Data Augmentation for Conversational AI

Fundamentals and Advances

Heydar Soudani, Faegheh Hasibi, Evangelos Kanoulas

It should be mentioned that we are still working on these slides and they are not the final version

Tutorial website: https://dataug-convai.github.io

Presenters

A first-year PhD candidate

Radboud University, Nijmegen

heydar.soudani@cs.ru.nl

Assistant professor

Radboud University, Nijmegen

f.hasibi@cs.ru.nl

Full professor

University of Amsterdam

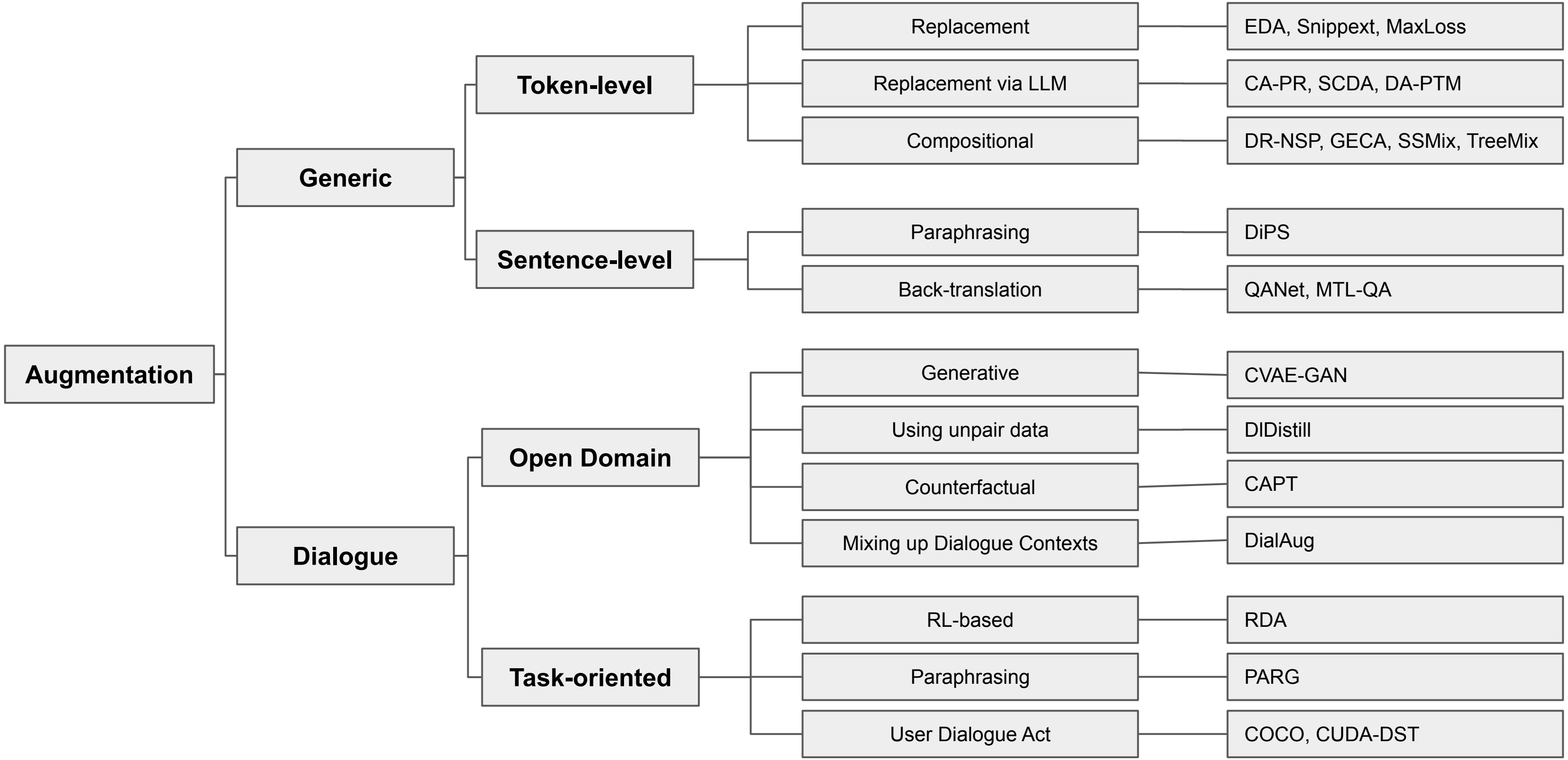
e.kanoulas@uva.nl

Agenda

- **Introduction (10 min)**
  - Conversational Systems
    - Open-domain
    - Task-oriented
  - Problem of Data Scarcity
  - Data Augmentation
    - Augmenting with Existing Conversations Data
    - Augmenting with Conversation Generation
- **Conversation Augmentation (30 min)**
  - Generic Token-level & Sentence-level Augmentation
  - Dialogue Data Augmentation
- **Conversation Generation: Open Domain (80 min)**
  - Single-turn QA Pair Generation
  - Multi-turn Dialogue Generation
  - Topic-aware Dialogue Agent
    - One-turn Topic Transitions
    - Target-oriented Dialogue Systems
- **Conversation Generation: Task-oriented (40 min)**
  - Schema-guided Generation
  - Simulator-agent Interaction
  - E2E Dataset Creation
- **Evaluation (10 min)**
- **Conclusion and Future Direction (10 min)**

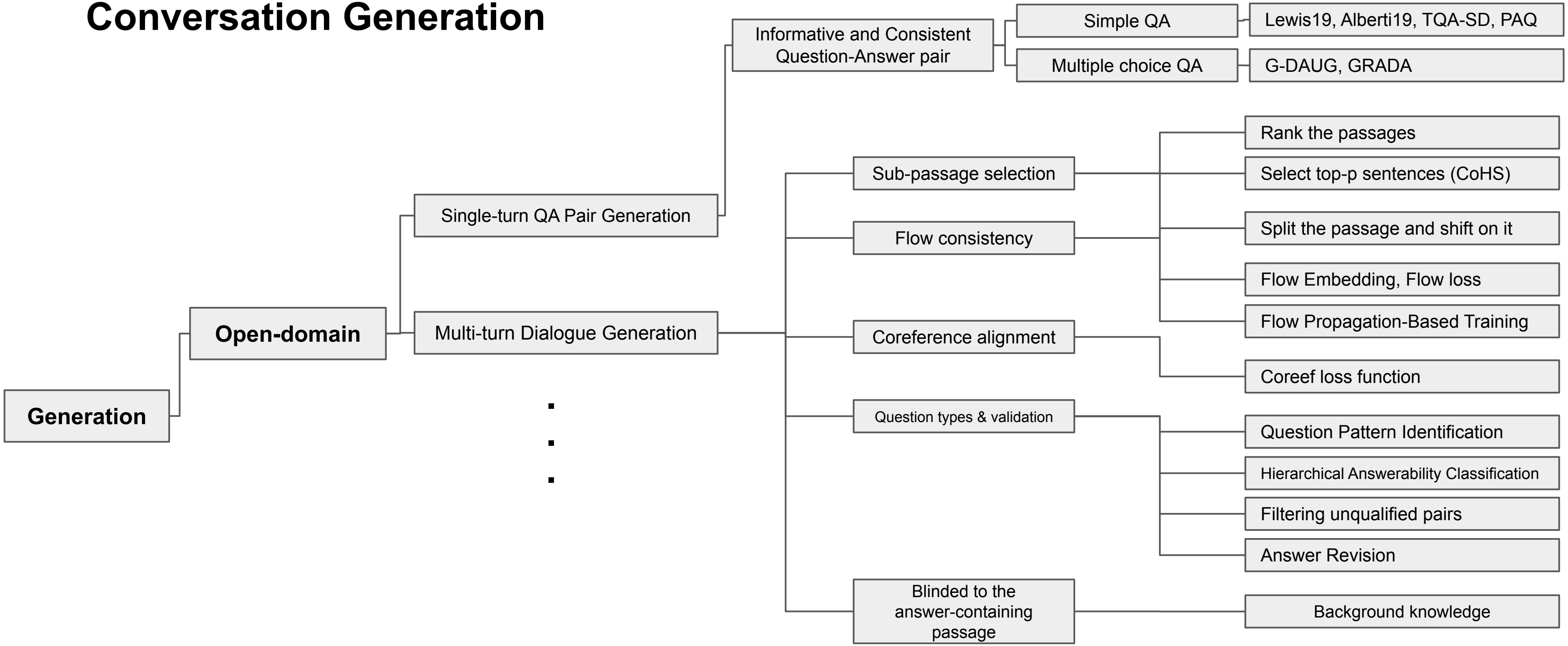
Tutorial overview

Conversation Augmentation



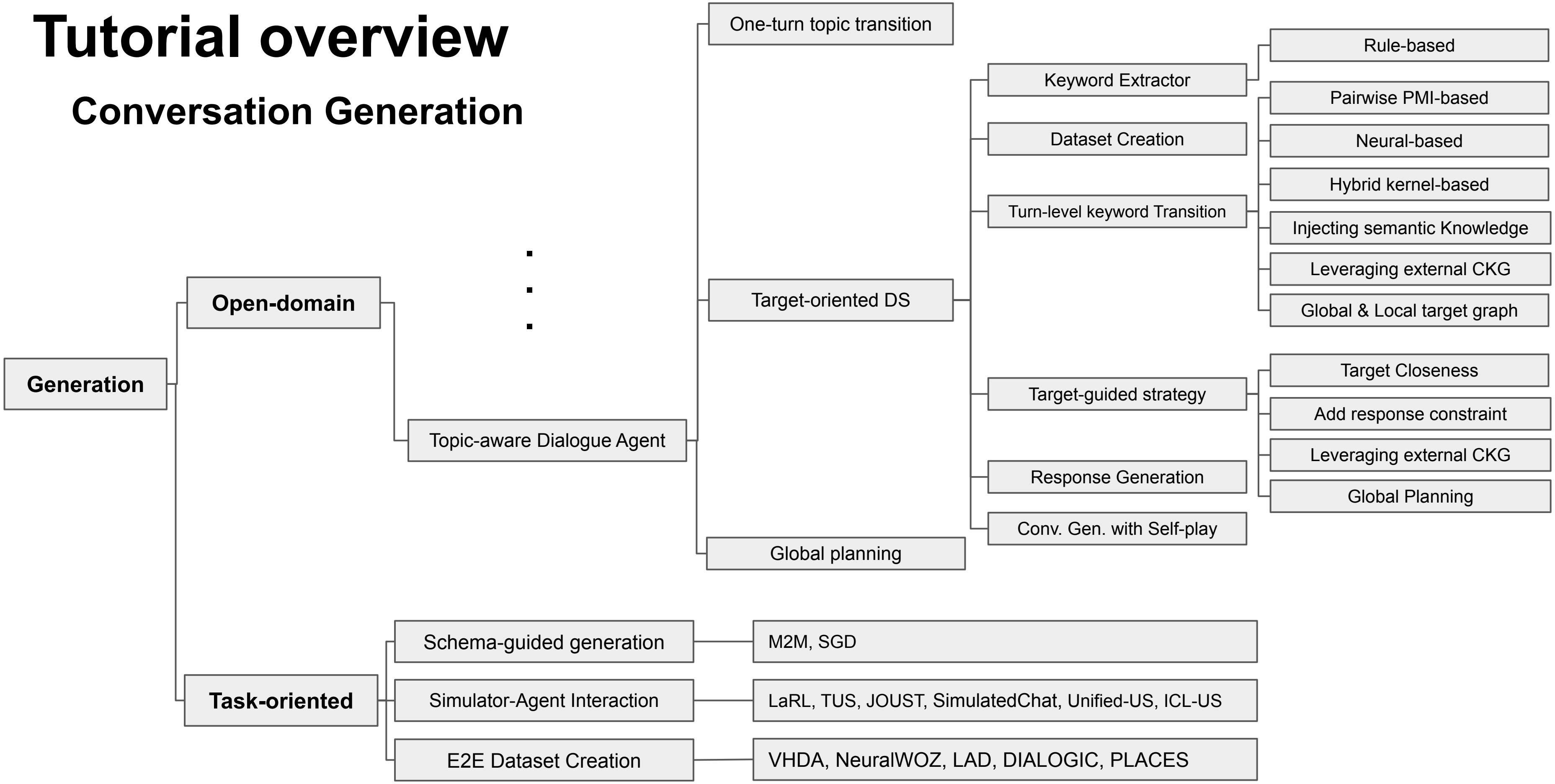
Tutorial overview

Conversation Generation



Tutorial overview

Conversation Generation



List of all papers

Table 1: List of current works for Dialogue Augmentation

Paper	Year	tr. Method	Level	Type	Challenges	Approach	Input
Aug: Generic Token-level & Sentence-level Augmentation							
MTL-QA [3]	2021-05	–	Sentence			Back-translation	
QANet [66]	2018-08	–	Sentence			Back-translation	
DiPS [22]	2019-06	–	Sentence		Inducing diversity	Paraphrasing	
TreeMix [68]	2022-07	–	Token		Meaningfulness, Diversity	Compositional	–
SSMix [65]	2021-08	–	Token		Meaningfulness, Diversity	Compositional	–
GECA [2]	2020-07	–	Token		Meaningfulness, Diversity	Compositional	–
DR-NSP [16]	2016-08	–	Token		Meaningfulness, Diversity	Compositional	–
DA-PTM [23]	2020-12	–	Token		Meaningfulness, Diversity	Replacement (LM)	–
SCDA [8]	2019-07	–	Token		Meaningfulness, Diversity	Replacement (LM)	–
CA-PR [21]	2018-06	–	Token		Meaningfulness	Replacement (LM)	–
Snippext [35]	2020-04	–	Token		Meaningfulness	Replacement (all)	–
EDA [54]	2019-11	–	Token		Meaningfulness	Replacement (all)	–
MaxLoss [17]	2019-03	–	Token		Meaningfulness	Replacement (Synonym)	–
Aug: Dialogue Data Augmentation							
DialAug [42]	2022-10	Generative	Dialogue	Open-domain	Diversity		
CAPT [40]	2022-10	Generative	Dialogue	Open-domain	Diversity	Counterfactual	H, CQ, CA
DlDistill [69]	2020-11	Both Ret. & Gen.	Dialogue	Open-domain	Diversity	Unpair data	–
CVAE-GAN [27]	2019-01	Generative	Dialogue	Open-domain	Diversity	Generative model	H, CQ, CA
CUDA-DST [24]	2022-09	Generative	Dialogue	Task-oriented	Generalization, Diversity	User Dialogue Act (UDA)	H
COCO [28]	2021-03	Retrieval	Dialogue	Task-oriented	Generalization, Diversity	UDA, Counterfactual	H
PARG [9]	2020-07	Generative	Dialogue	Task-oriented	Diversity	Paraphrasing	H, CA
RDA [63]	2020-04	Generative	Dialogue	Task-oriented	Quantity, Diversity	RL-based	P, CQ, CA

