P3- Introduction

The existing commonsense knowledge bases typically organize tuples in an isolated manner,

which lacks a commonsense conversation model to plan the next step.

To fill this gap, this paper designed a large-scale multi-turn human-written dialogue corpus

and created the first Chinese commonsense conversation knowledge graph

that contains social commonsense knowledge and dialogue flow information.

To demonstrate the potential of the proposed graph,

the authors developed a graph-dialogue matching method and benchmarked two graph-based conversation tasks.

接ppt ATOMIC

The "Relation" component of these triples defines the type of relationship between the "Subject" and "Object" entities.

接ppt例子， 点击ppt

Here, xEffect and xWant are two of nine relations defined in ATOMIC to infer people’s mental states for a given event, e.g., PersonX adopts a cat.

P4- Problem being solved

Ppt 例子

ATOMIC contains multiple tails like {finds out he has allergies}

and the tail {becomes less lonely}. 点击ppt

So the first of difficulty is the existence of multiple tails, which will confuse the chatbots

when inferring the cause behind the negative emotion.

P5- Problem being solved

the knowledge tuples in ATOMIC are isolated,

so it’s more difficult for chatbots to reason which tail(s) of knowledge

should be used to produce coherent responses.

接ppt

{PersonX adopts a cat, isAfter, finds a cat at the animal shelter} is detected from the dialogue history,

then the tuple {PersonX adopts a cat, xNeed, go to an animal rescue center} should not be considered anymore for future conversations. 点击ppt

Here comes the second difficulty: these issues hamper the application of ATOMIC

to multi-turn dialogue modeling where the conversational agents need

not only know the current state

but also plan the future dialog flow.

P6 – Related work

ConceptNet has a Chinese version with a relatively small set of knowledge.

TransOMCS: is built automatically by converting syntactic parses of Web sentences into structured knowledge.

接ppt

P7 – Contributions

Although the ways using in connecting knowledge and conversation seems plausible,

these knowledge graphs are built on the granularities of word or phrase,

which makes them hard to match the overall semantics of dialogue sentences.

The paper build a Chinese commonsense conversation knowledge graph

based on both multi-turn conversational corpus and event-centered knowledge base.

And also it propose to use Sentence-BERT,

a transformer-based semantic similarity model,

to construct dialog flow edges in our knowledge graph.

P8 – A Scenario-based Multi-turn Conversation Corpus

Our aim is to extract common dialog flow information from real conversations.

In this way, it is crucial to ensure the quality of the conversation corpus

and the reliability of the extraction method.

In the following, we firstly introduce the conversation corpus CConv we depend on.

CConv: collect a multi-turn human-written Chinese conversation corpus based on crowdsourcing.

- 32k sessions of high-quality two-party conversations (650k utterances in total)

on 200 scenarios of 15 daily topics.

- conversations are annotated with fine-grained emotional labels

including speaker’s emotion type, emotion cause, and response intention type.

The paper define 5 general emotion types: {joy, angry, sad, surprising, other}

And 6 commonly-adopted intent classes: {ask, advise, describe, opinion, console, other}

P9- ATOMIC  
Based on ATOMIC, the authors design a pipeline method to translate it into Chinese,

meanwhile ensuring the resulted knowledge graph is reliable and suitable for conversation grounding.

ATOMIC organizes commonsense knowledge in the form of triplet <head, relation, tail>,

where head often describes a daily event.

It collects knowledge about how people will react to a given event.

It also organizes knowledge using several inferential relations

and naturally supports if-then reasoning,

which I have mentioned initially.

在将ATOMIC翻译成中文方面，作者采用Regular Replacement和Joint Translation 的方法来提高翻译的质量，翻译后的ATOMIC表示为ATOMIC-zh。

P10- Conversation Knowledge Graph Construction

To supply dialog flow information for commonsense reasoning, we create a Chinese Commonsense Conversation Knowledge Graph, C3KG, whose statistics are summarized in below. The core is how to build new dialog flow relations.

P11- Conversation Knowledge Graph Construction (C3KG)

To address event extraction, a dependency parsing-based event detection pipeline is developed to extract salient events in each utterance.

In order to discover common dialog flows among the knowledge base,

the event mentions in the conversations are then linked to ATOMIC heads

using matching techniques.

Sentence-BERT: It encodes two given sentences separately and calculates the similarity between their representations, and thus performing efficiently in large-scale many-to-many matching.

