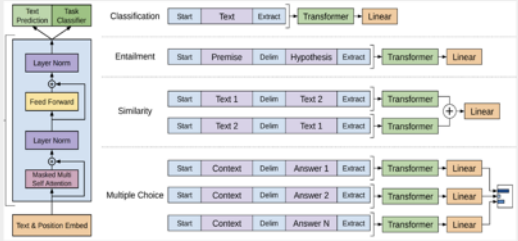
Outline

- Introduction and Background
- Problem Formalization
- Datasets and Evaluation
- **Conversational Recommendation Methods**
  - Three Architectures of Conversational AI Systems
  - Four Major Modules for Conversational Recommendation
  - Natural Language Understanding/Generation
  - Dialog State Managment
  - Recommendation
  - Explanation
- Tool-kits and Real-world Systems

Three Architectures of Conversational AI Systems

End-to-End Architecture

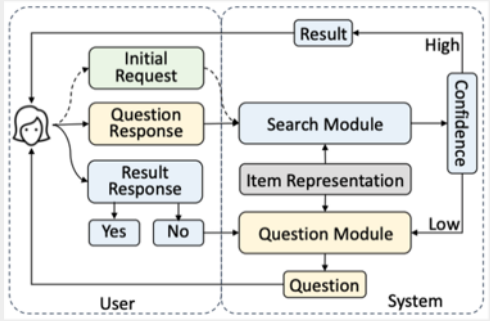
E.g., Sequence-to-Sequence models, Generative Language Models (GPT)



Radford, Alec et. al. Improving Language Understanding by Generative Pre-Training. arXiv 2018.

Modularized Architecture

e.g., Conversational Agent as Linked Functional Modules



Zhang, Yongfeng et. al. Towards Conversational Search and Recommendation: System Ask, User Respond. CIKM 2018.

Data-Flow Architecture

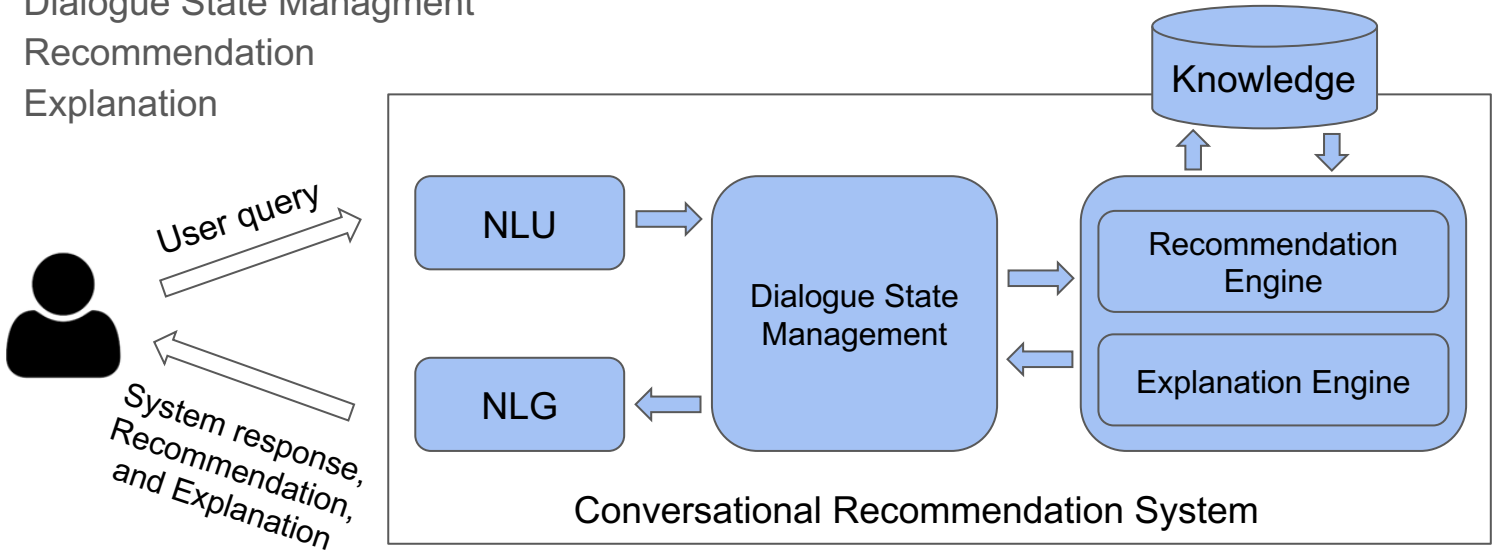
E.g., Dialogue State as Dataflow Graphs (DataFlow)

Andreas, Jacob et. al. Task-Oriented Dialogue as Dataflow Synthesis. TACL 2020.

In the following part of the tutorial we focus on the Modularized Architecture since it is widely used in practical industrial systems.

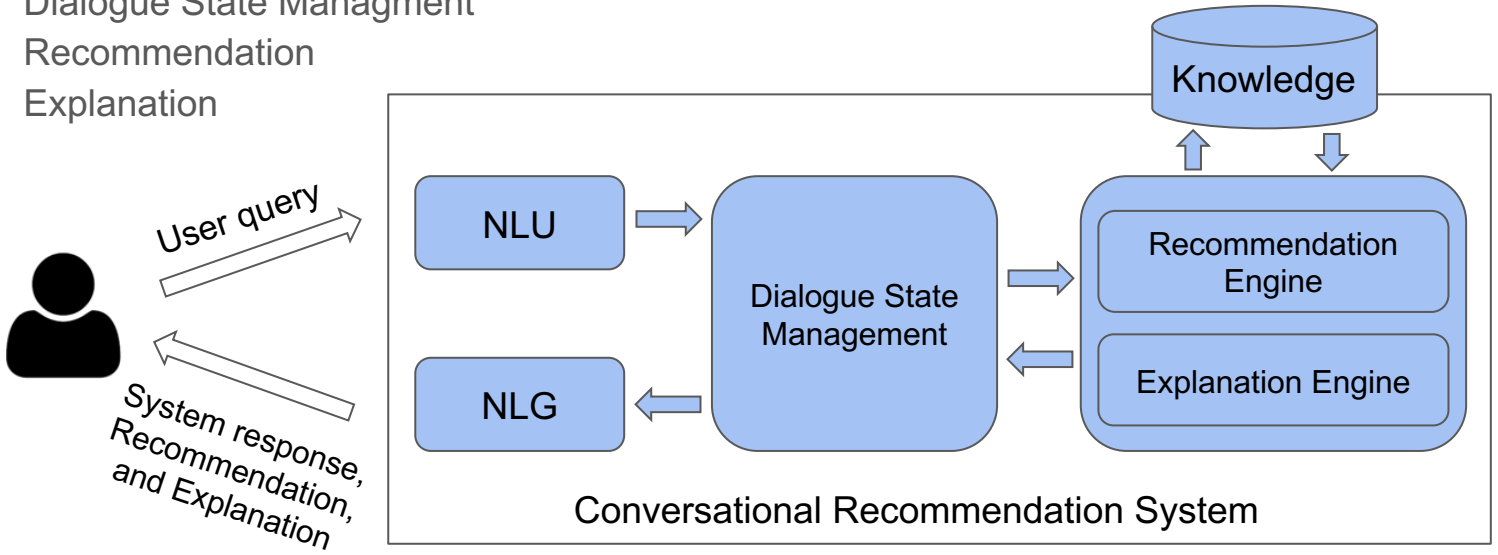
Modularized Architecture for Conversational Recommendation

- Four Major Modules
  - Natural Language Understanding/Generation
  - Dialogue State Managment
  - Recommendation
  - Explanation



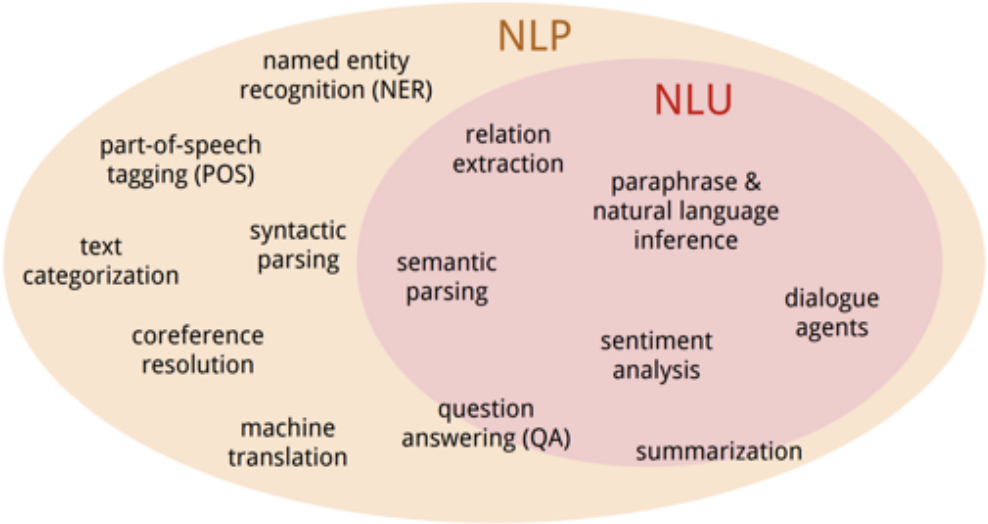
Modularized Architecture for Conversational Recommendation

- Four Major Modules
  - **Natural Language Understanding/Generation**
  - Dialogue State Managment
  - Recommendation
  - Explanation



Natural Language Understanding for CRS

- Interpret free-form text and any type of unstructured data
- Given an utterrance in CRS, the goal is to conduct:
  - Item Category Detection
  - Item Attribute Extraction
  - User Intent Extraction
  - Slot Value Extraction
  - Sentiment Analysis
  - ……



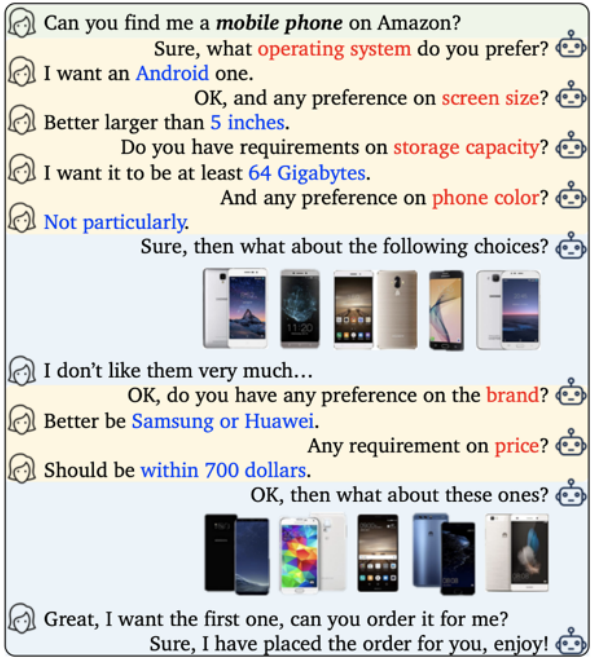
Jannach, Dietmar et. al. A Survey on Conversational Recommender Systems.