Table 2: List of current works for Dialogue Generation: Open-Domain

Paper	Year	tr. Method	Level	Type	Challenges	Approach	Input
ODD Gen: Single-turn generation							
GRADA [33]	2022-10	Generative, Augmentation	QA pair	Open-domain	Informative, Diverse	2, 3, 4	KG
G-DAUG [60]	2020-11	Generative	QA pair	–		2, 3, 4	–
PAQ [26]	2021	Hybrid-GR	QA pair	Open-domain		1, 2, 3, 4	KB
TQA-SD [43]	2020-02	Generative	QA pair	–		2, 3, 4	–
Alberti19 [1]	2019-07	Generative	QA pair	–		2, 3, 4	P
Lewis19 [25]	2019-07	Generative	QA pair	–		2, 3, 4	P
ODD Gen: Multi-turn generation							
Baize [56]	2023-04	Prompting	Dialogue	Open-domain			
SIMSEEK [18]	2022-12	Generative	Dialogue	Open-domain	Blinded to the answer-containing passage, rely on background information B	2, 3, 4	H, B, P
GCN [32]	2022-09	Hybrid-GR, RL	Dialogue	Open-domain	Real scenarios	1, 2, 3, 4	
DG2 [55]	2022-09	Generative	Dialogue	Open-domain	Consistency, Sub passage selection	1, 2, 3, 4	P, H
MultiCQAG [14]	2022-11	Generative	Dialogue	Open-domain	Consistency, Question Types	1, 2, 3, 4	H, P
CQAG-AR [15]	2022-10	Generative	Dialogue	Open-domain	Consistency, Answer revision	2, 3, 4	H, P
CoHS-CQG [6]	2022-10	Generative	Dialogue	Open-domain	Consistency, Sub Passage/History selection	1, 2, 3, 4	H, P
ChainCQG [11]	2021-04	Generative	Dialogue	Open-domain	Consistency	1, 2, 3, 4	H
SHSF [44]	2020-11	–	–	Open-domain		–	–
AU-CQG [37]	2019-11	Hybrid-GR	Dialogue	Open-domain	Consistency, Sub passage selection	1, 3, 4	H, P
ReDR [41]	2019-07	Generative	Dialogue	Open-domain	Consistency, Coreference alignment	2, 3	H
CFNet [10]	2019-07	Generative	Dialogue	Open-domain	Consistency, Coreference alignment	1, 3, 4	H, P
ODD Gen: Target-oriented dialogue systems							
COLOR [53]	2023-05	Generative	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Ext.KG, G
HiTKG [38]	2022-10	Generative	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Ext.KG, G
EGRL [57]	2021-01	Generative	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Ext.KG, G
KnowHRL [58]	2020-04	Generative	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Ext.KG, G
TopKG [62]	2022-10	Generative	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Ext.KG, G
CKC [71]	2021-12	Generative	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Ext.KG, G
DKRN [45]	2020-06	Retrieval	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Int.KG, G
TGODC [49]	2019-07	Retrieval	Dialogue	Target-oriented	Consistency, target achievement	1, 2, 3	H, Int.KG, G
TGCP [20]	2022-10		1-turn Dialogue	Target-oriented			
CRG [12]	2022-07		1-turn Dialogue	Target-oriented			
OTTers [47]	2021-08		1-turn Dialogue	Target-oriented			

Table 3: List of current works for Dialogue Generation: Task-oriented

Paper	Year	tr. Method	Level	Type	Challenges	Approach	Input
ToD Gen: Schema-guided generation							
SGD [46]	2020-06	Generative	Dialogue	Task-oriented	Task completion	–	–
M2M [48]	2018-01	Generative	Dialogue	Task-oriented	Task completion	–	–
ToD Gen: Simulator-Agent Interaction							
ICL-US [50]	2023-06	In-Context	Dialogue	Task-oriented	Task completion, Diversity	–	–
Unified-US [52]	2022-12	Generative	Dialogue	Task-oriented	Task completion	–	–
SimulatedChat [36]	2021-11	Hybrid-GR	Dialogue	Task-oriented	Task completion	–	–
JOUST [51]	2021-08	Generative	Dialogue	Task-oriented	Task completion	–	–
TUS [31]	2021-07	Generative	Dialogue	Task-oriented	Task completion	–	–
LaRL [70]	2019-06	Generative	Dialogue	Task-oriented	Task completion	–	–
ToD Gen: E2E dataset creation							
PLACES [4]	2023-05	In-context	Dialogue	Task-oriented	Multi-party conversations		
DIALOGIC [29]	2022-12	In-context	Dialogue	Task-oriented	Controllable Gen.		
LAD [34]	2022-09	Prompting	Dialogue	Task-oriented	ZSH. Generalisation, Diversity		
NeuralWOZ [19]	2021-08	Retrieval	Dialogue	Task-oriented	Task completion, Unseen domain	–	–
VHDA [64]	2020-11	Generative	Dialogue	Task-oriented	Task completion	–	–

Part 1: Introduction

1.1 Conversational Systems

1.2 Problem of Data Scarcity

1.3 Data Augmentation

Duration: 10 min

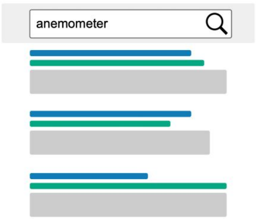
Presenter: —

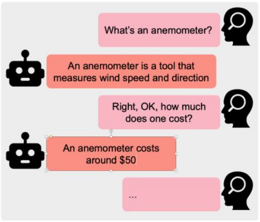
Conversational System

- Age of information explosion
- Conversation applications
  - Seeking relevant information
  - Need ongoing feedback from an agent
  - Engage in conversations for casual chit-chat
  - Engage in conversations for negotiating

Photo: Flight with query

Photo: Flight with Conversation





Types of Dialogue Systems

Open Domain Dialogue System (ODD)

- Aim to perform chit-chat with users without the task and domain restriction
- Engage in conversations with users across a wide range of topics and domains
- Task: generate meaningful and coherent responses based on the user's input and dialogue history
- Challenges:
    - Context awareness: response coherence
    - Response diversity: Avoid making dull responses
    - Controllable generation
        - Conversation topic
        - Mixed-initiative conversation
    - Informative responses: knowledge-grounded system

Types of Dialogue Systems

Task-oriented Dialogue System (ToD)

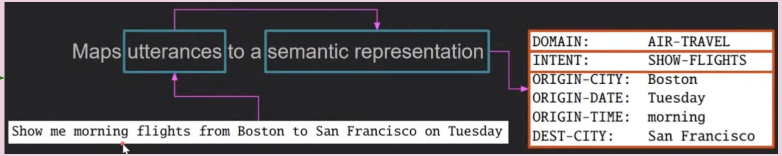
- Accomplish user-requested tasks
  - by accurately understanding the user's intent
  - providing suitable responses
- Domains:
  - Flight ticket / Restaurant / Hotel / Taxi / Movie ticket booking
  - Weather, navigation and scheduling domain
  - Customer service delivery
- Approaches: 1) Pipeline, 2) End-to-end

Goal: You want to book a taxi. The taxi should go to anatolia and should leave after 22:00. The taxi should depart from kymmoy. Make sure you get car type and contact number.
User: i need a taxi to |anatolia|>. **System:** what time would you like to leave? **User:** i would like to leave after |22:00|>. **System:** there are |9|> options available to you. where would you like to leave from? **User:** i 'd like to leave from |kymmoy|>. **System:** the fare estimate is |9.78|>. do you want to book this taxi now? **User:** no, i would like the contact number if possible. **System:** the contact number is |07356725299|> **User:** thank you very much for your help. **System:** you are welcome. do you need anything else today

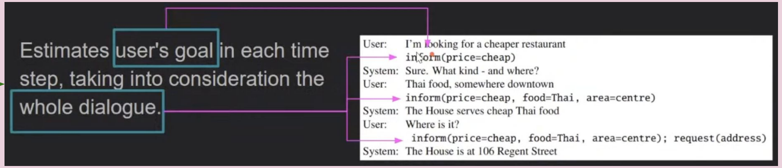
Table 1: A dialogue generated by our model with corresponding natural language user goal. The "|*anatolia*|>" with special tag indicates that the word is a value.

ToD: Pipeline approach

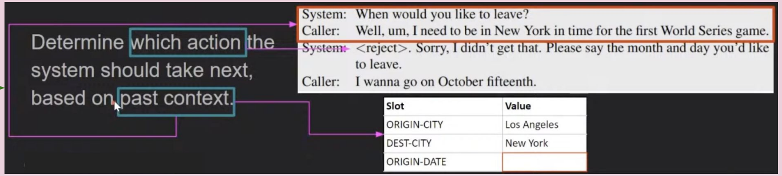
- 1) Natural Language Understanding (NLU)
    - Domain classification
    - Intent classification
    - Slot value Extraction (tagging)



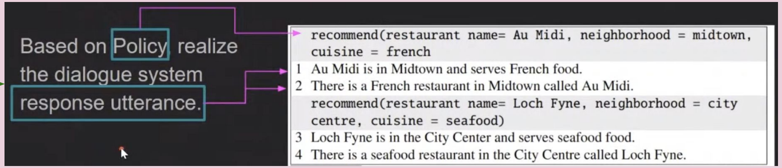
- 2) Dialogue management
    - Dialog state tracking
        - Conducts KB Query, make API calls
        - Provide feedback to DM in order to construct system response



    - Dialog policy optimization
        - RL framework:
            - The DM is the agent
            - The user is the environment
            - NLU are observations made by Agent
            - NLG are actions conducted by Agent
            - Agent tries to maximize Expected Future Rewards



- 3) Natural Language Generation (NLG)
    - Generated response:
        - Adequacy, fluency, readability, variation
    - Generate clarification question: instead of reject response



Problem of data scarcity

- Rapid development of conversation-based models
- Deep learning based
  - High number of trainable parameters
  - Requires a significant amount of training data
- Crowd-sourcing approach
  - Expensive and time-consuming
  - Hard to scale up to new domains
  - Human-authored data possess weaker performance on out of distribution samples
  - Do not necessarily consider the full diversity of question types
- Specific domain / language

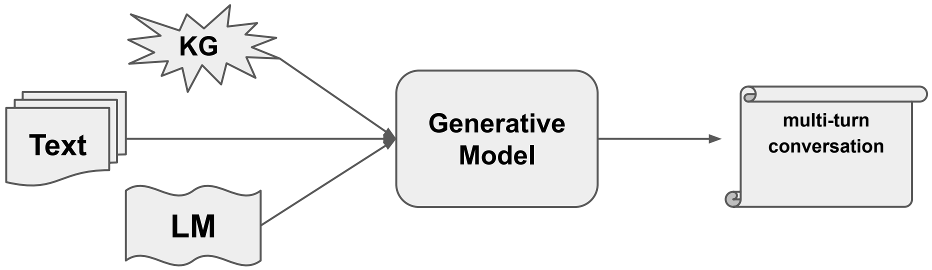
Learning from limited data

- Unsupervised learning: Training with unlabeled data
- Semi-supervised learning: Training with both labeled and unlabeled data
- Transfer learning: Leverage data from a different-but-related task
- Multi-task learning: Use information learned on different tasks for mutual benefit
- Few/Zero shot learning: Generalize to new tasks after seeing a few (or no) examples
- **Data augmentation:** Modify labeled data to with class-preserving transformations

[Chen, et al., 2022]

Data Augmentation (DA)

- Definition: A technique that increases the amount of training data
  - by either slightly modifying existing data or generating new synthetic
- DA Advantages
  - Increase amount of training data
  - Improve the diversity of the dataset
  - Dealing with the overfitting problem
- Data Generation
  - Using external and various resource



Part 2: Conversation Augmentation

2.1 Generic Token-level & Sentence-level Augmentation

2.2 Dialogue Data Augmentation

Duration: 30 min

Presenter: —

Generic Token-level & Sentence-level Augmentation

Overview

- Involve the replacement of original tokens or sentences with relevant alternatives
- Token-level: Modifying a few local words
  - Replacement: Random word, Synonym, Insertion, Deletion, Swapping, MixDA
  - Replacement via LM: Contextual augmentation
  - Compositional Augmentation
- Sentence-level: Modifying an entire sentence at once
  - Paraphrasing
  - Back-Translation

Token-level: Replacement

- **EDA:** Consists of four simple but powerful operations
  - Synonym Replacement (SR)
  - Random Insertion (RI)
  - Random Swap (RS)
  - Random Deletion (RD)
- **Sniptext**: Replacement analysis
- Problems:
  - A naive application of swap or delete may leave the sequence with an inconsistent state of tags
  - Replace or insert can change the meaning of tokens and make the original tags invalid

Operation	Sentence
None	A sad, superior human comedy played out on the back roads of life.
SR	A ***lamentable***, superior human comedy played out on the ***backward*** road of life.
RI	A sad, superior human comedy played out on ***funniness*** the back roads of life.
RS	A sad, superior human comedy played out on ***roads*** back ***the*** of life.
RD	A sad, superior human out on the roads of life.

