

Supplementary Material for “Transferring Domain-Agnostic Knowledge in Video Question Answering”

Tianran Wu¹
tianran.wu@ids.osaka-u.ac.jp

Noa Garcia¹
noagarcia@ids.osaka-u.ac.jp

Mayu Otani²
otani_mayu@cyberagentco.jp

Chenhui Chu³
chu@i.kyoto-u.ac.jp

Yuta Nakashima¹
n_yuta@ids.osaka-u.ac.jp

Haruo Takemura¹
takemura@cmc.osaka-u.ac.jp

¹ Osaka University
Osaka, Japan

² CyberAgent, Inc.
Japan

³ Kyoto University
Kyoto, Japan

This supplementary material contains the analysis of KnowIT-X VQA dataset as well as more results for knowledge retrieval on KnowIT-X v-3k and v-5k.


1 Dataset Statistics

1.1 Dataset examples

In the main paper, we constructed a new knowledge-based dataset KnowIT-X VQA, of which each sample includes a video clip, a related question, four candidate answers, subtitles and a piece of annotated ground-truth knowledge. We give some examples of the dataset in Figure 1 and Figure 2.

1.2 Dataset comparisons of the vocabulary

We compare KnowIT-X VQA for transfer knowledge with the source dataset KnowIT VQA. As shown in the word clouds in the Figure 3 of our main paper, the vocabularies in KnowIT and KnowIT-X differ a lot. As an extension for the figure, the vocabulary distributions of the two datasets are illustrated in Figure 3. Before calculating the distributions, we first remove the stopwords, which refer to the words that do not carry any information such as prepositions, conjunctions, etc. We use the stopwords list for English from Natural Language Toolkit (NLTK) and add several words like ‘n’t’ and ‘ah’. It can be seen in Figure 3 that the most frequent words are almost domain-specific entities like character names and locations.




Subtitles
Rachel: Uh, here's that trench coat you wanted.
Joshua: Oh, great. Great. Wow, yeah, it's comfortable.
Rachel: Ha, ha. They don't want you to put your hands in the pockets....until you are out of the store.
Joshua: Why not?

Why did Rachael does not want Joshua to put her hands in the coat?
A1: Because she put her phone number there.
A2: Because she does not want him to infect her.
A3: Because there are tags in the pocket.
A4: Because it is company policy.

Ground-Truth Knowledge: Rachel wanted Joshua to have her phone number so she put it in the jacket pocket but he tells her that he not looking to date anybody.

(a)




Subtitles
Monica: Listen, I did something kind of crazy tonight. Maybe I'm getting my period or something. I don't know. Anyway, I beeped into your machine and I heard this message....that freaked me out....and Michelle will tell you the rest.

Why is everyone dressed up?
A1: To go to Museum for an event.
A2: To go to a wedding
A3: To go to an engagement party.
A4: To go to Ross's boss's birthday party.

Ground-Truth Knowledge: Ross is taking everyone to a black tie event in his Museum where he works.

(b)




Subtitles
Rachel: Um, do you want to go inside and get some coffee?
Joshua: Yeah. Okay.
Rachel: [CLEARS THROAT]

Who is staring Rachel and Joshua while kissing?
A1: Joey
A2: Ross
A3: Chandler
A4: Monica

Ground-Truth Knowledge: They move to kiss, but realise that Chandler is staring at them. Chandler urges them on.

(c)




Subtitles
Paul: So you still think you might want that fifth date?
Monica: Yeah. Yeah, I do.

Who dates Richard?
A1: Rachel
A2: Carol
A3: Phoebe
A4: Monica

Ground-Truth Knowledge: Richard is the close family friend of the parents of one of the Friends. The two meet during a party that she is catering at, and hit it off.

(d)



Subtitles
Chandler: Oh, damn it!
Monica: Don't worry. After a while, you'll tune it out.
Doctor Connelly: Given your situation, the options with the greatest chances for success.....would be surrogacy or insemination using a sperm donor.

What do they decide to do in the end?
A1: They decide on surrogacy
A2: They decide to adopt
A3: They decide to continue trying to get pregnant
A4: They decide on insemination

Ground-Truth Knowledge: Monica and Chandler decide that they would rather adopt than go through with surrogacy.