Natural Language Geneartion for CRS

- Challenge
  - Lack of background knowledge
  - Consistency and Informative
- Retrieval-Based
  - Fluent and Informative conversational turns
  - Less flexible
- Generation-Based
  - Templated-based
  - Synthesize more flexible and tailored new sentence as reply
  - Insufficient semantics and information => solution: template + knowledge + generation (to make the dialog utterances more controllable)

System Ask, User Respond (SAUR)

- Belongs to Paradigm 1 (SAUP)
- Motivation
  - Learn about the user preferences on items by asking questions
- NLU
  - Multi-Memory Network
- NLG
  - Retrieval-based



Initiation: User initiates a conversation

Conversation: Ask about the user preferences on item aspects (get user feedback)

Display: Display recommenedations to user (when feels confident)

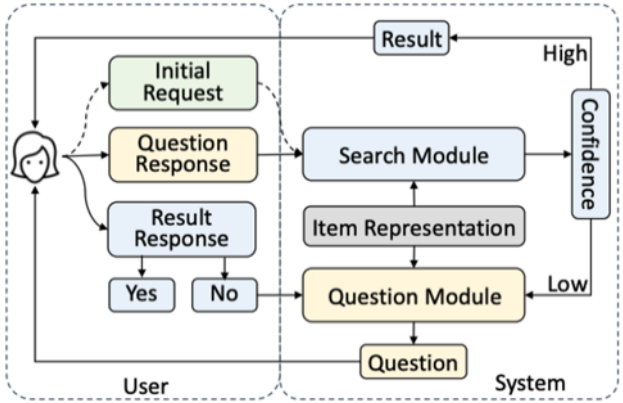
Zhang, Yongfeng et al., Towards Conversational Search and Recommendation: System Ask, User Respond, CIKM 2018.

System Ask, User Respond (SAUR)

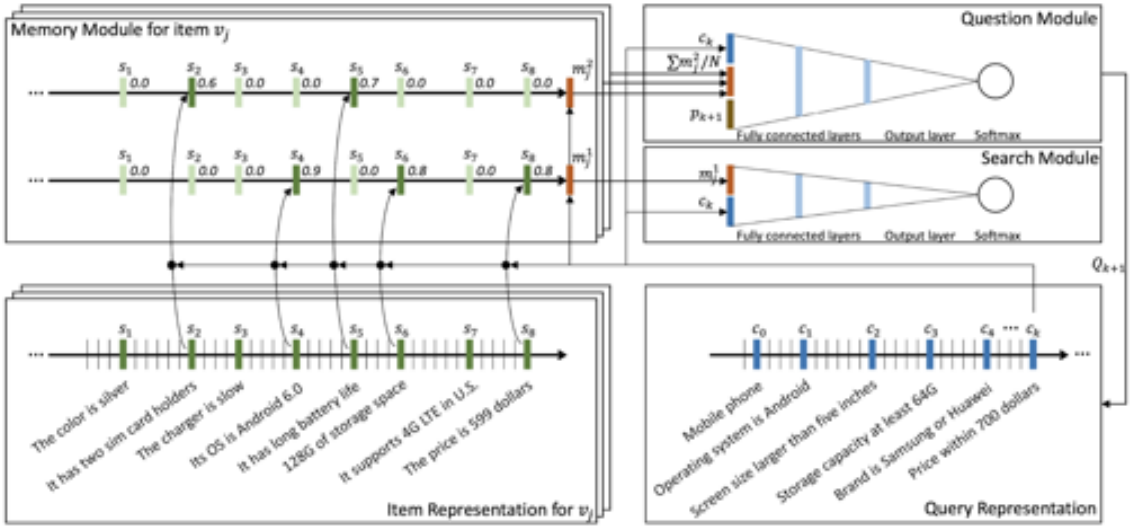
- SAUR formalization

$$u_i \rightarrow Q_0 \mid Q_1A_1, Q_2A_2 \cdots Q_KA_K \mid v_j$$
$$= Q_0(c) \mid Q_1(p_1)A_1(q_1) \cdots Q_K(p_K)A_K(q_K) \mid v_j$$

- The Unified MMN Architecture
  - Item Representations
  - Query Representation
  - Search Module
  - Question Module



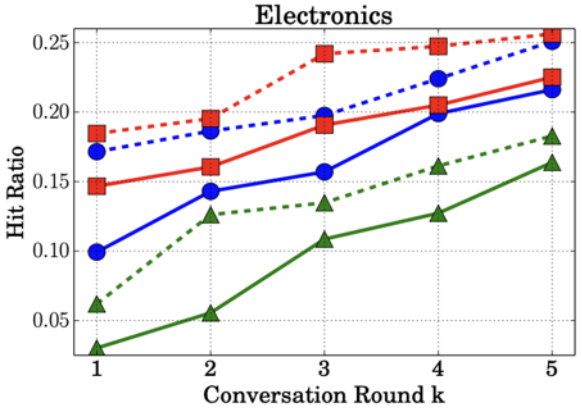
Workflow for SAUR



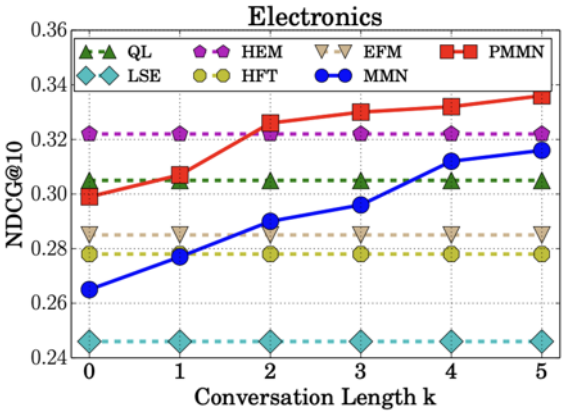
Model Architecture

System Ask, User Respond (SAUR)

- Advantages of Conversational Recommendation
  - Question Prediction (Hit-Ratio @ (n,k))
  - Recommendation (NDCG @ (n,k))



Hit-Ratio @ (n,k): Accuracy of Question Prediction over conversational rounds.



NDCG @ (n,k): Accuracy of Recommendation over conversational round.

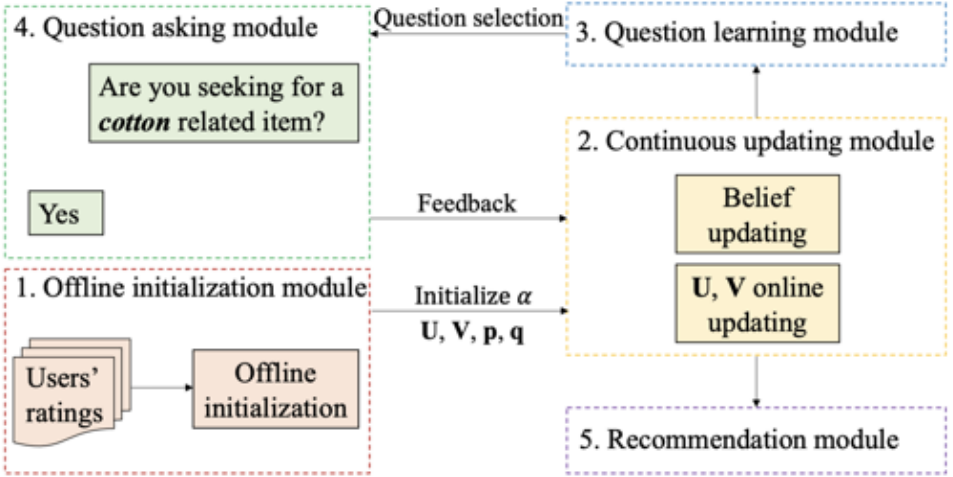
With more conversational rounds, the system can prediction better questions and make more accurate recommendations, which means the system can better understand user preference through conversations.

System Ask, User Respond (SAUR)

- The limitation of the SAUR model
  - Used a simple policy to decide when to ask questions and when to make recommendations (based on a confidence score threshold)
  - Strong assumption on the availability of item facets
  - Does not consider user historical behaviors
- In the following, we introduce approaches that advance SAUR

Question-based Recommendation (Qrec)

- Belongs to Paradigm 1 (SAUP)
- Motivation
  - Enhance the question space by considering item metadata
  - Enhance conversation model based on user interactions
- NLU
  - User response is highly structured (Yes/No/Not Sure), can be easily converted into a vector.
- NLG
  - Generate question by slot-filling



Framework of Qrec

extract short-phrases and entity linking

Zou, Jie et al., Towards Question-based Recommender Systems. SIGIR 2020.