Table 1: Sentences generated using EDA. SR: synonym replacement. RI: random insertion. RS: random swap. RD: random deletion.

[] EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks, EMNLP-IJCNLP, 2019

Token-level: Replacement

Operator	Description
TR	Replace **non-target token** with a new token.
INS	Insert before or after a **non-target token** with a new token.
DEL	Delete a **non-target token**.
SW	Swap two **non-target tokens**.
SPR	Replace a **target span** with a new span.

- **MIXDA:** Augment and Interpolate
- Types of tokens: 1) Target spans, 2) Non-target tokens
- Sampling & Post-sampling step: Determine a new token (or span) to insert or replace the original one
  - Uniform sampling
  - Importance-based sampling
  - Semantic Similarity (post-sampling only)

$$
\begin{aligned}
\mathrm{BERT}(x') &= \lambda \cdot \mathrm{BERT}(x_1) + (1-\lambda) \cdot \mathrm{BERT}(x_2) \\
y' &= \lambda \cdot y_1 + (1-\lambda) \cdot y_2
\end{aligned}
$$

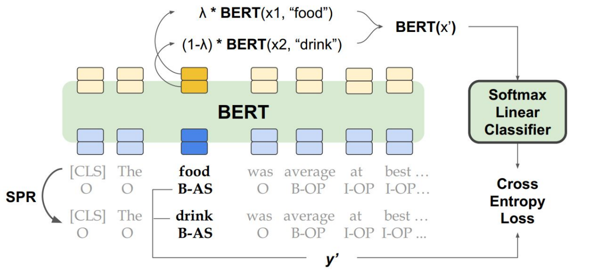
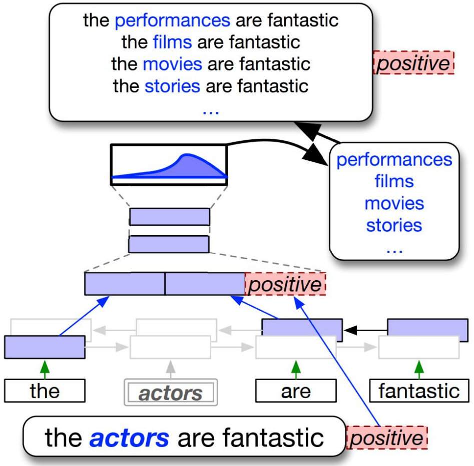


Figure 5: MixDA augments by interpolating over the BERT encoding.

[] Snippext: Semi-supervised Opinion Mining with Augmented Data, WWW, 2020

Token-level: Replacement via LLM

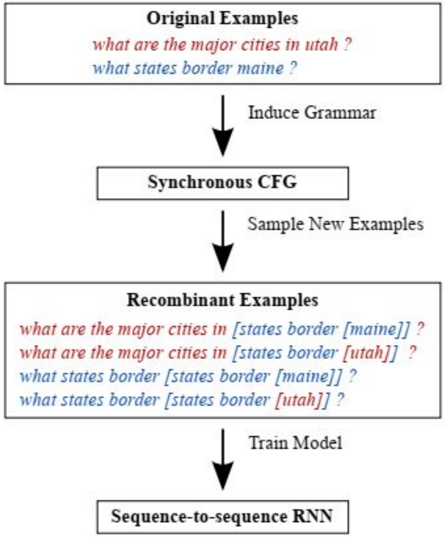
- Contextual augmentation: Replace the sampled words given their context
- Instead of the synonyms, we use words that are predicted by a LM given the context surrounding the original words to be augmented
- Conditional DA using Pre-trained LM [DA-PTM]
- Hard sampling [CA-PR]
- Soft sampling [SCDA]



Contextual augmentation: Data augmentation by words with paradigmatic relations, NAACL, 2018
Soft Contextual Data Augmentation for Neural Machine Translation, ACL, 2019

Token-level: Compositional

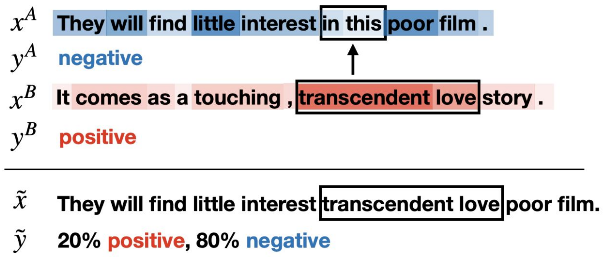
- Recombining various fragments from different sentences
- Requires more carefully designed rules
- Potential to greatly improve the generalization abilities to out-of-distribution data
- Methods:
  - Synchronous Context-Free Grammar (SCFG) [DR-NSP]
  - lexical environment [GECA]

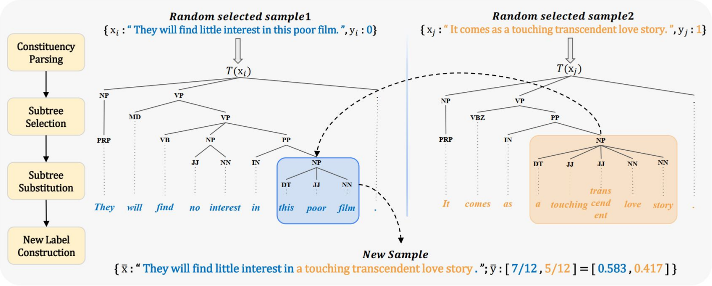


An Empirical Survey of Data Augmentation for Limited Data Learning in NLP, ACL, 2021

Token-level: Compositional

- Saliency-Based Span Mixup (SSMix)
  - Performed on input text rather than on hidden vectors
- TreeMix
  - Constituency parsing tree
    - Decompose sentences into constituent sub-structures
  - Mixup data augmentation technique
    - Recombine them to generate new sentences

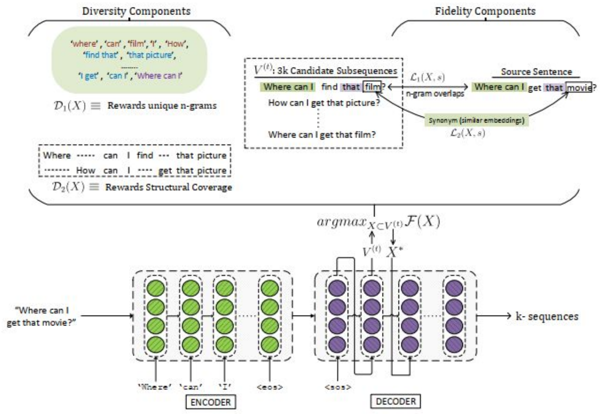




SSMix: Saliency-Based Span Mixup for Text Classification, ACL-IJCNLP 2021
TreeMix: Compositional Constituency-based Data Augmentation for Natural Language Understanding, NAACL, 2022

Sentence-level: Paraphrasing

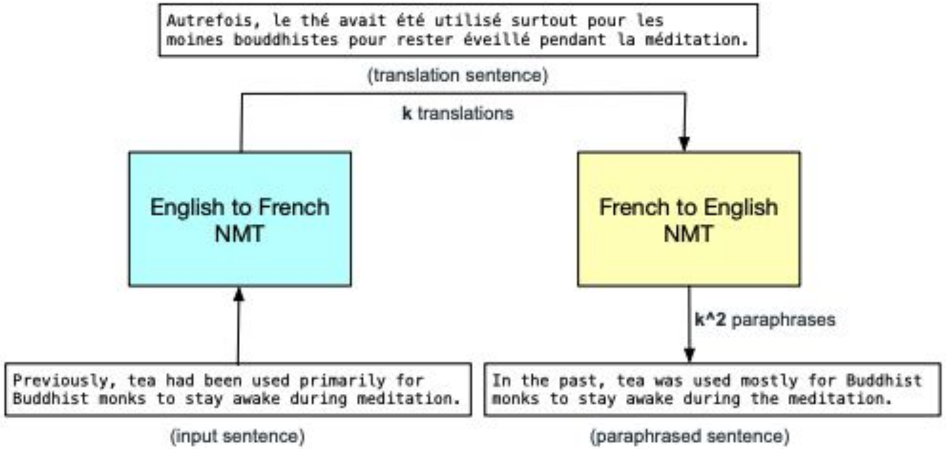
- Paraphrasing: Directly training end-to-end models to generate paraphrases
  - Generating semantically similar paraphrases, paying little attention towards diversity [top-k beam search]
  - Diverse Paraphraser using Submodularity [DiPS]



Submodular Optimization-based Diverse Paraphrasing and its Effectiveness in Data Augmentation, NAACL, 2019
Multilingual Transfer Learning for QA using Translation as Data Augmentation, AAAI, 2021

Sentence-level: Back-translation

- Translates a sentence into another language and then back again



Multilingual Transfer Learning for QA using Translation as Data Augmentation, AAAI, 2021

Dialogue Data Augmentation (DDA)

Overview

- Open Domain dialogue system
  - Generative: Generating different questions and answers given a dialogue data [CVAE]
  - Using unpair data: Dialogue Distillation
  - Counterfactual: [CAPT]
  - Mixing up Dialogue Contexts [DialAug]
- Task-oriented dialogue system
  - RL-based: Augmentation for dialogue state tracking (DST) module [RDA]
  - Paraphrasing: Response generation [PARG]
  - User Dialogue Act: Controllable user dialogue act augmentation [CoCo, CUDA-SDT]

DDA: Generative approach

- Type of dialogue: Open-domain, Query-Response
- CVAE: generating sufficient alternative expressions
- Discriminator: improve the relevance of each augmented query-response pair
- Distillation strategy: filter repetition query-response pairs
- Three different schemes for enhancing dataset
  - One-to-many (1-n)
  - Many-to-one (n-1)
  - Many-to-many (n-n)

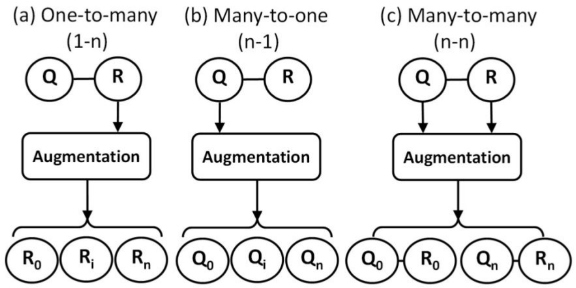
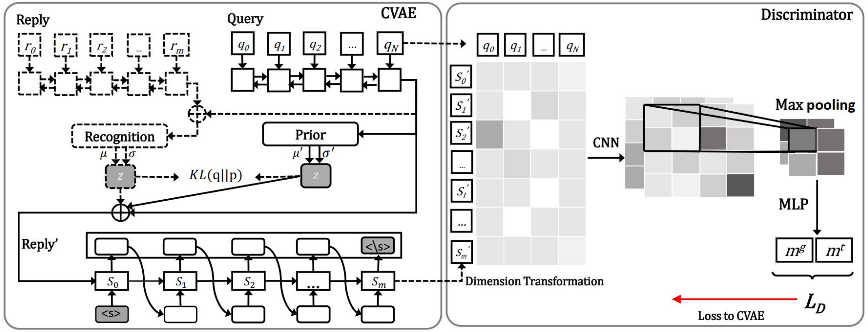


Figure 4 Three different paradigms of data augmentation for open-domain dialogue generation

Insufficient Data Can Also Rock! Learning to Converse Using Smaller Data with Augmentation, AAAI, 2019

DDA: Generative approach (cont.)

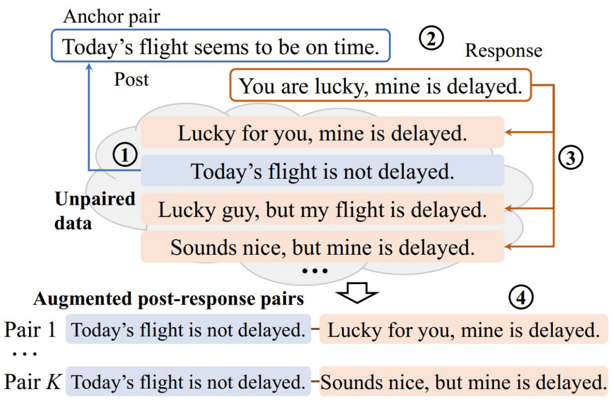
- Generator CVAE:
  - Encoder $(x, c) \mapsto z$. $q_\theta(z \mid x, c)$
  - Decoder $(z, c) \mapsto x$. $p_\theta(x \mid z, c)$,
- Discriminator
  - calculate the interaction (or matching) matrix
  - (CNN) to extract features
  - an adversarial fashion for training
- Filtering: Jaccard distance
  - depict the word-level semantic similarity



Insufficient Data Can Also Rock! Learning to Converse Using Smaller Data with Augmentation, AAAI, 2019

DDA: Using unpair data

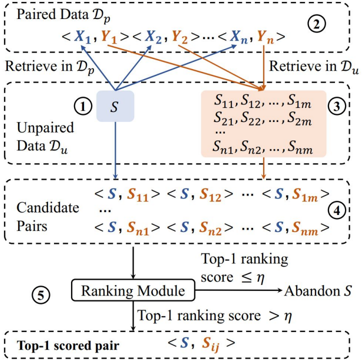
- Type of dialogue: Open domain, post-response pairs
- High-quality unpaired data are generally easier to collect
- The method involves two phases of distillation
  - Data level: constructs post-response pairs by matching sentences retrieved from a set of unpaired data
  - Model-level: distills a teacher model using the augmented data



Dialogue Distillation: Open-Domain Dialogue Augmentation Using Unpaired Data, EMNLP, 2020

DDA: Using unpair data, Data-level

- Constructing a set of new post-response pairs
- 1) Constructing candidate pairs
    - 1) The sentence S is randomly selected in the unpaired data Du.
    - 2) A set of posts X1; : : : ; Xn that are similar to S are retrieved from the paired data Dp
    - 3) Each corresponding response Yi is then used to retrieve m sentences Si1; : : : ; Sim that are similar to Yi from Du.
    - 4) Then n × m candidate pairs can be formed by grouping S with each sentence: hS; Siji, (i = 1; : : : ; n,
    - j = 1; : : : ; m)
- 2) Filtering low-quality candidates
    - ranking module is used to rank these candidate pairs
    - score function as a text matching model
    - only extract the top-1 scored pair hS; Siji among all its n × m candidate pairs