(e)




Subtitles
Rachel: But he can't stay too mad at me. I mean, I just had his baby. Ha, ha.
Joey: That's not fair. I can't do that.
Rachel: I'm not sure you should be here when he comes up.
Joey: I didn't think I should be here either. But somebody said he'd be over it.

Who was late when the friends went to dinner to celebrate Monica's promotion?
A1: Rachel
A2: Chandler
A3: Ross
A4: Joey

Ground-Truth Knowledge: When Chandler go to the restaurant, he told everybody he was sorry that he was late and congratulated Monica. He then whispered to Ross that he was not sorry he was late because he had an incredible afternoon with Jade.

(f)




Subtitles
Joey: Oh, I am going to go for a walk in the rain. Ha, ha.
Rachel: Oh, yeah, me too.
Phoebe: That's weird.

Why did Rachel and Joey decide to go in the rain?
A1: To escape Phoebe.
A2: They will kiss in the rain.
A3: To get Phoebe confused.
A4: Because they are hot.

Ground-Truth Knowledge: Rachel finds out about the Monica and Chandler affair and has a conversation with Joey about, they promise not to tell anyone what they know.

(g)




Subtitles
Bitter lady: How could he leave me?!

Which television channel was Ross going to be appearing on?
A1: Sports Channel
A2: Wild Animal
A3: National Geographic
A4: Discovery Channel

Ground-Truth Knowledge: Ross stayed to help Rachel get ready for the dinner party with her boss. Rachel was touched when she found out that Ross had missed his appearance on television for her.

(h)

Figure 1: Examples from KnowIT-X VQA Dataset. We show in the figure three frames of the given video clips, QA pairs, subtitles and the annotated ground-truth knowledge. The correct answers are illustrated in green.




Subtitles
Ross: I wanna stay. I wanna talk about this.
Rachel: All right. How was she?
Chandler: Oh-oh.
Rachel: What? Was she good?
Joey: Don't answer that.

What is gang hearing from outside of door in the scene?

A1: Monica and Rachel's fight
A2: Ross and Rachel's fight
 A3: Phoebe and Joey's fight
 A4: Ross and Rachel's love

Ground-Truth Knowledge: Ross and Rachel fight always in the show.

(i)




Subtitles
Ross: It's 72 long-stemmed, red roses. One for each day I've known and loved Emily... ..cut up into mulch.
Rachel: Honey, that's awful.
Ross: It's not so bad. Monica's going to make potpourri.

Why did Emily the roses back to him?

A1: He lied to her about his past
 A2: He said he wanted to date other people
 A3: He cheated on her with someone else
A4: He said someone else's name at the alter

Ground-Truth Knowledge: During Ross and Emily's wedding, Ross said Rachel's name instead of Emily's name

(j)




Subtitles
Announcer: Presenting the Favorite Supporting Actress award: Joey Tribbiani from Days of Our Lives.
Joey: Any one of the brilliant actresses nominated for this award tonight...

Who got the favorite actress's award for DOOL?

A1: Rachel
 A2: Ross
A3: Joey
 A4: Monica

Ground-Truth Knowledge: Presenting the award for Favorite Supporting Actress is Joey Tribbiani from Days of Our Lives. (Applause as Joey angrily walks to the podium and eyes the trophy.)

(k)




Subtitles
Monica: I mean, women have needs. Do it. Get yours.
Phoebe: Oh, I don't know about that. I think if the two of you had sex, the repercussions would be catastrophic.

Who is the girl in the pink talking about?

A1: Her brother
 A2: Monica's x boyfriend
A3: Ross
 A4: Her x boyfriend

Ground-Truth Knowledge: Monica say's that if Ross can't do it then Rachel can borrow Chandler.

(l)




Subtitles
Rachel: Oh, yeah, no, what's that? **Ross:** If the four of us could all, you know, hang out together. Emily's coming into town this weekend. Why don't, uh, we all have dinner Sunday night?
Rachel: That would be great! - Yeah?