Question-based Recommendation (Qrec)

- Why Better Performance
  - Considers users' historical preference (through user ratings over items)
  - Ask questions about both item aspects in reviews and item aspects in product metadata

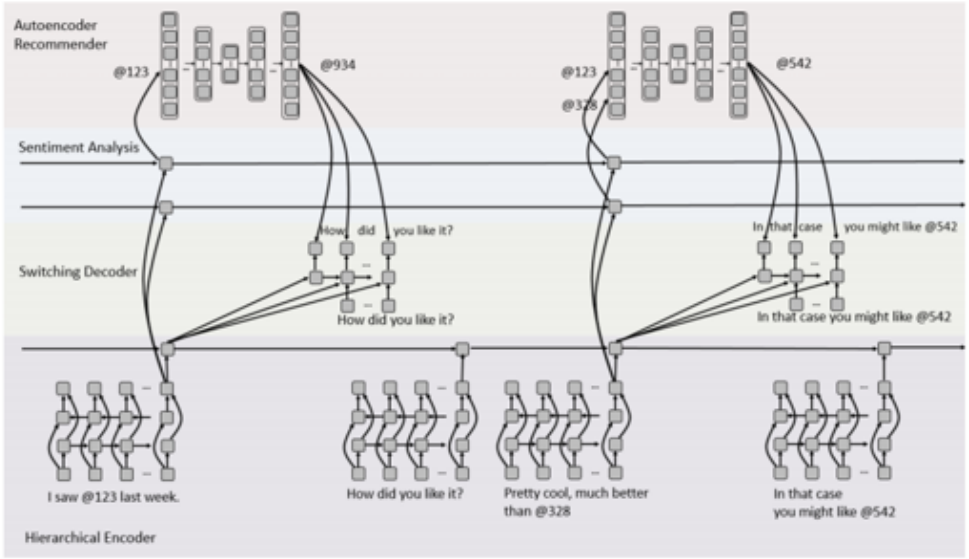
	recall@ 5						
#.	PMF	QMF	NeuMF	QMF+Rand.	SBS	PMMN	Qrec
2	0.011	0.062	**0.222**	0.075	0.060	0.073	0.130
5	0.011	0.062	0.222	0.095	0.353	0.194	**0.443**
10	0.011	0.062	0.222	0.121	0.883	0.216	**0.943**
15	0.011	0.062	0.222	0.151	0.933	0.216	**0.982**
20	0.011	0.062	0.222	0.188	0.962	0.216	**0.995**
	NDCG						
#.	PMF	QMF	NeuMF	QMF+Rand.	SBS	PMMN	Qrec
2	0.036	0.082	0.183	0.113	0.181	0.212	**0.215**
5	0.036	0.082	0.183	0.131	0.389	0.300	**0.443**
10	0.036	0.082	0.183	0.158	0.749	0.310	**0.915**
15	0.036	0.082	0.183	0.184	0.899	0.310	**0.980**
20	0.036	0.082	0.183	0.211	0.935	0.310	**0.993**

Overall Recommendation performance on Amazon Review dataset, The results of PMMN comes from SAUR.

More rounds of conversation brings better performance.

Recommendation through Dialog (ReDial)

- Belongs to Paradigm 2 (SAUE)
- NLU
  - Hierarchical Recurrent Encoder-Decoder (HRED)
- NLG
  - Switching Decoder
- Limitations of Method
  - Does not consider user historical behaviors
  - Simple dialog managment strategy

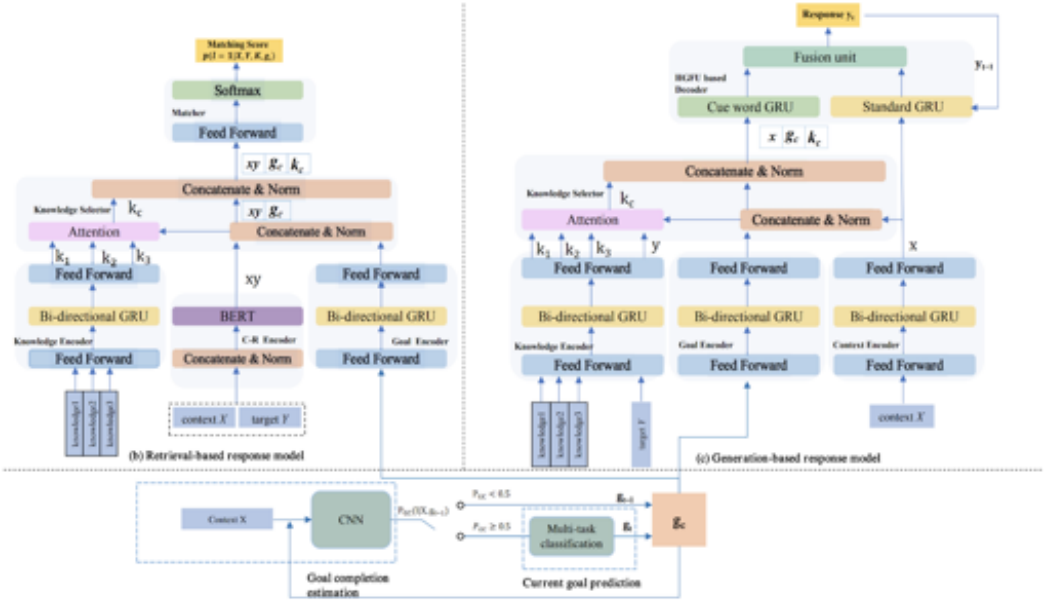


Proposed method for ReDail dataset

Li, Raymond et al., Towards Deep Conversational Recommendations. NerIPS 2018.

Multi-Goal driven Conversation Generation (MGCG)

- Belongs to Paradigm 3 (SAUA)
- Motivation
  - Learn strategies to repurpose the conversation to recommendation when the conversation topic drifts
- NLU
  - Context-response representation module (C-R Encoder)
  - Knowledge representation module (Knowledge Encoder)
- NLG
  - Mixture of Retrieval-based and Generation-based methods



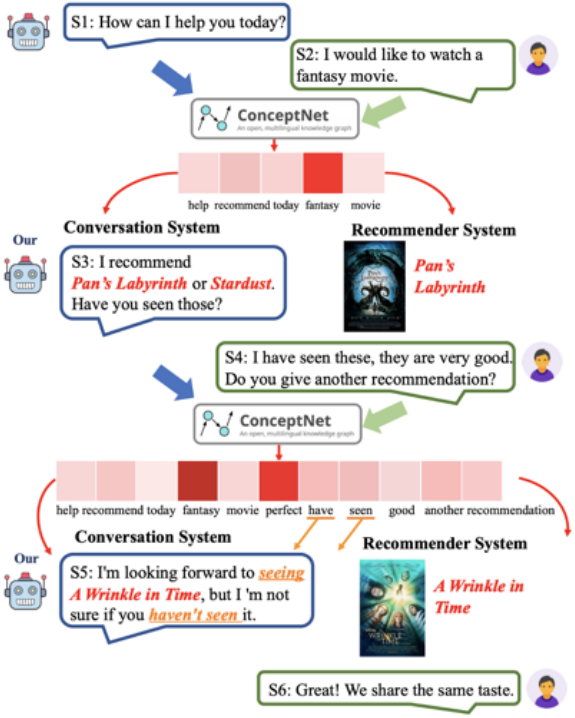
goal planning module

Proposed method

Liu, Zemin., Towards Conversational Recommendation over Multi-Type Dialogs. ACL 2020.

KG-based Semantic Fusion (KGSF)

- Belongs to Paradigm 3 (SAUA)
- Motivation
  - Undertanding dialog semantics by Sentence-KG aligment
- NLU
  - KG fusion to encode Word-KG and Item-KG
- NLG
  - KG-enhanced response generation

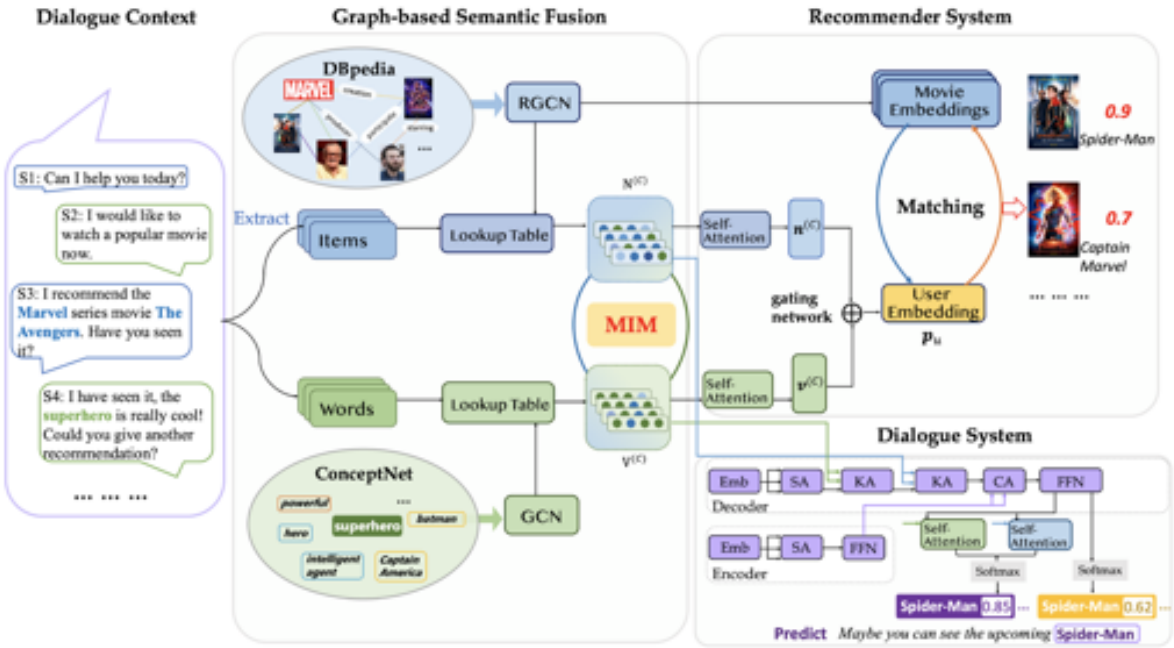


A sampled conversations

Zhou, Kun et al., Improving Conversational Recommender Systems via Knowledge Graph based Semantic Fusion. KDD 2020