Dialogue Distillation: Open-Domain Dialogue Augmentation Using Unpaired Data, EMNLP, 2020

DDA: Using unpair data, Model-level

- Training with both original and augmented data
- The augmented pairs in Da might not be as high-quality as these human-crafted pairs in Dp.
- Prevent the dialogue models from being affected by the noise in Da
- Retrieval-based Dialogue Model
  - A matching function that predicts whether a response Y matches a given post X
    - Optimizing a negative log likelihood (NLL) loss

$$\mathcal{L}_{m-nll}(\theta) = -(1-l)\log\mathcal{P}_\theta(0|X,Y) - l\log\mathcal{P}_\theta(1|X,Y)$$

  - A teacher model
    - Teacher model is fixed
    - Used to compute a knowledge distillation (KD) loss

$$\mathcal{L}_{m-kd}(\theta) = -\sum_{i=0}^{1}\mathcal{P}_{\theta_t}(i|X,Y)\cdot\log\mathcal{P}_\theta(i|X,Y)$$

  - The final matching model is trained on the following loss

$$\mathcal{L}_M(\theta) = \mathcal{L}_{m-nll}(\theta) + \alpha_m\mathcal{L}_{m-kd}(\theta),$$

- Generation-based Dialogue Model
  - Try to capture the distribution of the response sentences Y given the post sentence X

Dialogue Distillation: Open-Domain Dialogue Augmentation Using Unpaired Data, EMNLP, 2020

DDA: Counterfactual

- Type of dialogue: Open domain, post-response pairs
- RQ: Given an observed dialogue, what the response would happen if we change the reply perspective, while keeping the current environment unchanged?
- Background: Structural Causal Model
    - Definition
    - Intervention
    - Counterfactual Inference

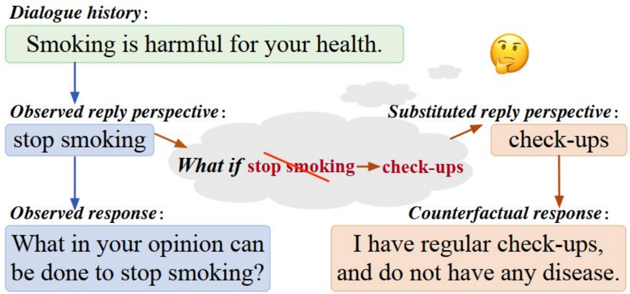
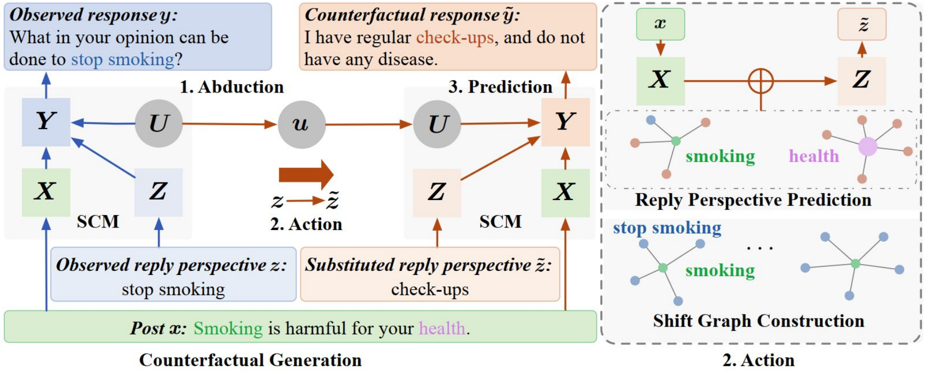


Figure 1: An example of a counterfactual response, which is a semantically different response re-inferred by changing the observed reply perspective.

Counterfactual Data Augmentation via Perspective Transition for Open-Domain Dialogues, EMNLP, 2022

DDA: Counterfactual (cont.)

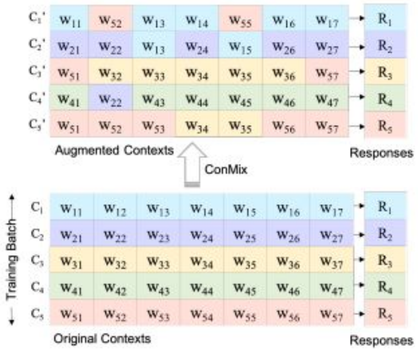
- Counterfactual Generation via Perspective Transition
  - Abduction: Estimate the unobserved variable given the observed sample
  - Action: Replace the observed reply perspective z with a substituted reply perspective
  - Prediction: Generate the counterfactual response given the posterior sample
  - Model Training: Reply perspective predictor and the counterfactual generator

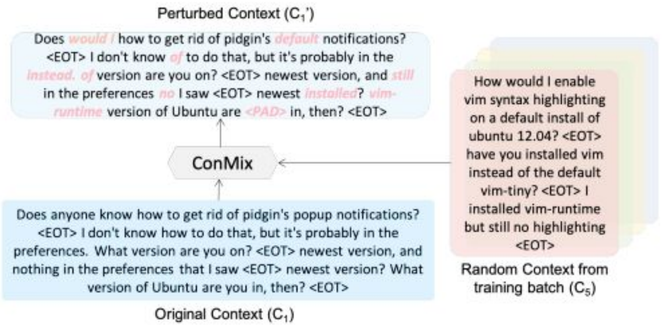


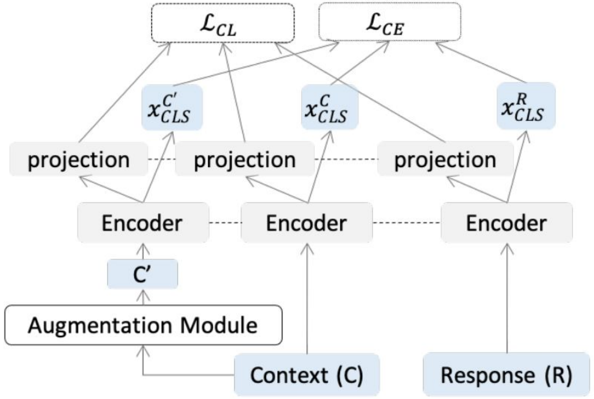
Counterfactual Data Augmentation via Perspective Transition for Open-Domain Dialogues, EMNLP, 2022

DDA: Mixing up Dialogue Contexts

- ConMix DA method: creates augmentations through dynamic mixing of words from other contexts in the batch
- Adapt the Bernoulli MixUp approach



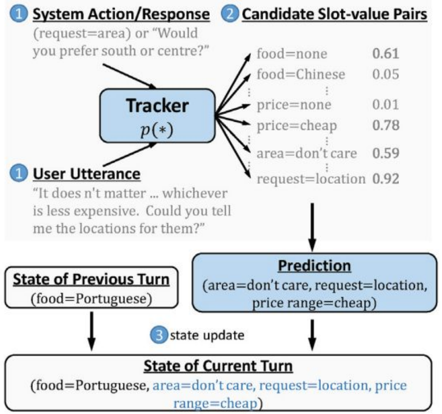




DialAug: Mixing up Dialogue Contexts in Contrastive Learning for Robust Conversational Modeling, COLING, 2022

DDA: RL-based

- Tracker Module: Track the user's goal during the dialog process
- Generator Module: Formulate DA as an optimal text span replacement problem
- Reward Design
  - Bag-level reward: Re-train the Tracker with each sampled bag
  - Instance-level reward: Evaluate each generated instance in the bag

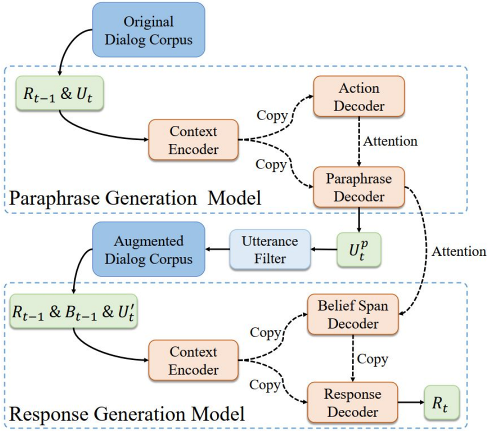


Dialog State Tracking with Reinforced Data Augmentation, AAAI, 2020

DDA: Paraphrasing

- Type of dialogue: Task-oriented, Augmenting next user utterance
- PARG model:
  - 1) Paraphrase Data Construction
  - 2) Paraphrase Augmented Response Generation
    - Paraphrase Generation Model
    - Paraphrase Filter
    - Response Generation Model

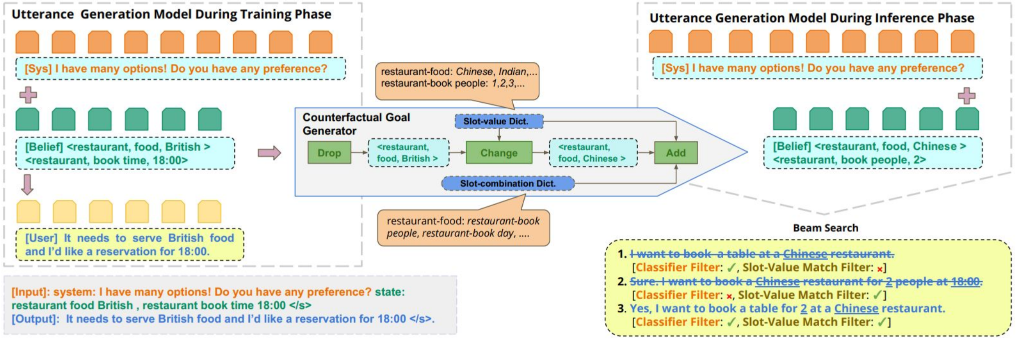
$$loss = loss_a + loss_p + loss_b + loss_r$$



Paraphrase Augmented Task-Oriented Dialog Generation, ACL, 2020

DDA: User Dialogue Act, counterfactual approach

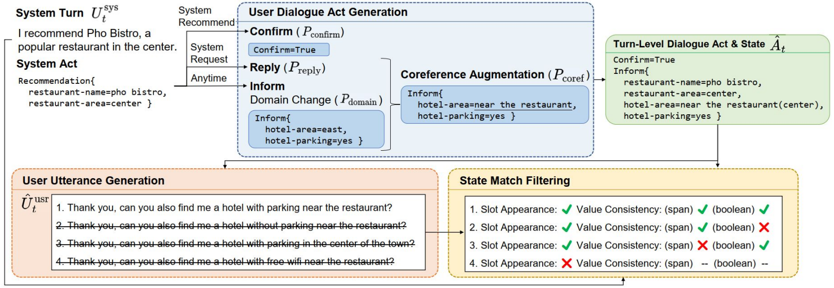
- Type of dialogue: Task-oriented, Augmenting next user act
- COCO consists of three main pillars:
  - Value Substitution
  - Controllable Counterfactual Generation
    - Operation
    - Slot-Value Dictionary
    - Slot-Combination Dictionary
  - Filtering
    - Slot-Value Match Filter
    - Classifier Filter

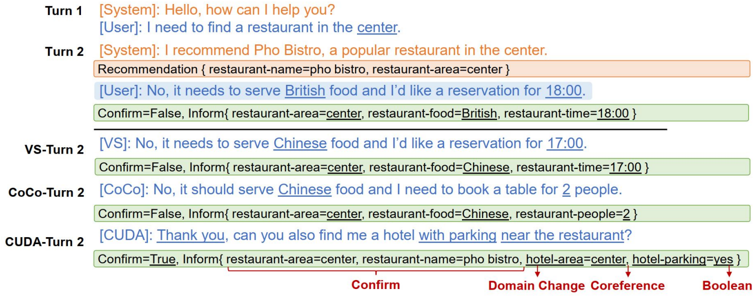


CoCo: Controllable Counterfactuals for Evaluating Dialogue State Trackers, ICLR, 2021

DDA: User Dialogue Act, user’s behavior

- Type of dialogue: Task-oriented, Augmenting next user act
- A user dialogue act represents the core meaning of the user’s behavior
- Controllable User Dialogue Act Augmentation (CUDA)
  - User Dialogue Act Generation
  - User Utterance Generation
  - State Match Filtering





Controllable User Dialogue Act Augmentation for Dialogue State Tracking, SIGDIAL, 2022

Token-level & Sentence-level Shortcomings

- Provide variety to the training data without greatly altering the semantics of the original sentences
  - Consistency between turns and coreference
- Only operate on the existing dialogue data
  - Fail to take advantage of the available external resources
  - Not suitable for low-resource languages and domains
- No Control on conversation flow
  - Cannot generate different forms of questions
    - Open-ended / Close-ended / unanswerable
  - Clarification question
- Next Idea: Using Generation process

Part 3: Conversation Generation, ODD

3.1 Single-turn QA Pair Generation

3.2 Multi-turn Dialogue Generation

3.3 Topic-aware Dialogue Agent

Duration: 80 min
Presenter: —

Single-turn QA pair Generation

Overview

Input (C)		... in 1903, boston participated in the first modern world series, going up against the pittsburgh pirates ...
(1)	$C \rightarrow A$	1903
(2)	$C, A \rightarrow Q$	when did the red sox first go to the world series
(3)	$C, Q \rightarrow A'$	1903
(4)	$A \stackrel{?}{=} A'$	Yes

Table 1: Example of how synthetic question-answer pairs are generated. The model's predicted answer (A') matches the original answer the question was generated from, so the example is kept.