With whom did Ross invite Rachel to come with?

A1: Tag
 A2: Paulo
 A3: Chip
A4: Joshua

Ground-Truth Knowledge: Ross invited Rachel on a double date with him and Emily and Rachel and Joshua.

(m)




Subtitles
Chandler: Now you're just a girl in a tub. Hi, bubbles. Manly. I thought I'd let you know how it went. She pulled it out of me!

Which of the following choices was used as a 'free turn' in the game?

A1: Angel Pass
 A2: Phoebe Pass
 A3: Monica Pass
 A4: Rachel Pass

Ground-Truth Knowledge: After Ross got his question wrong, Joey went to Chandler but Ross interrupted him and said, 'Wait a minute, I believe I'm entitled to use my Angel Pass for a free turn?'

(n)




Subtitles
Phoebe: Kurt? I thought, you know, abrasive drunk. Um, Lola, mind-numbingly stupid. And, okay, you guys: Gold digger, cradle-robbing perv.

How does Phoebe know Kurt?

A1: Kurt lives in the apartment building.
 A2: Kurt is Phoebe's client.
 A3: Kurt is Phoebe's next door neighbor.
 A4: Kurt is Phoebe's co-worker.

Ground-Truth Knowledge: They are all together for Howard's retirement party.

(o)



Subtitles
Fredrick: Thank you, sweetheart.
Dina: I can't believe she's really gone. Look around you. All of this is ours.
Joey: I don't think so.

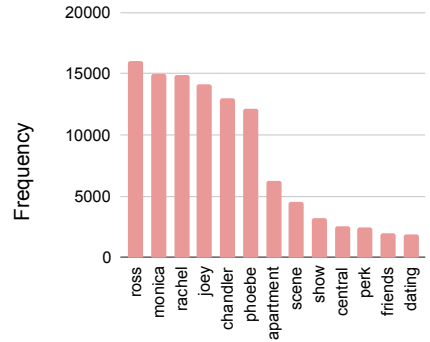
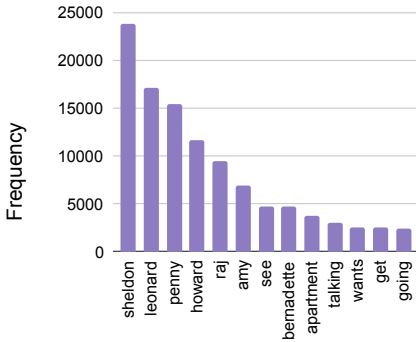
What's going on in TV?

A1: The Joey series.
 A2: Movie.
 A3: A TV show.
 A4: A TV commercial.

Ground-Truth Knowledge: Joey is an actor and is working on a TV series.

(p)

Figure 2: More examples from KnowIT-X VQA Dataset.



(a) Vocabulary distribution in KnowIT.

(b) Vocabulary distribution in KnowIT-X.

Figure 3: Vocabulary distribution in the two datasets, where x-axis shows words in the vocabulary, and y-axis their frequency in the questions, answers and annotated knowledge.

Table 1: More knowledge retrieval module results on KnowIT-X v-3k and v-5k. Each training configuration is computed with and without DET and DA.

Source	Target	Learning	v-3k		v-5k	
			R@1	R@10	R@1	R@10
KnowIT	-	Direct	0.005	0.010	0.005	0.010
		w/ DET+DA	0.244	0.482	0.190	0.447
KnowIT-X	-	Direct	0.242	0.523	0.304	0.598
		w/ DET+DA	0.279	0.563	0.312	0.592
Both	-	Direct	0.299	0.569	0.278	0.612
		w/ DET+DA	0.239	0.574	0.277	0.608
KnowIT	KnowIT-X	Transfer	0.294	0.589	0.270	0.592
		w/ DET+DA	0.299	0.579	0.302	0.612

2 More Experiment Results

Table 1 is the extension of Table 3 in the main paper. R@1 and R@10 are shown in the table. Although not all the experiments with the proposed transfer learning method show the best result, it is the R@5 that contributes most to the video reasoning as only the first five most relevant pieces of knowledge are used as inputs.