KG-based Semantic Fusion (KGSF)

- Contribution
  - Mutual Information Maximization to bridge the semantic gap
- Limitations
  - Monotonous user hisotry
  - Lack of Dialogue state Management

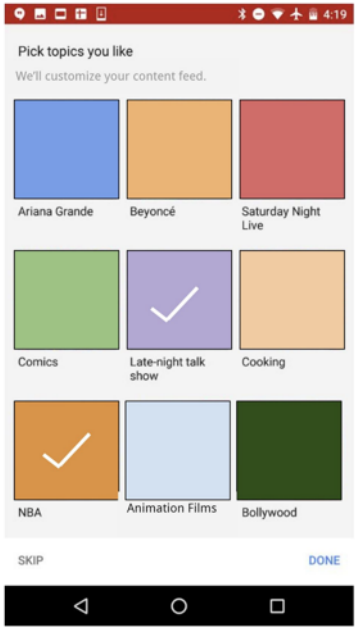


Overvell of the model

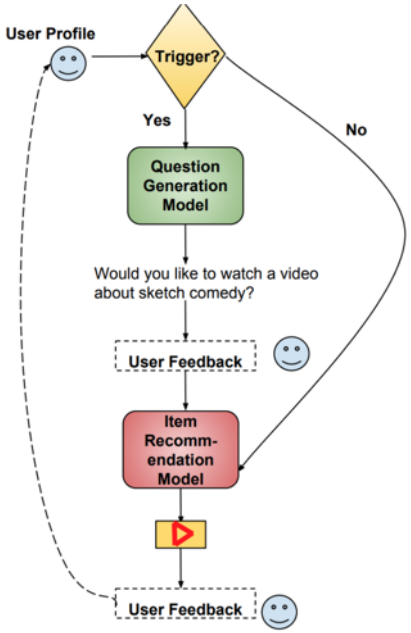
Zhou, Kun et al., Improving Conversational Recommender Systems via Knowledge Graph based Semantic Fusion. KDD 2020

Deployment in Commercial Systems (YouTube)

- Q&R: A Two-Stage Approach toward Interactive Recommendation
- Belongs to Paradigm 1 (SAUP)
- Motivation
  - NLU/NLG can be replaced by other similar funtional modules in commercial systems depending on available UI
- NLU
  - Mulpile choice
- NLG
  - Retrieve from a question pool



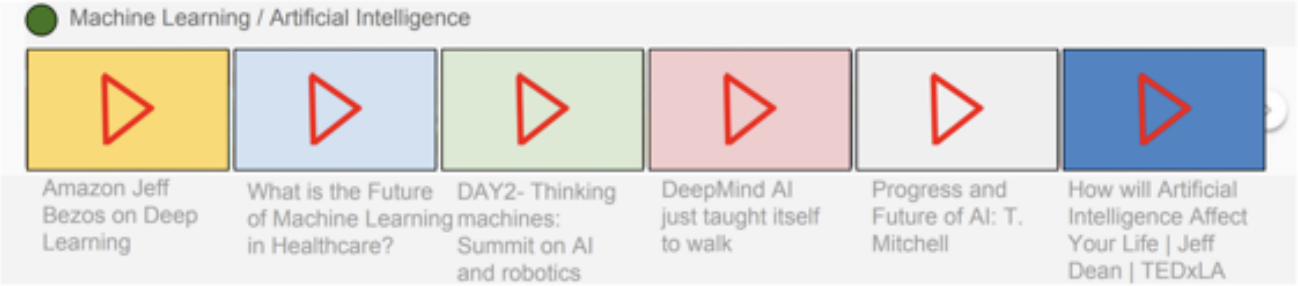
User Onboarding UI



Framework Overview

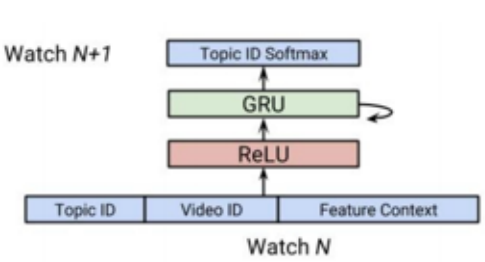
Christakopoulou, Konstantina, et al. "Q&R: A two-stage approach toward interactive recommendation." KDD. 2018.

Deployment in Commercial Systems (YouTube)

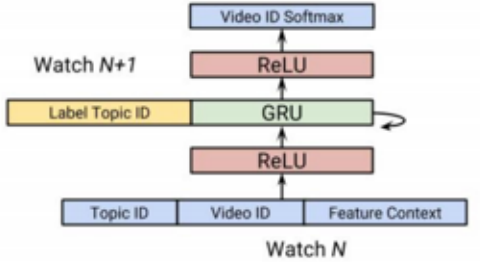


Topic Shelf related to the topic 'Machine Learning'

- Method: Sequential Neural Models
  - Watch history
  - Clicked topic
  - Good response



Question Ranking Module

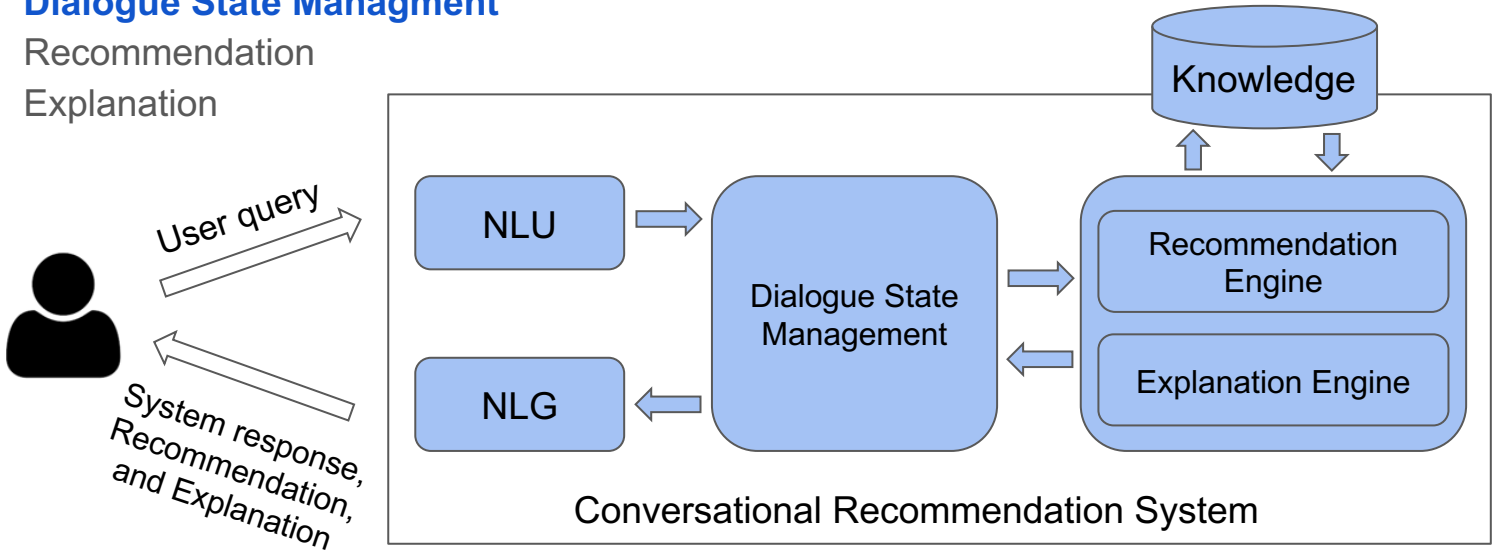


Response Module

Christakopoulou, Konstantina, et al. "Q&R: A two-stage approach toward interactive recommendation." KDD. 2018.

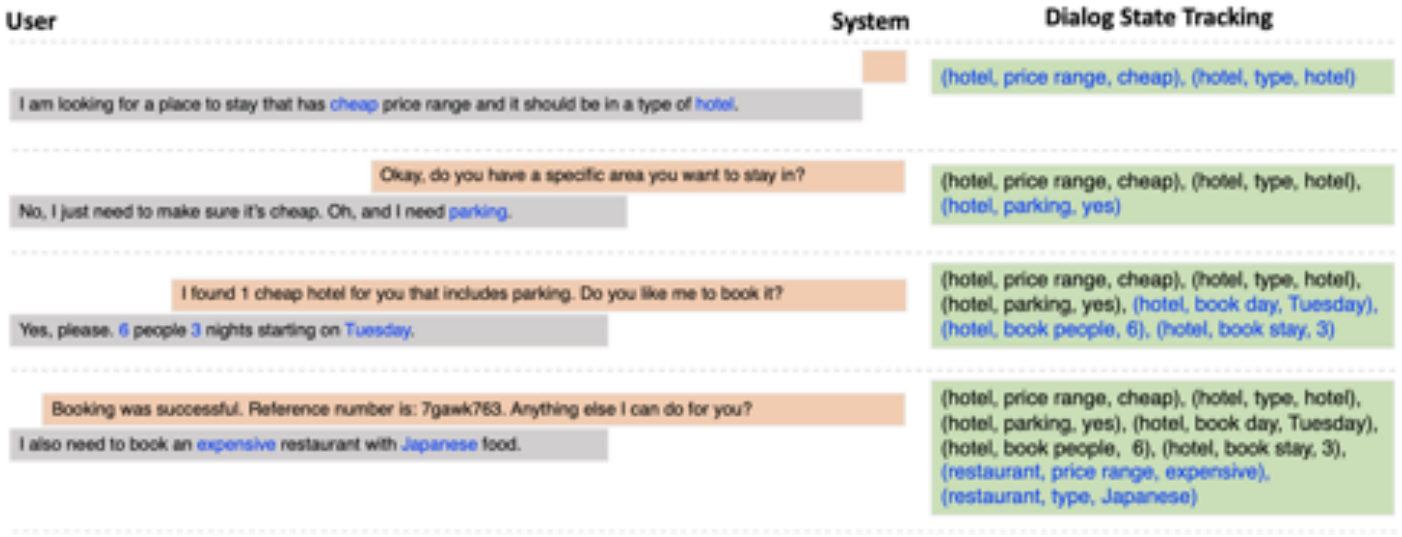
Modularized Architecture for Conversational Recommendation