- Creating new Question-Answer pair based on Passage
- Main goal: Informative and Consistent Question-Answer pair
- Principles of Pipeline method
  - Passage selection: identifying passages which humans are likely to ask questions about
  - Answer Extraction: an answer is extracted from the selected passage
  - Question Generation: generating a question based on the passage C and the extracted answer
  - Filtering: to ensure the accuracy and quality of the generated question-answer pair

Synthetic QA Corpora Generation with Roundtrip Consistency, ACL, 2019

Passage Selection

- Used to find passages that are likely
  - To contain information that humans may ask about,
  - Make good candidates to generate questions from
- Dense Passage Retrieval" method
  - Training passage and question Encoders
  - At run-time, calculate the similarity
  - For training the encoder: Metric learning

$$\mathrm{sim}(q, p) = E_Q(q)^\intercal E_P(p).$$

$$L\left(q_i, p_i^+, p_{i,1}^-, \cdots, p_{i,n}^-\right) = -\log \frac{e^{\mathrm{sim}\left(q_i, p_i^+\right)}}{e^{\mathrm{sim}\left(q_i, p_i^+\right)} + \sum_{j=1}^{n} e^{\mathrm{sim}\left(q_i, p_{i,j}^-\right)}}$$

Dense Passage Retrieval for Open-Domain Question Answering, EMNLP, 2020
PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them, TACL, 2021

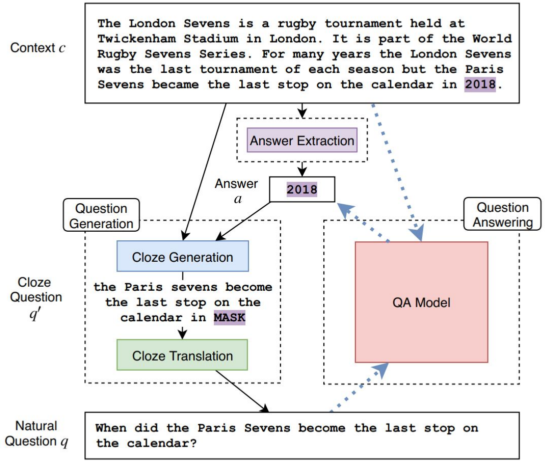
Answer Extraction

- Given a passage, identify spans that are likely to be answers to questions
- Types:
    - Question Unconditional Extractive [Alberti]
    - Question Conditional Extractive [Alberti]
    - NER answer extractor [PAQ][Lewis]

PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them, TACL, 2021
Synthetic QA Corpora Generation with Roundtrip Consistency, ACL, 2019

Question Generation

- Given a passage and an answer, generate likely questions with that answer
- Types:
    - As a left-to-right language model [Alberti][PAQ]
    - By Cloze Translation [Lewis19]



Unsupervised question answering by cloze translation, ACL, 2019
Synthetic QA Corpora Generation with Roundtrip Consistency, ACL, 2019

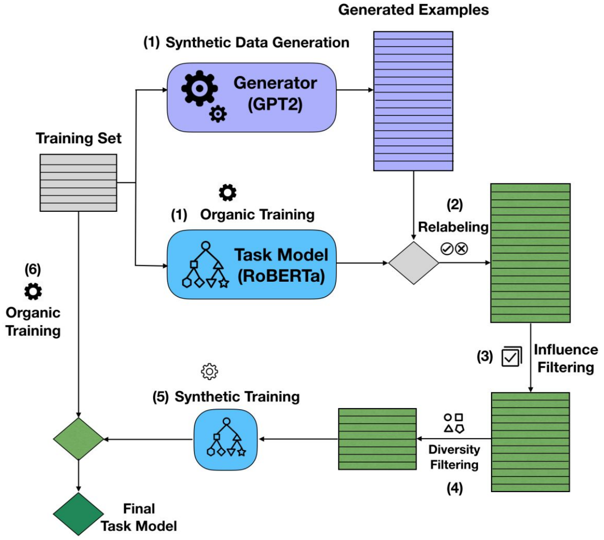
Filtering

- Improves the quality of generated questions
  - By ensuring that they are consistent
  - The answer they were generated is likely to be a valid answer to the question
- Types:
  - Local filtering [Alberti, RTQA]
    - Based on the selected passage ot sub-passage
    - will not remove questions that are ambiguous
  - Global filtering [PAQ]
    - Based on the whole corpus

PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them, TACL, 2021
Synthetic QA Corpora Generation with Roundtrip Consistency, ACL, 2019

Multiple choice QA

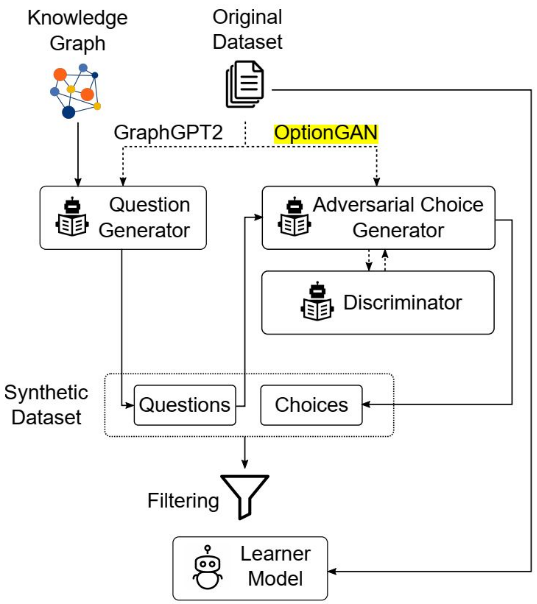
- Selects the most informative and diverse set of examples for data augmentation
- Synthetic Training Data Generation
  - Generating Synthetic Questions
  - Generating Synthetic Answers and Distractors
  - Data Relabeling
- Filtering: Synthetic Data Selection
  - Filtering with Influence Functions
  - Selecting Diverse Examples
- Training with Synthetic Data
  - Synthetic Training
  - Organic Training



Generative Data Augmentation for Commonsense Reasoning, EMNLP, 2020

Multiple choice QA (KG-based)

- GRADA: a graph-generative data augmentation framework
    - To synthesize factual data samples from knowledge graphs
    - For commonsense reasoning datasets
- Graph-to-Text Generation
    - Linearized Graph Input
    - GraphGPT2 for Structure-aware Graph Input
- Answer & Distractor Generation
- Filtering and Selection of Samples

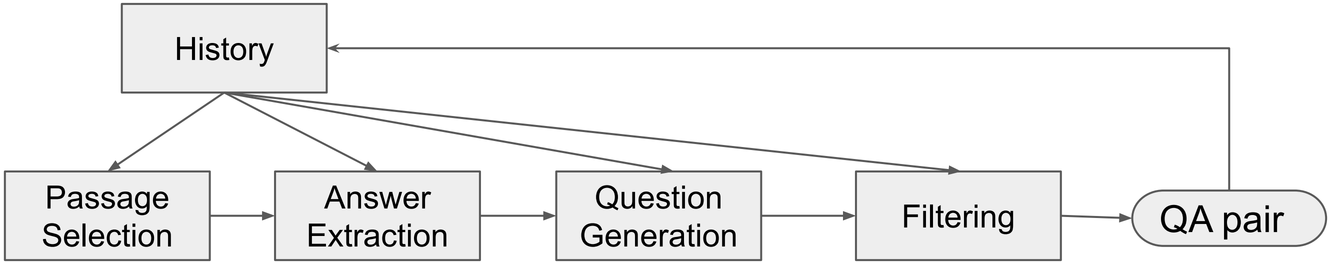


GRADA: Graph Generative Data Augmentation for Commonsense Reasoning, COLING, 2022

Multi-turn Conversation Generation

Overview

- Differences with single-turn QA
  - The history of the conversation are also fed to model
  - The generated QA is added to conversation history
  - The use of some components has been slightly changed
- Additional challenges
  - deeper understanding of the context and the dialogue history
  - specific challenges unique to conversational settings, such as coreference alignment and conversation flow



Passage Selection

- Modified to selecting a sub passage
- 1) Rank the passages [DG2]: A document can often be very long
    - Contrastive loss
- 2) Shortening the context and history for the model
    - Highlight a sentence of the given passage that contains answer rationale, in answer-aware [ChainCQG]
    - Context and History Selection (CoHS) [CoHS-CQG]
        - Naive solution: only simply selected the last k turns
        - CoHS: select the top-p of sentences and QA pairs from C and Hn

$$minimize(u + k)$$

$$\sum_{i=v-u}^{v-1} \sum_{j=n-k}^{n-1} \mathbf{T}[i][j] \geq p$$

$$(q_{n-1}, a_{n-1}) \in H_{sub}, c_s \in C_{sub}$$

Passage Selection (cont.)

- 3) Guarantee & Control the flow of the generated conversation
  - Assumption: A passage usually starts with some general concepts and as going through, it gets in details
  - Split the passage in some chunks [CFNet, AU-CQG]
  - Document Positional Information [DG2]

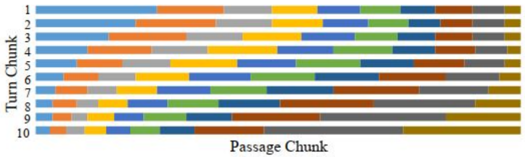
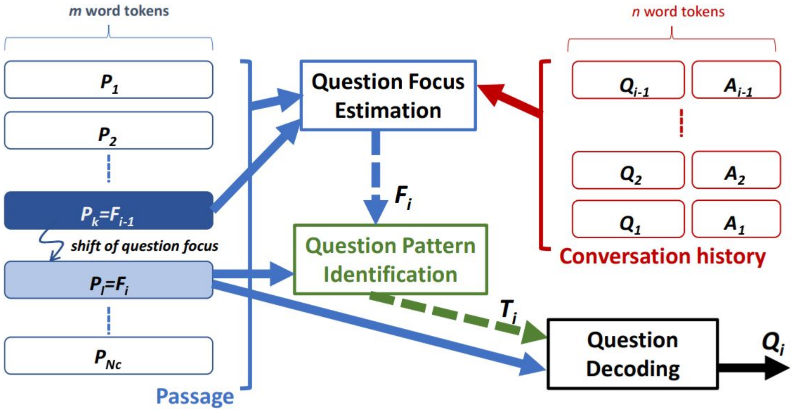


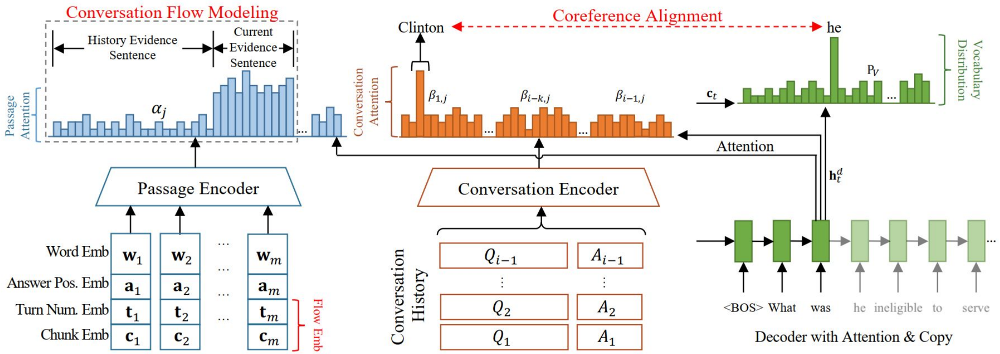
Figure 1: Passage chunks of interest for each turn chunks. Each row contains 10 bands distinguished by different colors. Each band represents a passage chunk. The width of a passage chunk indicates the concentration of conversation in that turn. The y-axis indicates turn chunk number. Same passage chunks share the same color across different turn chunks. *(Best viewed in color)*



Conversation Flow

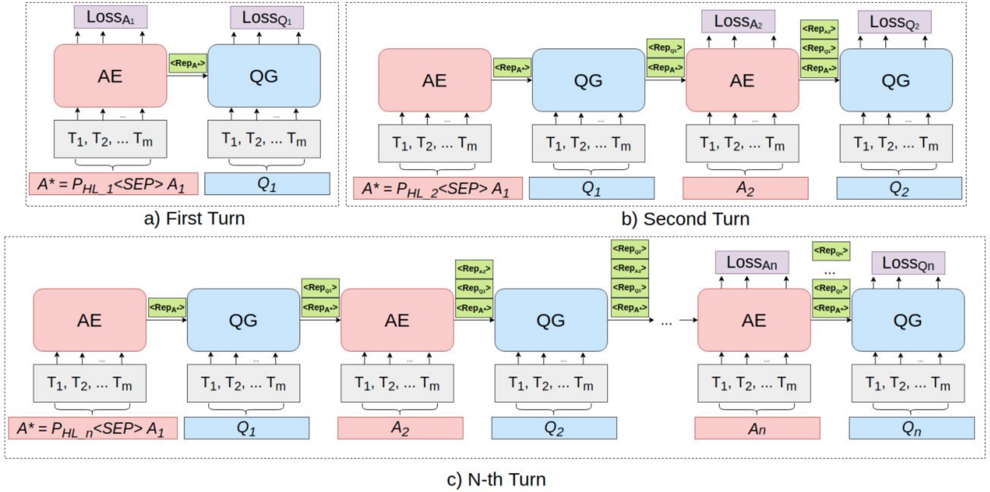
- Mainly by filtering sub-module
- Flow embedding [CFNet]
- Flow loss [CFNet]
- Conversation Flow Modeling Analysis [CFNet]

$$\mathcal{L}_{\text{flow}} = -\lambda_3 \log \frac{\Sigma_{j:w_j \in \mathtt{CES}} \alpha_j}{\Sigma_j \alpha_j} + \lambda_4 \frac{\Sigma_{j:w_j \in \mathtt{HES}} \alpha_j}{\Sigma_j \alpha_j}$$



Conversation Flow (Cont.)