P12 – Head-Head

Event Flow

If two event mentions are detected together within in a conversation,

the cooccurrence can be regarded as a dialog flow example.

Following the flow, it is then intuitive to connect the ATOMIC heads linked by the mentions, as illustrated in Figure.

By connecting intrautterance and inter- utterance mentions,

we acquire the event flows of next-sub-utterance and next-utterance.

Concept Flow

ATOMIC also has entity-level heads in addition to the phrase-level events.

To utilize them, we perform entity linking by detecting word entities with POS tag

belonging to {verb, noun, adjective} in the original conversations,

and match them with the entity-level ATOMIC heads to construct concept flow edges similarly.

P13 – Tail-Tail Edge construction

The emotional labels are used to construct two kinks of emotion-based edges connection tails in knowledge graph.

Emotion-cause dialog f low reflects the reasons for a specific emotion,

which is useful for fine-grained emotion understanding.

And emotion-intent empathy flow indicates what response intentions are proper to use

when the other one is in a specific emotion,

which is critical for response empathy.

To construct emotion-based edges, tails are categorized into 3 classes according to their connecting relations, as listed in Table.

The first class of tails are linked by relations xAttr or xReact,

which reflects people’s psychological reaction towards a certain event (head).

For instance, {PersonX runs out of steam, xAttr, tired} indicates that someone is lacking energy. We denote the first class as Tailemotion.

The second class Tailbefore states the events commonly happen before the heads,

e.g., {PersonX runs out of steam, isAfter, PersonX exercises in gym}.

On the contrary, the last class Tailafter contain the events following the head events

like {PersonX runs out of steam, xWant, to get some energy}.

点击ppt

This figure depicts the process of constructing the labeled emotion-cause edge.

Firstly, we match the tail angry in Tailemotion to the utterance emotion label "angry".

Then, we detect that the tail insomnia in Tailbefore shows up in the previous utterance.

So we build a emotion\_cause edge from the tail angry to tail insomnia.

This kind of tail-tail emotion\_cause flows is supportive for chatbots to have a better understanding of users’ emotional mood by reasoning its cause.

P14- Evaluation

Manual Assessment

­­­ We randomly choose 100 utterances to evaluate our event extraction and matching methods. We denote our proposed method as Parsing.

POS employs POS tagging-based templates to extract events, and Simple only splits and filters utterances according to punctuation before matching.

Similarity stands for the averaged matching degree,

and Number for the average number of matched ATOMIC heads of the chosen utterances, which can be seen as an indicator for matching recall.

Parsing gets an obvious similarity improvement after finetuning.

Scenario Graph Visualization

We also build up scenario graphs based on matching results and the scenario descriptions. By visualizing the matched result for each topic of scenarios, we are able to better understand the matching quality.

We visualize a snippet of the scenario graph “sickness” in Figure. After annotation, the matching accuracy based on 3 annotators reaches 0.71, which indicate a fair quality of scenario graph.

P15- Evaluation

Node Evaluation

We firstly evaluate the quality of our graph in terms of translation accuracy.

the significant increases on both Fluency and Logic aspects

clearly demonstrate the superiority of joint translation method.

Edge Evaluation

To validate the quality and robustness of these relations, we utilize another open-domain multiturn Chinese dialogue dataset, MOD.

We extract event mentions from MODutterances and match them to our graph.

Then we evaluate the connectivity and average distance of the matched results.

The comparing result shows the effectiveness of our event flow,

which leads the matching of context within a dialogue has higher connectivity and shorter distance.

P16- graph based conversation tasks

Task 1: Emotion Classification requires to produce an emotion label conditions on the conversations. Following common practice, we choose the BERT model, and sample the xAttr, xReact tails from our matching head as extra input.

Task 2: Intent Prediction requires to predict a proper type of response intent for the conversations. Wechoose BERT model, and sample the oReact, AVG\_Dist. 1.90 10.81 oEffect tails from our matching heads.

点击ppt

The accuracies of baseline methods are reported in Table.

Base denotes only using the utterance to do prediction.

Knowledge and History denote whether to add knowledge we sampled and dialogue history to the model.

While adding knowledge improves the model performances,

it seems problematic to directly concatenating history dialogues, which may bring noises.

The moderate scores also indicate that there is still a room

to improve for graph-grounded conversation understanding.