- Four Major Modules
  - Natural Language Understanding/Generation
  - **Dialogue State Managment**
  - Recommendation
  - Explanation



Dialogue State Management (DSM)

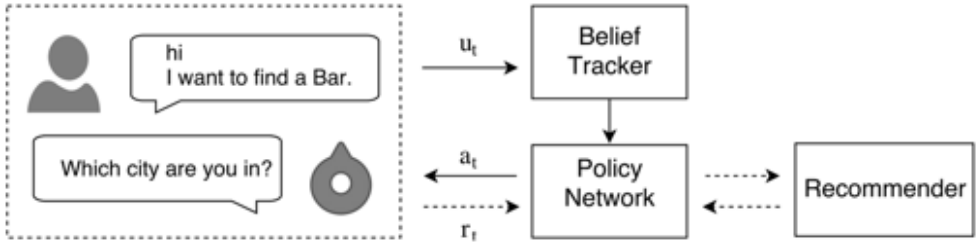
- Keep track of the dialogue state to decide the next conversational move
  - e.g., whether to make a recommendation or to ask a question



Zhang, Jianguo et al., Find or Classify? Dual Strategy for Slot-Value Predictions on Multi-Domain Dialog State Tracking. *SEM 2020.

Conversational Recommendation Model (CRM)

- Belongs to Paradigm 1 (SAUP)
- Motivation
  - Learn a model to decide when to recommendation and when to ask
- NLU
  - Deep belief tracker to analyze user's current utterance
  - Extract the facet values of the target item
- DSM
  - RL-based policy network for dialog state management

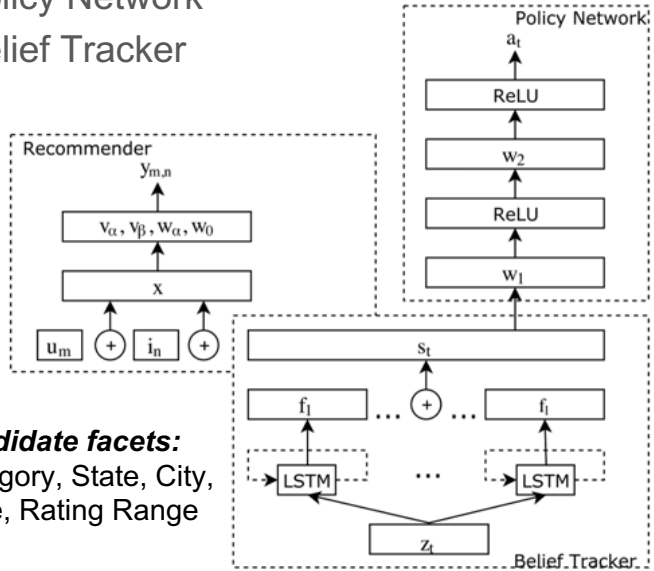


Overview of the CRM Framework

Sun, Yueming, et al. "Conversational Recommender System." SIGIR. 2018.

Conversational Recommendation Model (CRM)

- Key Components of the Model
  - Policy Network
  - Belief Tracker



Candidate facets:
Category, State, City, Price, Rating Range

The structure of the proposed model

User: I NEED MEXICAN CATEGORY *(category="Mexican")*
Agent: Which city are you in?
User: GILBERT *(city="Gilbert")*
Agent: Which rating range do you want?
User: 4.0 *(rating_range>="4.0")*
Agent: <make recommendation, target ranked 7 of 12>

Note: The RL agent asked for three slots and then make an recommendation, no need to ask for the rests.

User: juice bars & smoothies *(category="Juice Bars & Smoothies")*
Agent: Which city are you in?
User: henderson *(city="Henderson")*
Agent: Which rating range do you like?
User: 4.0 *(rating_range>="4.0")*
Agent: <make recommendation, target ranked 2 of 4>

Sampled Example

Sun, Yueming, et al. “Conversational Recommender System.” SIGIR. 2018.

Conversational Recommendation Model (CRM)

- Limitations of the CRM model
  - Less scalable to large size of item attributes
  - Incapable of updating user preference in multi-round setting where user may provide negative feedbacks on recommendations
- In the following, we introduce approaches that advance CRM

The structure of the proposed model

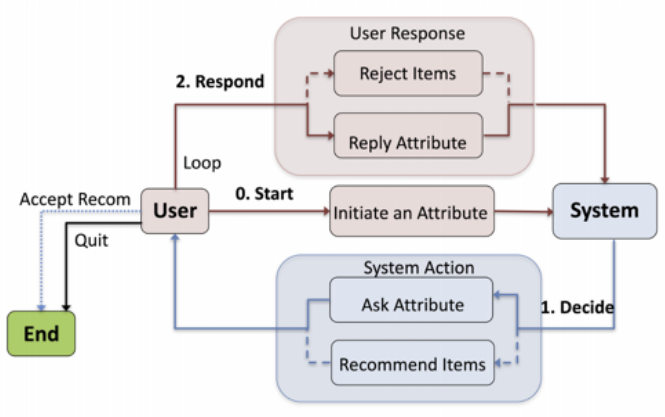
Sampled Example

Sun, Yueming, et al. “Conversational Recommender System.” SIGIR. 2018.

Estimation–Action–Reflection (EAR)

- Belongs to Paradigm 1 (SAUP)
- Motivation
  - What Attributes to ask? When to recommend items? How to adapt to user's online feedback?

Example of the conversation

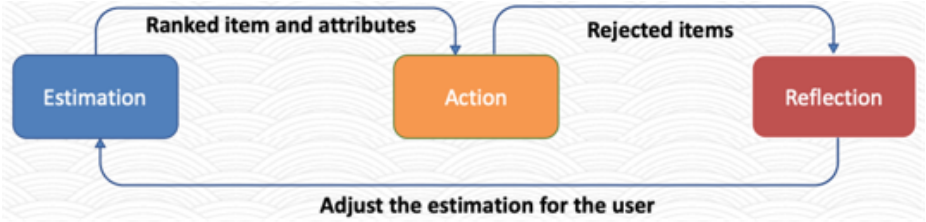


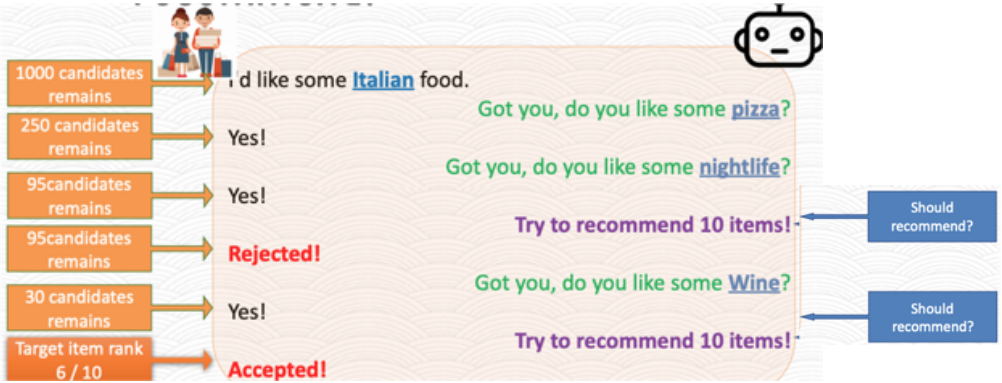
Workflow of the system

Lei, Wenqiang et al., Estimation–Action–Reflection: Towards Deep Interaction Between Conversational and Recommender Systems. WSDM 2020.

Estimation–Action–Reflection (EAR)

- DSM
  - A policy network that integrates conversational component and recommender component
- Limitations
  - Strong assumption that user expresses the preferences clearly
  - Excluded user negative feedbacks





example: Strategy to ask and recommend?

Lei, Wenqiang, et al. "Estimation-action-reflection: Towards deep interaction between conversational and recommender systems." WSDM. 2020.

Conversational Path Reasoning (CPR)

- Belongs to Paradigm 1 (SAUP)
- Motivation
  - Utilize user's attribute feedback explicitly (Compared to EAR)
  - Hetergeogeneous fine-grained user preference
- Method
  - Model conversational recommendation as an interactive path reasoning problem over graph.