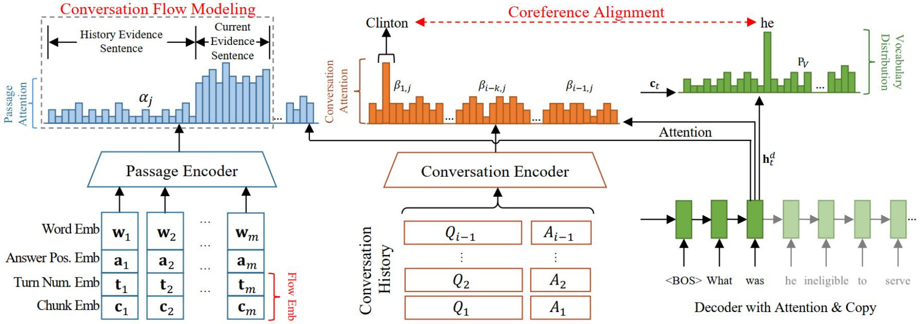
- Flow Propagation-Based Training [ChainCQG]



Coreference Alignment

- Coref loss function
- History as an input
- Coreference Alignment Analysis

$$\mathcal{L}_{\text{coref}} = -(\lambda_1 \log \frac{\Sigma_j \beta^c_j}{\Sigma_{k,j} \beta_{i-k,j}} + \lambda_2 \log p_{\text{coref}}) * s_c,$$



Passage: … however , mccain has a very different life story . he grew up in a navy family and was a pilot during the vietnam war in the 1960s …

Conversation History:

<q>	what	war	was	mccain	in	?
0.0000	0.0001	0.0049	0.0138	0.7710	0.0055	0.0069
<a>	vietnam	war				
0.0000	0.0140	0.0095				
<q>	was	he	in	the	army	?
0.0000	0.0045	0.1303	0.0005	0.0139	0.0001	0.0250
<a>	no					
0.0000	0.0000					

Question (Human): what was his job ?

Question (Our Model): what was his job ?

Passage: … incumbent democratic president bill clinton was ineligible to serve a third term due to term limitations in the 22nd amendment of the constitution …

Conversation History:

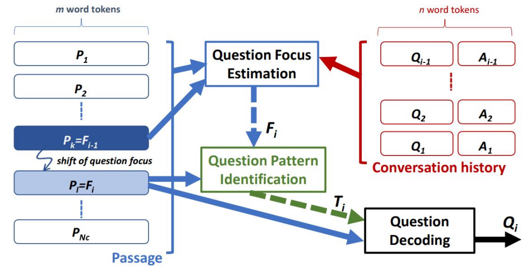
<q>	what	political	party	is	clinton	a
0.0000	0.0000	0.0002	0.0063	0.0045	0.9260	0.0430
member	of	?	<a>	democratic		
0.0008	0.0006	0.0026	0.0000	0.0160		

Question (Human): what was he ineligible to serve ?

Question (Our Model): what was he ineligible for ?

Question Types

- Question Pattern Identification [AU-CQG]
  - Question Pattern Classification
  - Question Pattern Generation



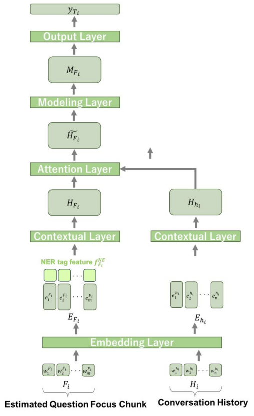


Figure 3: Question pattern classification model.

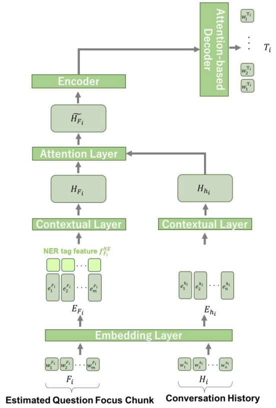


Figure 4: Question pattern generation model.

Question Types (cont.)

- Hierarchical Answerability Classification [MultiCQAG]
    - filters out invalid Q–A pairs and acquires unanswerable questions
    - determines whether a question can be answered based on the passage
    - If not, the module replaces the answer of an unanswerable question with "unknown"
- Classify synthetic questions into three categories:
    - 1) Answerable in correct context
    - 2) Answerable in different context
    - 3) Unanswerable question
- Modeling

Filtering unqualified pairs & Answer Revision

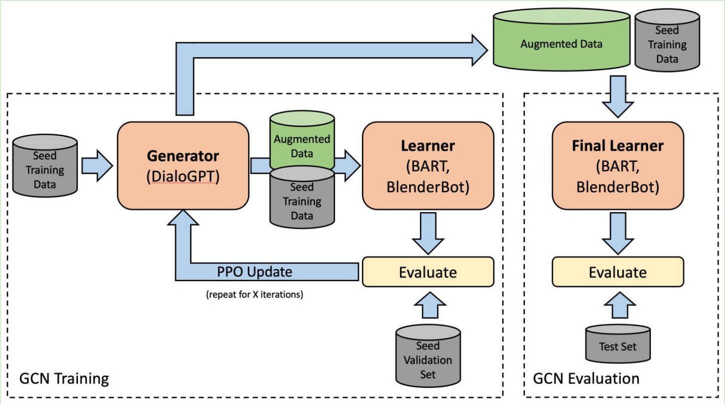
- Discarding unqualified pairs
  - Question Filtering [CoHS-CQG]
  - Roundtrip Filtration for CQA [SIMSEEK]
- Revision module:
  - First considers that the extracted answer span is proper for use as an answer
  - Then modifies it if it is inappropriate
- Two negative sampling techniques:
  - To collect improper answer spans from proper ones
  - Generating Negative Samples
    - 1) Answer Span Expansion
    - 2) Answer Span Reduction
  - Modeling

Blinded to the answer-containing passage

- Rely on background information B [SIMSEEK]
- Two opposite ways to simulate synthetic conversations from unlabeled documents
  - SimSeek-sym
  - SimSeek-asym

Generative and RL based

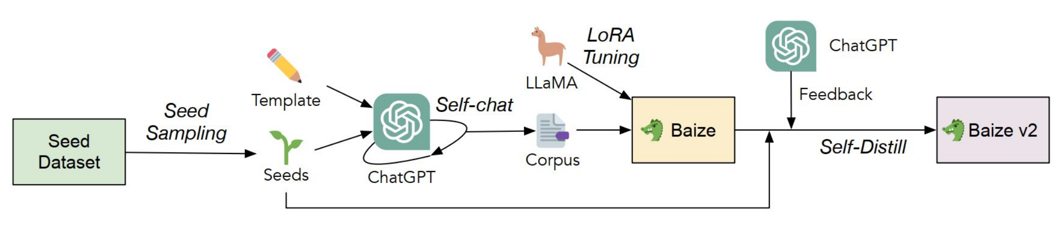
- T



Knowledge-Grounded Conversational Data Augmentation with Generative Conversational Networks, SIGDIAL, 2022

Prompting

- T



Baize: An Open-Source Chat Model with Parameter-Efficient Tuning on Self-Chat Data, EMNLP, 2023

Multi-turn Dialogue Generation: Summary

- The main challenges:
  - Sub-passage selection
  - Flow consistency
  - Coreference alignment
  - Question types & validation
  - Blinded to the answer-containing passage
- Text-based approach
- No annotation needed
  - Just the consistency and correctness are evaluated

Topic-aware Dialogue Agent

- No controllability in the pipeline method
- Proactive dialogue systems
  - Add more control over the conversations
  - Proactiveness: the agent introducing new elements into the conversation
- How can this type of dialogue system be beneficial in generating full conversations?
  - Having a high-level plan or strategy
  - Planning techniques that involve defining paths on a knowledge graph
- Tasks
  - Incorporating mix-initiative in control flow & facilitating topic shifting
  - asking clarifying questions,
  - suggesting useful search queries,
  - integrating chit-chat with task-oriented interactions,

One-turn Topic Transition

- Task definition: Identifying a bridging path of entities to link the context and the target
- OTTers
    - Incorporates a reasoning module
    - Connects the two sentences using a Knowledge Graph (KG) path
        - The outputs can be more controlled and interpretable
- Sub-task for mixed-initiative dialogue system

User A	**Source Topic**:	I spend a lot of time **outside**.
User B	**Transition**:	I like the outdoors as well, especially **gardening**. It destresses me.
	Target Topic:	I enjoy relaxing and getting **flowers**.
	Entity Path:	**outside** - **garden** - **flower**
User A	**Source Topic**:	I like **seafood** a lot.
User B	**Transition**:	Since you like seafood, is **Swedish fish** a candy that you might enjoy?
	Target Topic:	I have no self control when it comes to **candy**.
	Entity Path:	**seafood** - **Swedish fish** - **candy**
User A	**Source Topic**:	I think I am getting **engaged** soon.
User B	**Transition**:	I have two children from a previous **marriage**
	Target Topic:	My **children** are my life.
	Entity Path:	**engagement** - **marriage** - **child**

OTTers: One-turn Topic Transitions for Open-Domain Dialogue,

Commonsense Path and Data

- Two approaches for the transition response generation task
  - 1) Commonsense-guided response generation (CRG)
  - 2) Data augmentation to tackle data sparsity

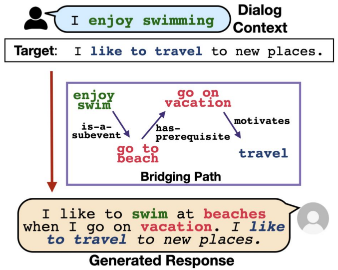
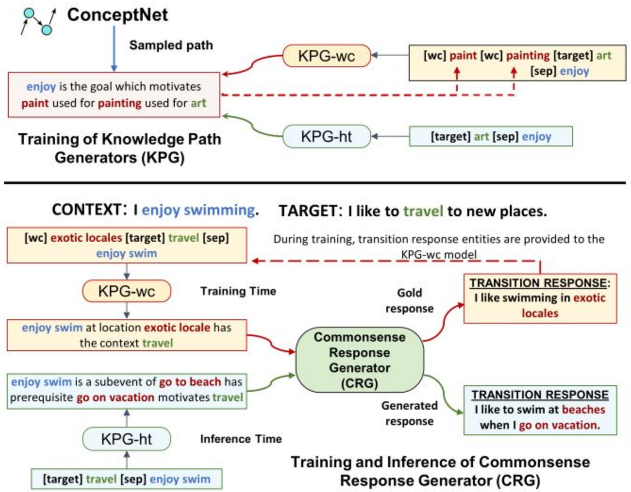


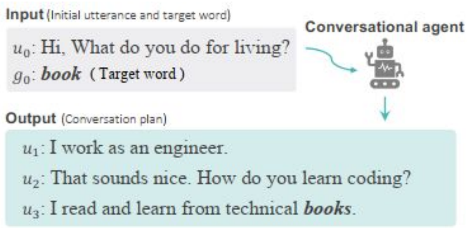
Figure 1: Given a dialogue context and a target sentence, our goal is to generate a dialogue response that smoothly transitions the conversation from context towards the target. Our proposed approach involves identifying a bridging path of entities to link the context and the target.

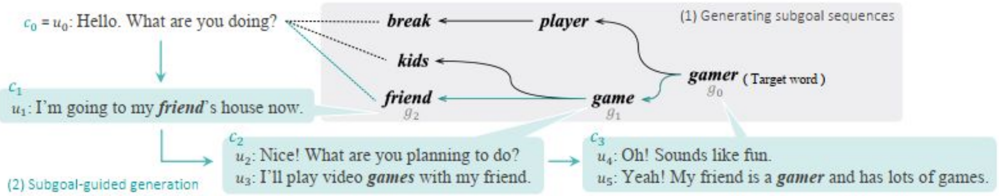


Target-Guided Dialogue Response Generation Using Commonsense and Data Augmentation

Target-Guided Open-Domain Conversation (TGCP)

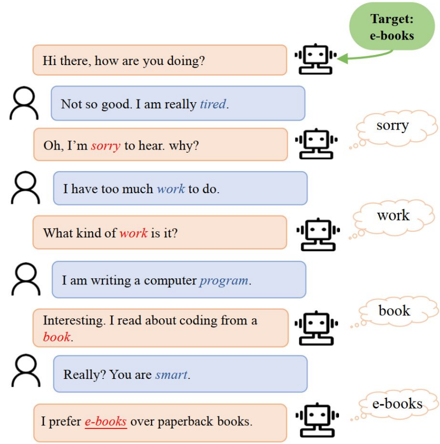
- Definition: Evaluate whether neural conversational agents have goal-oriented conversation planning abilities





Target-Guided Dialogue Response Generation Using Commonsense and Data Augmentation

Target-oriented Dialogue System



- Task Definition
  - Chat naturally with human
  - Proactively guide the conversation to a designated target subject
- Types of goal
  - Designated target topic: A word (an named entity, or a common noun)
  - Pre-defined topic thread
- The challenge of this problem: how to balance the tradeoff between
  - Transition smoothness: maximizing transition smoothness between turns in the dialogue
  - Target achievement: minimizing the number of turns taken to reach the target
- Sub-problems:
  - Next turn keyword selection
  - Keyword-augmented response retrieval / generator
- Dataset problem for this task
  - Add the proposed challenge to existed dataset

Target-Guided Open-Domain Conversation, ACL, 2019

Keyword Extractor

- Rule-based keyword extractor
  - Combines TF-IDF and Part-Of-Speech features for scoring word salience
- Steps:
  - 1) Regards each utterance in a conversation as a document
    - each word as a term to calculate the TF-IDF value of each word
  - 2) Ignores the words
    - Appearing less than 10 times in all corpus
    - Have been mentioned in the former utterance of the current conversation
  - 3) Sets different weights to distinguish the importance of each part of speech
  - 4) Multiplies the TF-IDF value with the part-of-speech weight
  - 5) Considers the words left as the keywords of an utterance in an conversation
- Pros:
  - Simplicity and interpretability
- Cons:
  - Data sparsity
  - Perform poorly with a priori unseen transition pairs

Target-Guided Open-Domain Conversation, ACL, 2019

Dataset Creation

Dialog	**A:** I spend a lot of time **outside**. **B:** I like the outdoors as well, especially **gardening** . **A:** Wow! I used to have a **garden** too. **B:** I love sipping coffee while enjoying **flowers** in my garden. **A:** Flowers are always beautiful and **colorful** ! **B:** I like anything with **art**, especially colorful things.
Entity Path	**Outside-Garden-Flower-Color-Art**
Target	**Art**