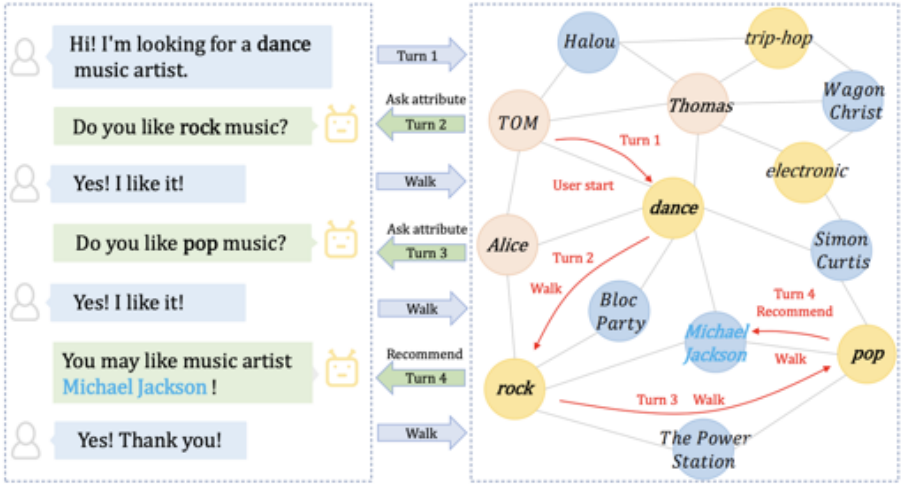
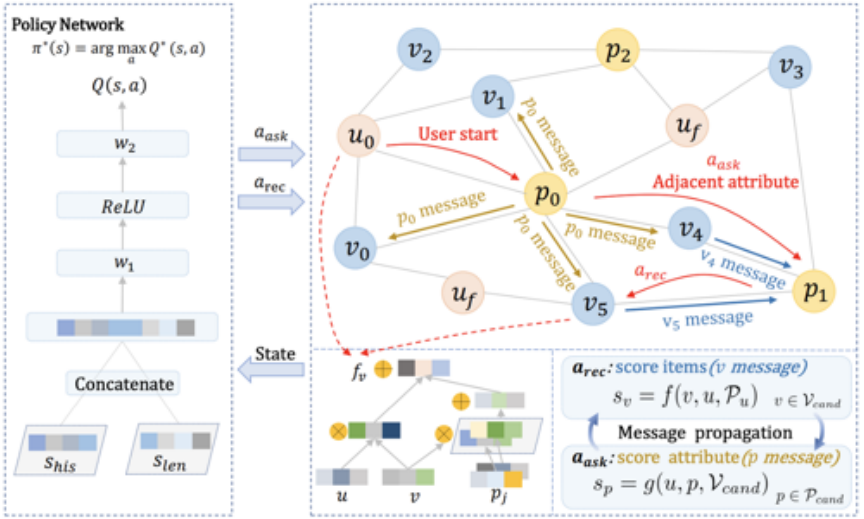


illustration of interactive path reasoning in CPR.

Lei, Wenqiang, et al. "Interactive Path Reasoning on Graph for Conversational Recommendation." KDD. 2020.

Conversational Path Reasoning (CPR)

- Method
  - Adjacent attributes: reduces the search space for selecting which attribute to ask
  - Policy function: whether to ask an attribute or to recommend items
- DSM
  - knowledge-ground transparent dialogue state tracking



Overview of CPR framework

Conversational Path Reasoning (CPR)

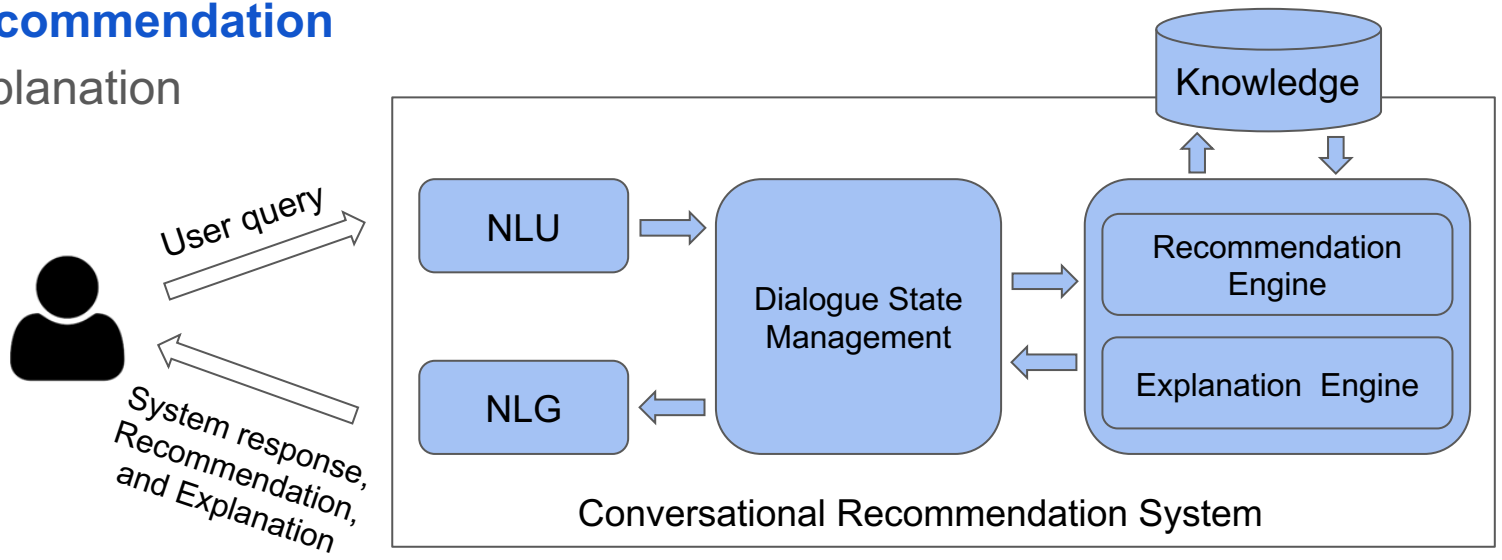
- Conversational Path Reasoning(CPR)

Scenerio Example Conversation of CPR vs EAR

https://dl.acm.org/doi/pdf/10.1145/3336191.3371769
Interactive Path Reasoning on Graph for Conversational Recommendation

Modularized Architecture for Conversational Recommendation

- Natural Language Understanding/Generation
- Dialogue State Managment
- **Recommendation**
- Explanation

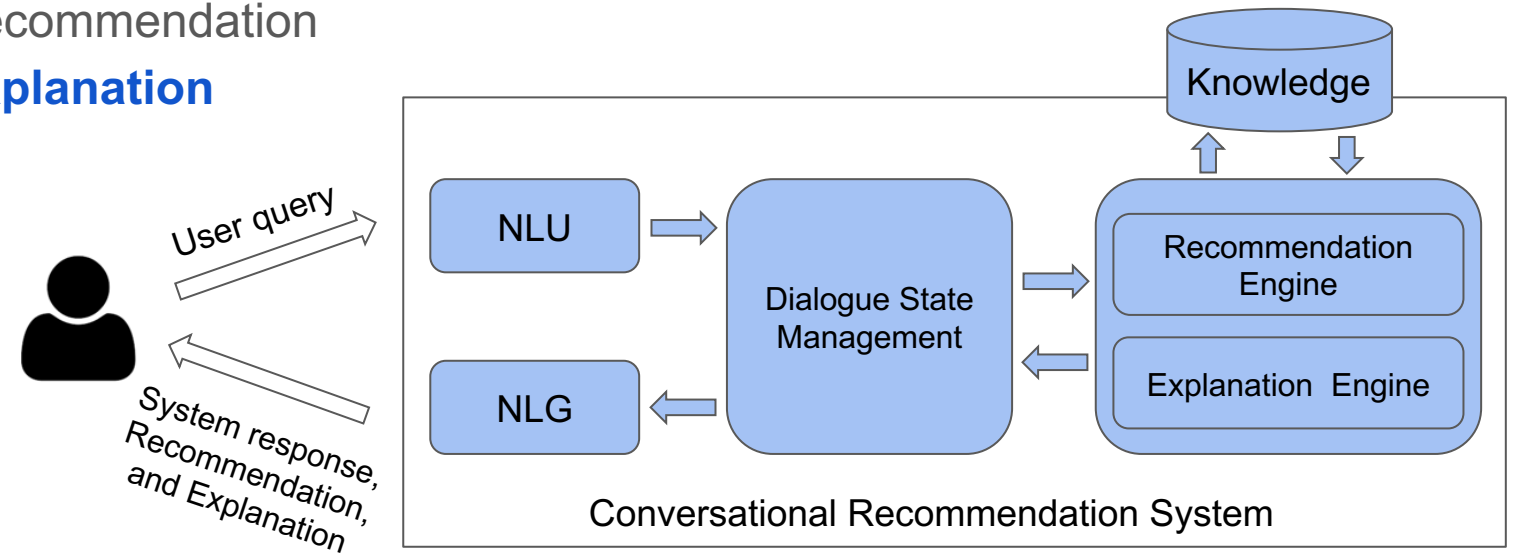


Recommendation

- Key Differences from Traditional Recommendation Models
  - How to make use of the **information** encoded in the dialog?
  - How to model the user preference **dynamically** during the dialog?
- Make use of the information encoded in the dialog
  - Extract **structured information** from dialog (e.g., user preference on aspects)
  - Encode dialog into **continous embeddings** (e.g., word/sentence embeddings)
- Model the user preference dynamically during the dialog
  - Update **user profile** based on the structured information from dialog
  - Update **user embedding** based user profiles and/or continous dialog embeddings
- Many recommendation models can be used based on above information
  - Content-based, Collaborative Filtering, Matrix Factorization, Neural Networks, Knowledge Graph Reasoning, Graph Neural Networks, etc.

Modularized Architecture for Conversational Recommendation

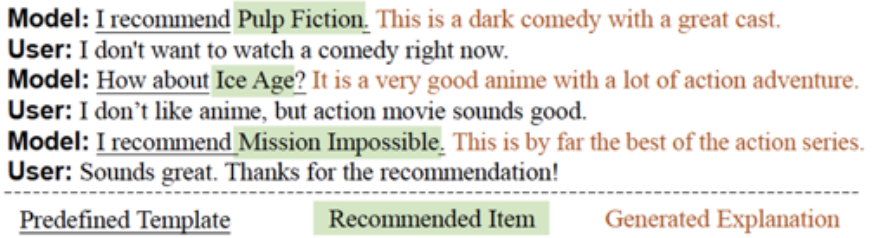
- Natural Language Understanding/Generation
- Dialogue State Managment
- Recommendation
- **Explanation**



Explanations are especially important in conversational recommendation settings, because users can better lead the conversation if they know why certain items are recommended, e.g., by reconsidering what question to ask the system according to the explanations provided by the system.

Explainable Conversational Recommendation (ECR)

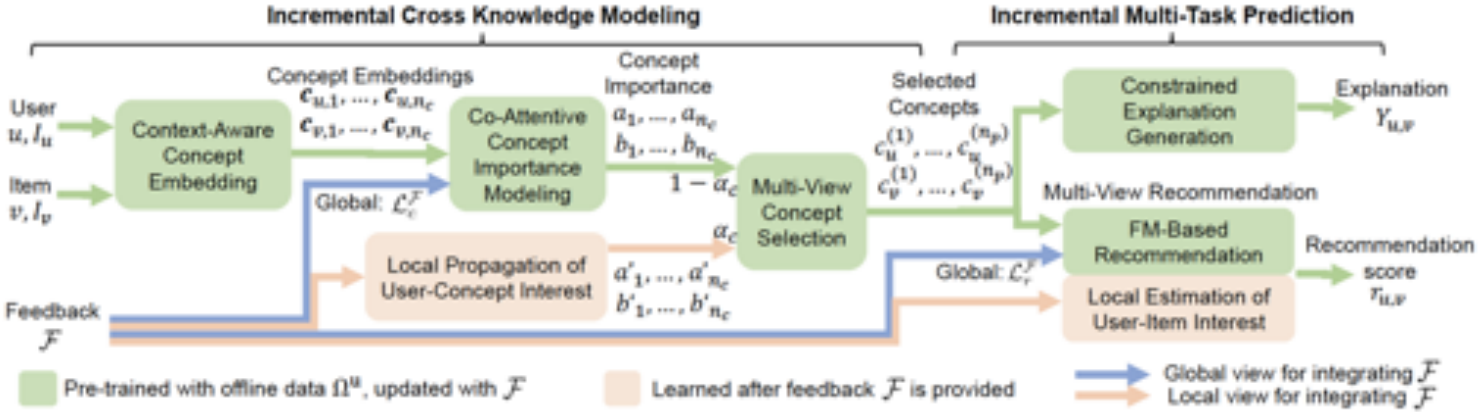
- Belongs to Paradigm 2 (SAUE)
- Motivation
  - Provide explanations to help users understand the recommendation and dialog
  - Collect user feedback from explanations to understand user needs
- NLU
  - Context-aware concept embedding
- NLG
  - Template-based and GRU
- Explanation
  - Constrained explanation generation via bidirectional GRU



Chen, Zhongxia, et al. "Towards Explainable Conversational Recommendation." IJCAI, 2020.

Explainable Conversational Recommendation (ECR)

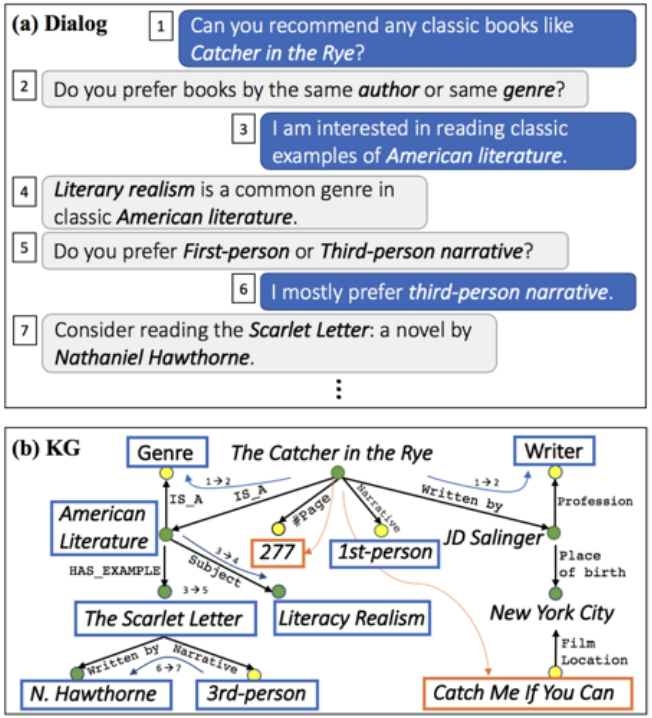
- Incremental multi-task learning framework



- Limitation
  - System always recommend items at each round
  - Strong assumption on user feedbacks (users must explicitly provide positive/negative feedback)

OpenDialKG

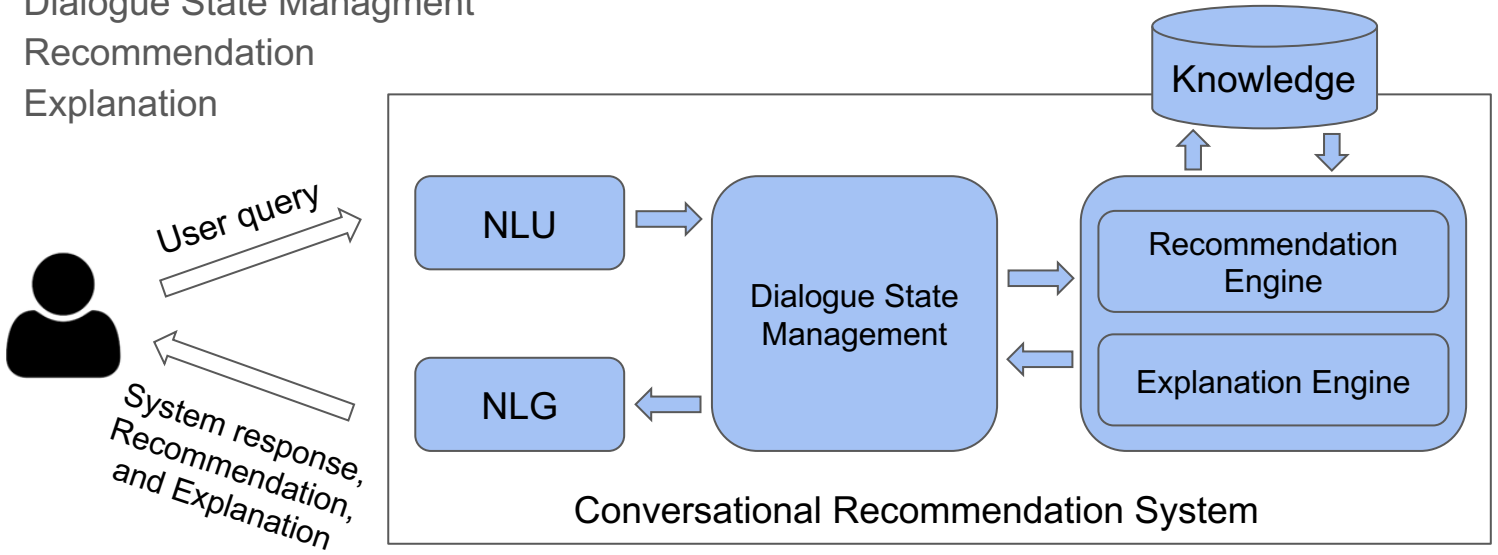
- Belongs to Paradigm 2 (SAUE)
- Motivation
  - Associate the dialogue utterances with KG
  - Transparent state tracking via explainable path
- NLU
  - Hierarchical Bi-LSTM
- Explanation
  - KG Path Walker, which prunes unattended paths to effectively reduce the search space
  - More transparent and explainable



Moon, Seungwhan, et al. "Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs." ACL. 2019.

Brief Summarization of CRS Models

- Four Major Modules
  - Natural Language Understanding/Generation
  - Dialogue State Managment
  - Recommendation
  - Explanation
- Three Paradigms
  - SAUP, SAUE, SAUA