Table 1: A target-oriented example dialog in TGConv

- Target-Guided Conversation (TGC) [TGODC]
  - is constructed from Persona-Chat
  - The conversations cover a broad range of topics such as work, family, personal interest, etc
  - the discussion topics change frequently during the course of the conversations
- Chinese Weibo Conversation(CWC) [DKRN]
  - derived from a public multi-turn conversation corpus1 crawled from Sina Weibo
  - covers rich real-world topics in our daily life, such as shopping, disease, news, etc
- Reddit-based [CKC]
  - collected from casual chats on the CasualConversation and CasualUK subreddits,
    - where users chat freely with each other in any topic
  - use TF-IDF and part-of-speech (POS) features to extract keywords, same as TGC
- TGConv [TopKG]
  - from chit-chat corpus ConvAI2
  - a consistent reasoning path of words linking all the utterances in their order in the dialog
  - is distinguished into "easy-to-reach/hard-to-reach" targets

A survey on Proactive Dialogue Systems: Problems, Methods, and Prospects, Arxiv, 2023

Component 1: Turn-level keyword transition module

- Task: Guarantee the smooth transition
- Approaches
  - A) Pairwise PMI-based transition [TGODC]
  - B) Neural-based prediction [TGODC]
  - C) Hybrid kernel-based method [TGODC]
  - D) Inject semantic knowledge relations among candidate keywords [DKRN]
  - E) Leverage external commonsense knowledge graphs (CKG) as an another input [CKC]
  - F) Global-target graph and local-target graph as external inputs [TopKG]

Component 1: Turn-level keyword transition module

A) Pairwise PMI-based transition

- Construct a keyword pairwise matrix
- Characterizes the association between keywords in the observed conversation data
- Pointwise mutual information (PMI)
  - Computes likeliness
  - p(wi|wj) is the ratio of transitioning to wi in the next turn given wj in the current turn
  - p(wi) is the ratio of wi occurrence
  - Both quantities can be directly counted from the conversation data beforehand

$$\mathrm{PMI}(w_i, w_j) = \log p(w_i|w_j)/p(w_i),$$

Component 1: Turn-level keyword transition module

B) Neural-based prediction

- Steps:
  - Use a recurrent network to encode the conversation history
  - Feed the resulting features to a prediction layer to obtain a distribution over keywords for the next turn
- Pros:
  - Straightforward
- Cons:
  - Rely on a large amount of data for learning

Target-Guided Open-Domain Conversation, ACL, 2019

Component 1: Turn-level keyword transition module

C) Hybrid Kernel-based Method

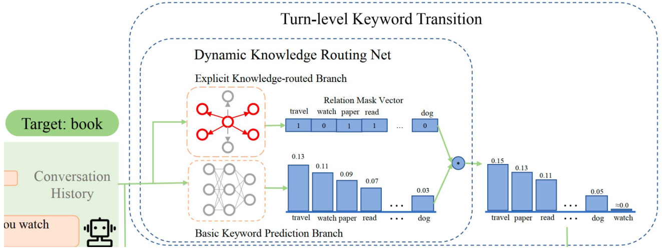
- Combines neural feature extraction with pairwise closeness measuring
- Steps:
  - 1) Measuring the cosine similarity of their normalized word embeddings
  - 2) Feeding the quantity to a kernel layer consisting of K RBF kernels
  - 3) The output of the kernel layer is a K-dimension kernel feature vector
    - Fed to a single-unit dense layer for a candidate score
  - 4) The score is finally normalized across all candidate keywords to yield the candidate probability distribution
  - 5) If the current turn has multiple keywords
    - The corresponding multiple K-dimension kernel features are first summed up before feeding to the dense layer
    - The intermediate kernel layer serves as a soft aggregation mechanism to account for multiple-to-one keyword transition

Target-Guided Open-Domain Conversation, ACL, 2019
End-to-end neural ad-hoc ranking with kernel pooling, SIGIR, 2017

Component 1: Turn-level keyword transition module

D) Inject semantic knowledge relations among candidate keywords

- Previous methods' limitation
  - Considering semantic or factual knowledge relations among candidate topics/keywords
- There are two branches in our DKRN
  - Basic keyword prediction branch
  - Explicit knowledge routing branch
- Mask those keywords that are uncorrelated directly to the conversation context/history

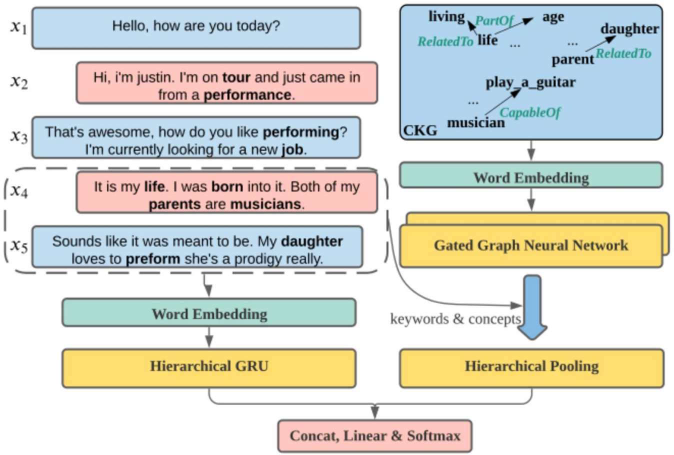


Dynamic Knowledge Routing Network For Target-Guided Open-Domain Conversation, AAAI, 2020

Component 1: Turn-level keyword transition module

F) Leverage external commonsense knowledge graphs (CKG)

- A multi-label classification problem
- Steps:
  - 1) Utterance Representation
  - 2) CKG Graph Representation
  - 3) Keyword and Concept Representation
  - 4) Classification

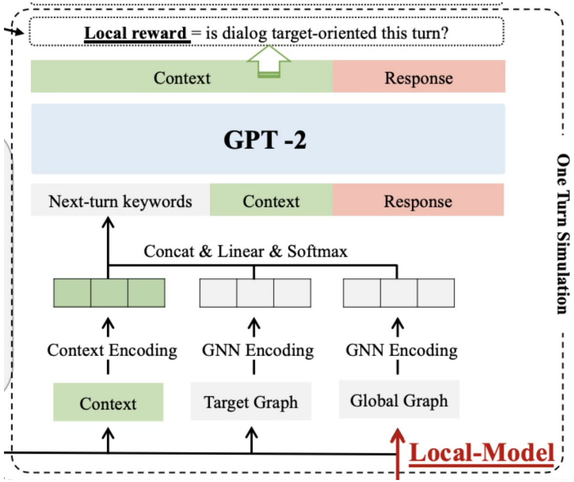


Keyword-guided neural conversational model, AAAI, 2021

Component 1: Turn-level keyword transition module

E) Global-target graph and local-target graph as external inputs

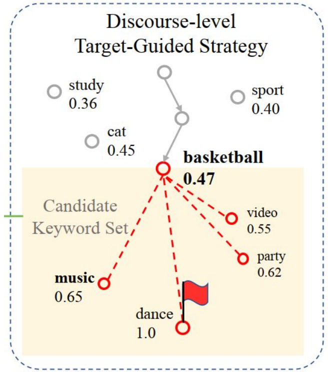
- Previous methods shortcoming:
  - They adopt a short-sighted and greedy strategy
  - instead of global planning to optimize the process towards the global target
- Steps:
  - Graph-based Encoder
  - Conversation Context Encoder
  - Classification



TopKG: Target-oriented Dialog via Global Planning on Knowledge Graph, COLING, 2022

Component 2: Discourse-level Target-Guided Strategy module

- Task: Taking an effective step towards the target
- Approaches:
- A) Calculate the closeness between keywords candidates and the target [TGODC]
    - Higher target closeness as a constraint, cosine similarity
- B) Dual Discourse-level Target-Guided Strategy [DKRN]
    - Not only constrains the keyword but also constrains the next chosen response (retrieval model)



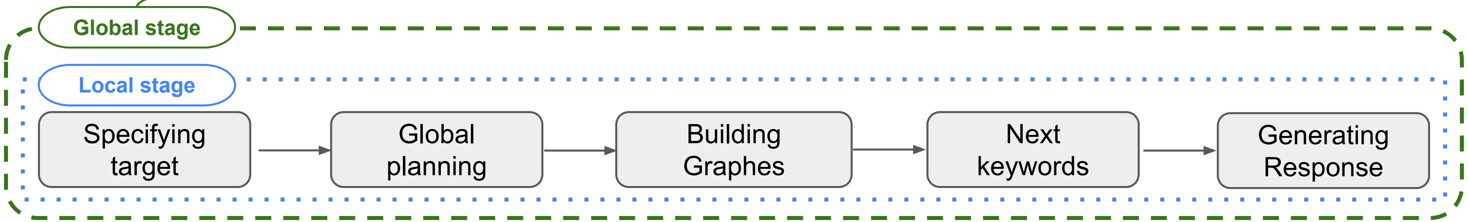
Component 2: Discourse-level Target-Guided Strategy module…

- Approaches:
- C) Leverage external commonsense knowledge graphs (CKG) as an another input
    - The distance between keywords is measured as the weighted path length between keywords on the CKG
    - Computed by the **Floyd-Warshall algorithm**
    - The edge weights on ConceptNet correlate positively with concept relatedness
        - we apply a reciprocal operation to the weights before computing path lengths

Component 2: Discourse-level Target-Guided Strategy module…

- Approaches:
- D) Global planning:
  - Leveraging RL based global-model
  - Dialogue Policy Learning

Guide with RL:
- State: G_global, G_local, Context
- Action: The next turn keyword
- Action space: The potential paths obtained by global planning
- Reward:
  - Local: Encourages the contextual consistency at each turn of dialog
  - Global: Encourages the global target-oriented response
- Policy: Proximal Policy Optimization



TopKG: Target-oriented Dialog via Global Planning on Knowledge Graph, COLING, 2022

Component 3: Keyword-augmented Response Retrieval/Generator

- Produce a response
  - Conditioning on both the conversation history and the predicted keyword
- Approaches:
  - 1) Retrieval based [TGODC]
    - Recurrent networks to encode the input conversation history and keyword, and each of the candidate responses
  - 2) The sequential matching network (SMN) + the Hadamard product [DKRN]
  - 3) Retrieval + matching score [CKC]
  - 4) Keyword Augmented GPT [TopKG]

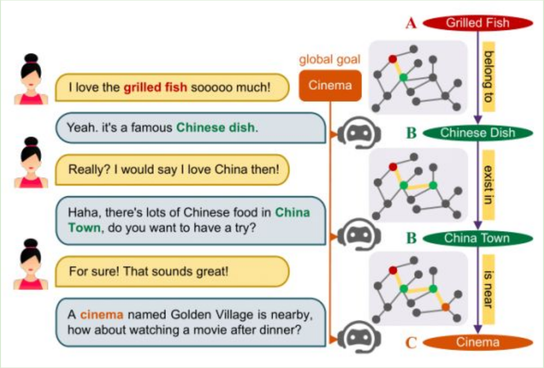
Multi-turn Conversation Generation by Self-play

- To simulate the process of two virtual agents taking turns talking with each other
- For global evaluation
    - To assess that does the model reach the target
- The base Retrieval/Generator agent to play the role of human
    - Retrieves/Generates a response without knowing the end target
- The simulator randomly picks a keyword as the end target, and an utterance as the starting point

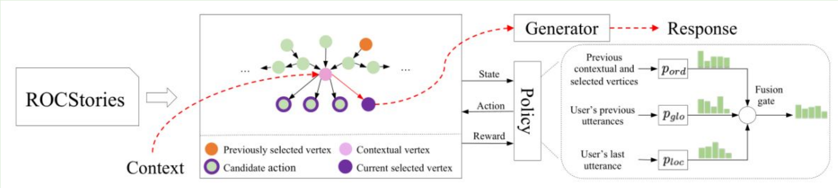
Deal or No Deal? End-to-End Learning of Negotiation Dialogues, EMNLP, 2017

Global Planning

- Enhance multi-turn consistency
- Mainly KG-based: Generating more informative responses
- Guide the response/conversation generation flow toward the target

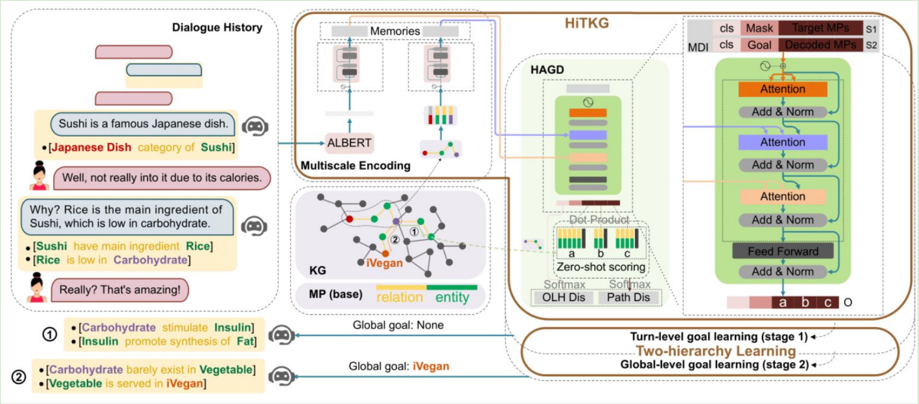


Event Graph Grounded Content Planning



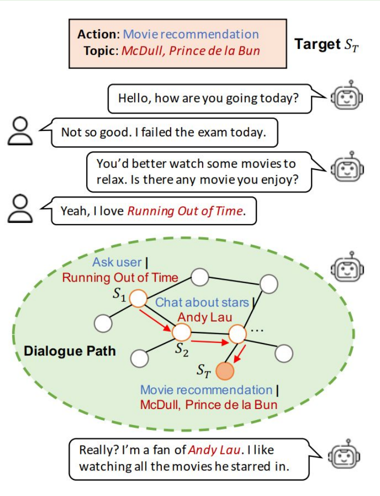
Multi-Hierarchy Learning

-



Brownian Bridge Stochastic Process

•



Topic-aware dialogue agent: Summary

- Add new responsibility to dialogue agent
- Improve the controllability on dialogue generation
- The dialogue is annotated by
  - Initial sentence and keyword, and final Target
  - The full transition keyword path
- Using a knowledge graph seems to be effective in guiding the conversation towards the Target
  - Unlike most dialog system models that use KG to generate informative responses
- Using the self-play tool completes the process of conversation generation
  - Although the main purpose of the topic-aware conversation agent is not to create a conversation dataset