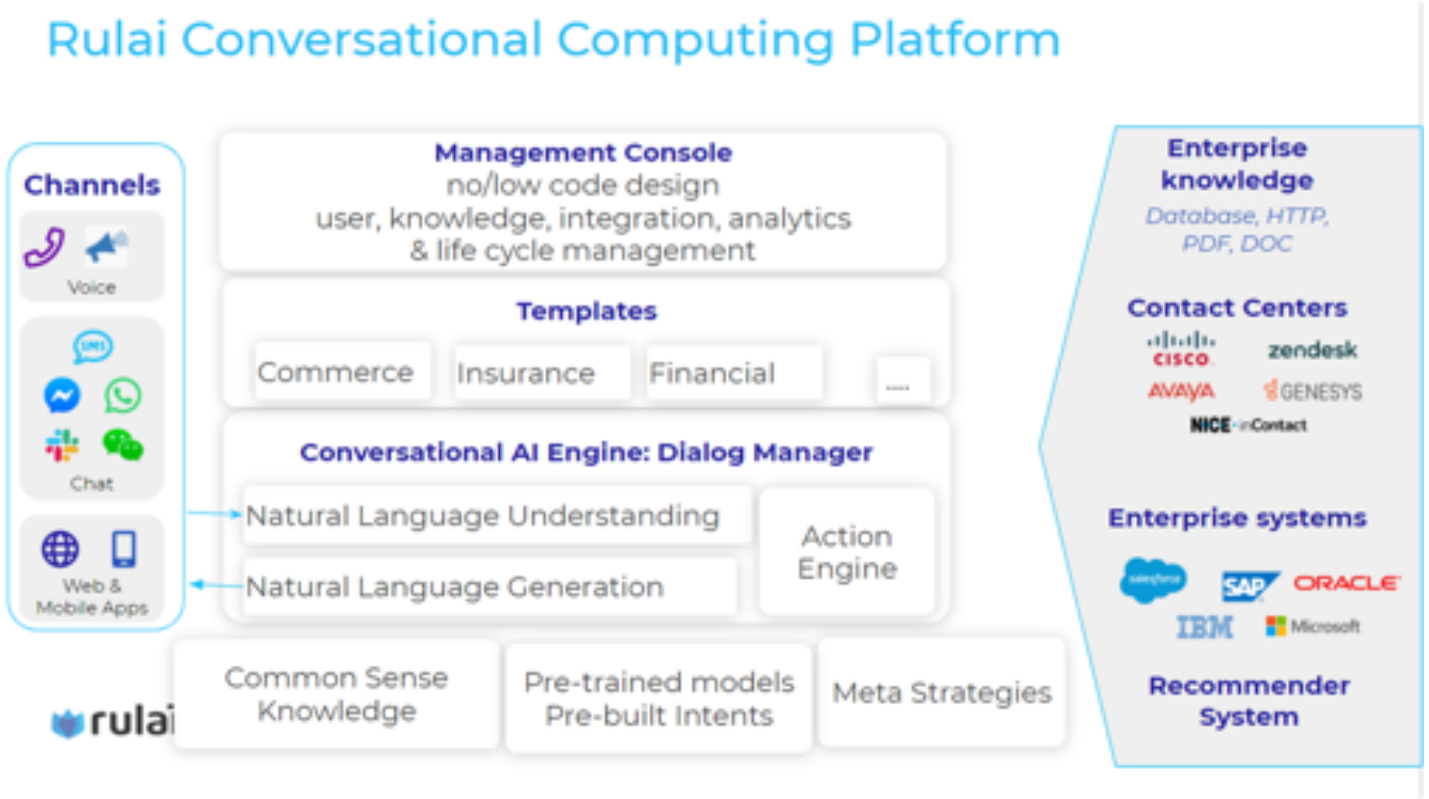
Outline

- Introduction and Background
  - Introduction to Conversational Recommendation
  - A Brief History of Conversational Recommendation Research
- Problem Formulation
- Datasets and Evaluation
- Conversational Recommendation Methods
- **Tool-kits and Real-world Systems**

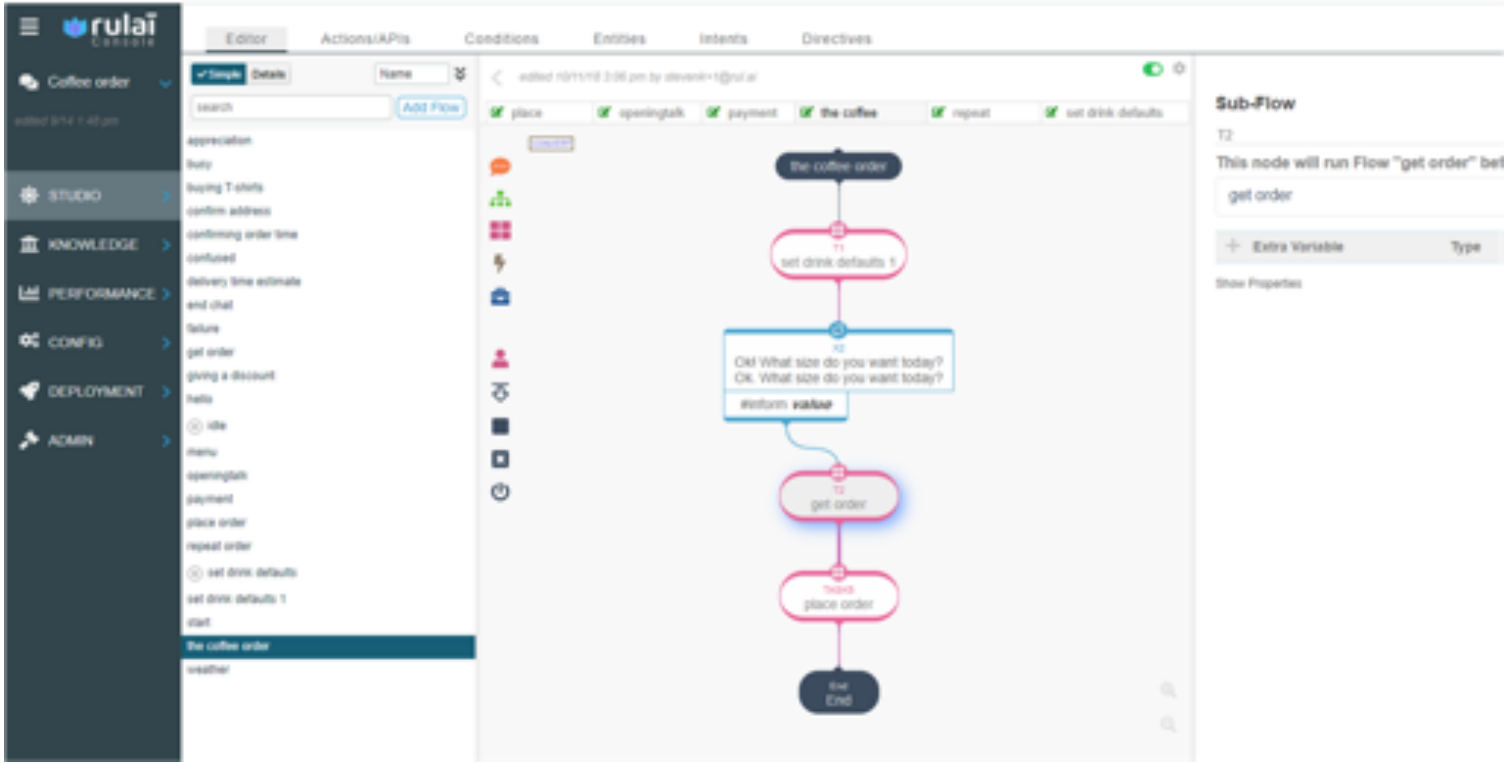
Tool-Kit and Real-world Systems

- Open source dialog systems + open source recommener systems
  - https://github.com/evison/Conversational
  - Dialog systems: CMU Olympus, Deeppavlov, Uber Plato, Cisco Mindmeld, Rasa, etc.
- Commercial conversational AI tool kits for developers
  - Microsoft Bot framework, Google Dialogflow, Amazon Lex, IBM Watson etc.
  - Apple iMessage for business, Facebook messenger bot, etc.
- Commercial all-in-one Conversational AI platforms
  - https://rul.ai/request-a-free-account/ etc.

Example: All-in-One Conversational Platform



Example: Drag & Drop Conversational Design



Conclusion & QA

- Introduction and Background
- Problem Formalization
- Datasets and Evaluation
- Conversational Recommendation Methods
- Tool-kits and Real-world Systems

References

1. Croft and Thompson. 13R: A New Approach to the Design of Document Retrieval Systems. JASIST 1987.
2. Belkin et al. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. 1995.
3. Goker and Thompson. The adaptive place advisor: A conversational recommendation system. 2000.
4. Bridge, D.G. Towards Conversational Recommender Systems: A Dialogue Grammar Approach. ECCBR workshops. 2002.
5. McGinty and Smyth. On the role of diversity in conversational recommender systems. Inter. Conf. on Case-Based Reasoning. 2003
6. Smyth, McGinty, Reilly and McCarthy. Compound critiques for conversational recommender systems. In Web Intelligence. 2004.
7. Wärnestål. User evaluation of a conversational recommender system. Knowledge and Reasoning in Practical Dialogue Systems. 2005.
8. Mahmood, Mujtaba and Venturini. Dynamic personalization in conversational recommender systems. Info. Sys. and e-Business. 2014.
9. Christakopoulou, Radlinski and Hofmann. Towards conversational recommender systems. KDD. 2016.
10. Sun, Yueming, et al. “Conversational Recommender System.” SIGIR. 2018.
11. Zhang, Yongfeng, et al. "Towards conversational search and recommendation: System ask, user respond." CIKM. 2018.
12. Christakopoulou, Konstantina, et al. "Q&R: A two-stage approach toward interactive recommendation." KDD. 2018.
13. Li, Raymond, et al. "Towards deep conversational recommendations." NeuIPS. 2018.
14. Moon, Seungwhan, et al. "Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs." ACL. 2019.
15. Chen, Qibin, et al. "Towards knowledge-based recommender dialog system." EMNLP. 2019.
16. Jannach, Dietmar, et al. "A Survey on Conversational Recommender Systems." arXiv. 2020.
17. Zou, Jie, et al. "Towards Question-based Recommender Systems." SIGIR. 2020.
18. Zhang, Jian-Guo, et al. "Find or classify? dual strategy for slot-value predictions on multi-domain dialog state tracking." SEM. 2020.
19. Lei, Wenqiang, et al. "Estimation-action-reflection: Towards deep interaction between conversational and recommender systems." WSDM. 2020.
20. Lei, Wenqiang, et al. "Interactive Path Reasoning on Graph for Conversational Recommendation." KDD. 2020.
21. Liu, Zeming, et al. "Towards Conversational Recommendation over Multi-Type Dialogs." ACL. 2020.
22. Zhou, Kun, et al. "Improving Conversational Recommender Systems via Knowledge Graph based Semantic Fusion." KDD. 2020.
23. Chen, Zhongxia, et al. "Towards Explainable Conversational Recommendation." IJCAI, 2020.
24. Xu, Hu, et al. "User Memory Reasoning for Conversational Recommendation." arXiv. 2020.
25. Fu, Zuohui, et al. "COOKIE: A Dataset for Conversational Recommendation over Knowledge Graphs in E-commerce." arXiv. 2020.
26. Cai, Wanling, and Li Chen. "Predicting User Intents and Satisfaction with Dialogue-based Conversational Recommendations.” UMAP. 2020.

RUTGERS

The ACM Conference Series on
Recommender Systems

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Tutorial on Conversational Recommendation Systems

Zuohui Fu, Yikun Xian, Yongfeng Zhang, Yi Zhang
zuohui.fu@rutgers.edu
yikun.xian@rutgers.edu
yongfeng.zhang@rutgers.edu
yiz@soe.ucsc.edu