Part 3: Conversation Generation, TOD

3.1 Schema-guided Generation

3.2 Simulator-Agent Generation

3.3 E2E Dataset Creation

Duration: 40 min

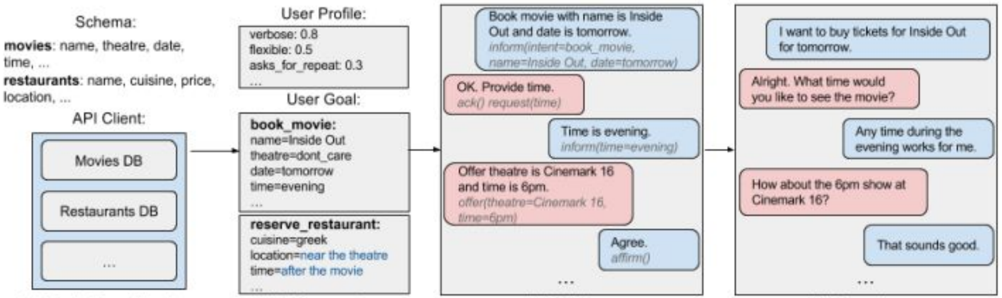
Presenter: —

Introduction

- Human2Human: MultiWOZ
- Schema-guided Generation: M2M, SGD
  - Combining automation and crowdsourcing
  - Generate outline or schema by simulator
  - Paraphrase template utterances into more natural sounding utterances by crowdsourcing
- Simulator-Agent Interaction
  - The dialogue system (DS) and user simulator (US) discourse with each other
  - A domain-independent (general) user simulator with transformers [TUS]
  - Interaction between a prompt-based user simulator and a dialog system [ICL-US]
  - The context encoder is shared between the two agents [JOUST]
- E2E dialogue generation
  - A model generates a whole dialogue

Schema-guided Generation: M2M

- A framework combining automation and crowdsourcing
- Machine-to-Machine:
  - 1) Generate dialogue templates
    - Creating an environment with a simulated user
  - 2) These templates can be mapped to a natural language
- self-play
- user simulator: employ an agenda-based
- system agent: a finite state machine based



(a) Task Specification (b) Scenario (c) Outline (d) Paraphrase

A

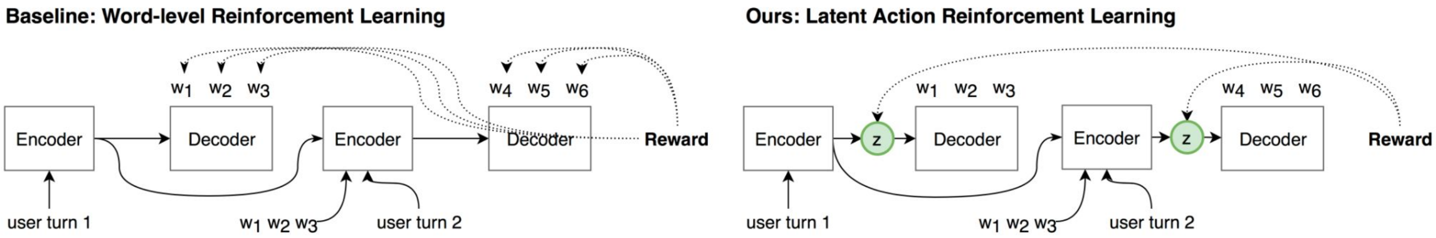
Schema-guided Generation: SGD

- Schema-guided paradigm for task-oriented dialogue
  - simplifies the integration of new services and APIs with large scale virtual assistants
- Crowdsourcing: to paraphrase template utterances into more natural sounding utterances
- Steps:
  - 1) Services and APIs
  - 2) Dialogue Simulator Framework
  - 3) Dialogue Paraphrasing

A

Simulator-Agent Interaction: LaRL

-



A

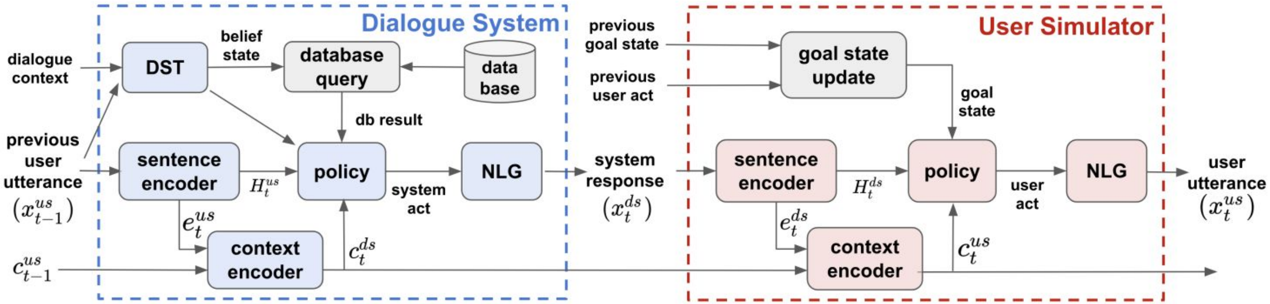
Simulator-Agent Interaction: TUS

-

A

Simulator-Agent Interaction: JOUST

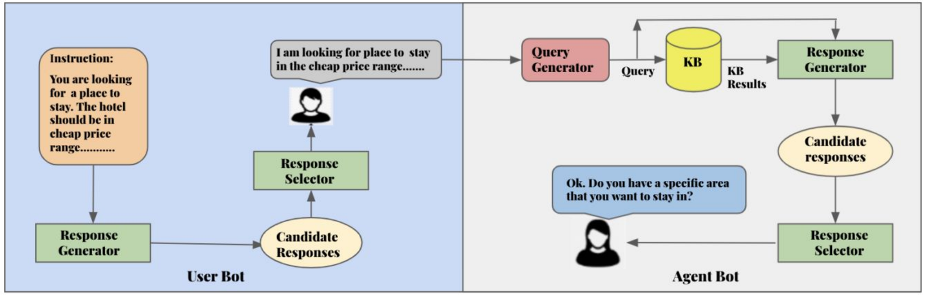
-



A

Simulator-Agent Interaction: SimulatedChat

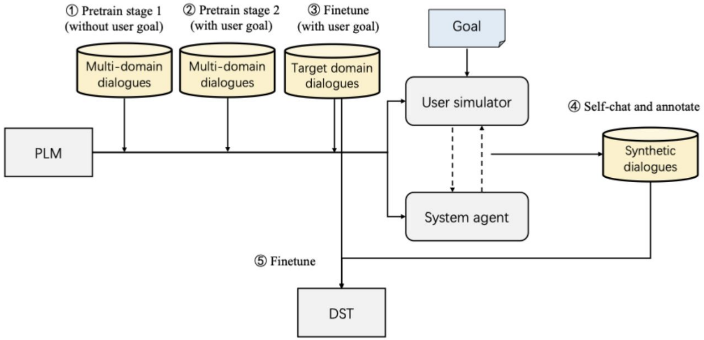
-



A

Simulator-Agent Interaction: Unified-US

-



A

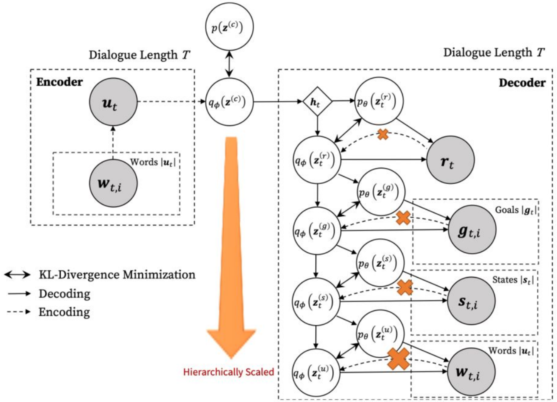
Simulator-Agent Interaction: ICL-US

-

A

E2E Dataset Creation: VHDA

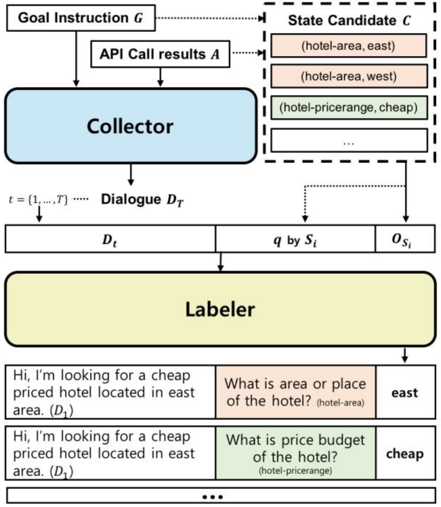
-



A

E2E Dataset Creation: NeuralWOZ

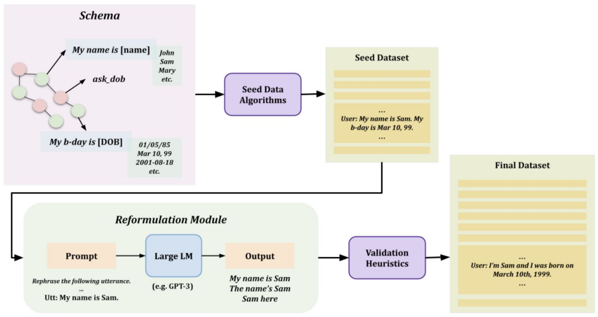
-

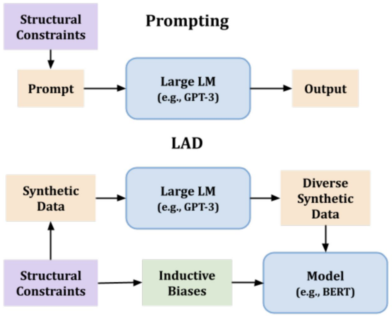


A

E2E Dataset Creation: LAD

-

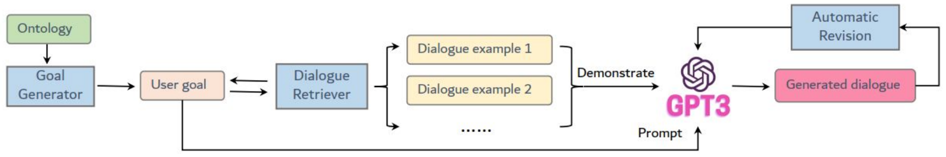




A

E2E Dataset Creation: DIALOGIC

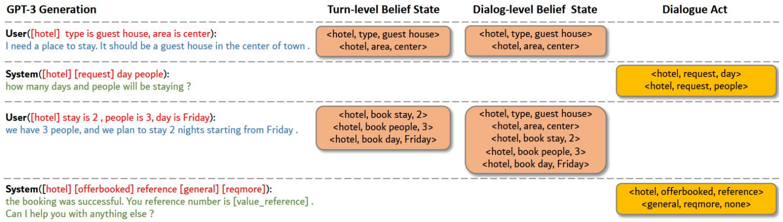
-



A

E2E Dataset Creation: PLACES

-



A

Part 3: Evaluation

How to evaluate

- Open-domain Dialogue system
    - Measure response correctness
        - Automatic Evaluation metrics:
            - F1_score, EM, BLEU, ROUGE, METEOR, Perplexity, BERTscore, MoverScore
    - Measure the whole conversation
        - Human Evaluation
            - Naturalness, Relevance, Coherence, Richness, Fluency, Informativeness
    - For models that generate conversation indirectly
        - Make conversation with human or user simulator
            - Evaluate the generated conversation with
                - Automatic Evaluation metrics
                - Human Evaluation
    - TARGET-COHERENCE [CRG]

How to evaluate …

- Task-oriented Dialogue system
  - Automatic Evaluation metrics
  - Human Evaluation
    - Consistency, Grammar, Fluency
  - Use of generated data in End-task
    - End-to-end dialogue modeling (E2E)
    - Dialogue state tracking (DST)
    - Metrics:
      - Inform, Success, BLEU,
      - An overall metric Combined Score: BLEU +0.5×(Inform+Success)

Part 4: Limitations and Future work

Open Challenges

- Generate More challenging and realistic dataset
  - With clarification questions
  - Mixed Initiative dialogues
  - Real-world scenarios, Negotiation and Persuasion dialogues
- Define strategies grounded on external resource
  - Use them for generating full conversation
- How to evaluate the added challenges
- For task-oriented dialogue system:
  - Out of domain generation
  - Multi domain dialogue generation

Building and evaluating open-domain dialogue corpora with clarifying questions, EMNLP, 2021

Why we generate new data

- [GCN] many challenges still remain
    - Handling idioms, humour, expressing empathy, processing unstructured knowledge
    - One big factor for this is the lack of large and rich conversational data that include these complex aspects of human communication
    - These expensive data collections usually target a single phenomenon at a time, and hence do not necessarily scale to the richness of human conversations
    - Another challenge, privacy, preventing the use of much of the publicly available conversational data
- [Baize] Fine-tune chat models to be specialized in specific areas, such as
    - healthcare or finance
- [Baize] Enhance the knowledge and ability of the chat model for a specific domain
    - One could use questions or phrases extracted from a domain-specific dataset, medical domain
    -

Conclusion

- We introduce approaches, tools, … that can be a part of full conversation generation process
- Study different user behaviour
- Adding more challenges to the conversation
    - (i.e. user satisfaction, clarification question)
- Multimodal dialogue systems

References