Dialogue Plus

乐然

2019.3.29

Dialogue Plus

1.User Modeling

Addressee Identification

Speaker Identification

2.Dialogue Based Application

Recommendation

Image Retrieval

3.Dialogue Content Mining

Dialogue Act Classification

Structure Mining

Interest Mining

Inference&Understanding

Addressee Identification

Problem

User	Addressee	Utterance
User 1	-	I have a problem when I install ...
SYSTEM	-	did you set initial params ?
User 2	-	Show the error message, and ...
User 1	*SYSTEM*	how ?
User 1	User 2	ok just a moment !
SYSTEM	[To Whom?]	[What?]
	1. User 1	1. see this URL : http://xxxx
	2. User 2	2. It 's already in os

Formulation

	Type	Notation
Input	Responding Agent	a_{res}
	Context	$\mathcal{C}$
	Candidate Responses	$\mathcal{R}$
Output	Addressee	$a \in \mathcal{A}(\mathcal{C})$
	Response	$\boldsymbol{r} \in \mathcal{R}$

Table 1: Notations for the ARS task.

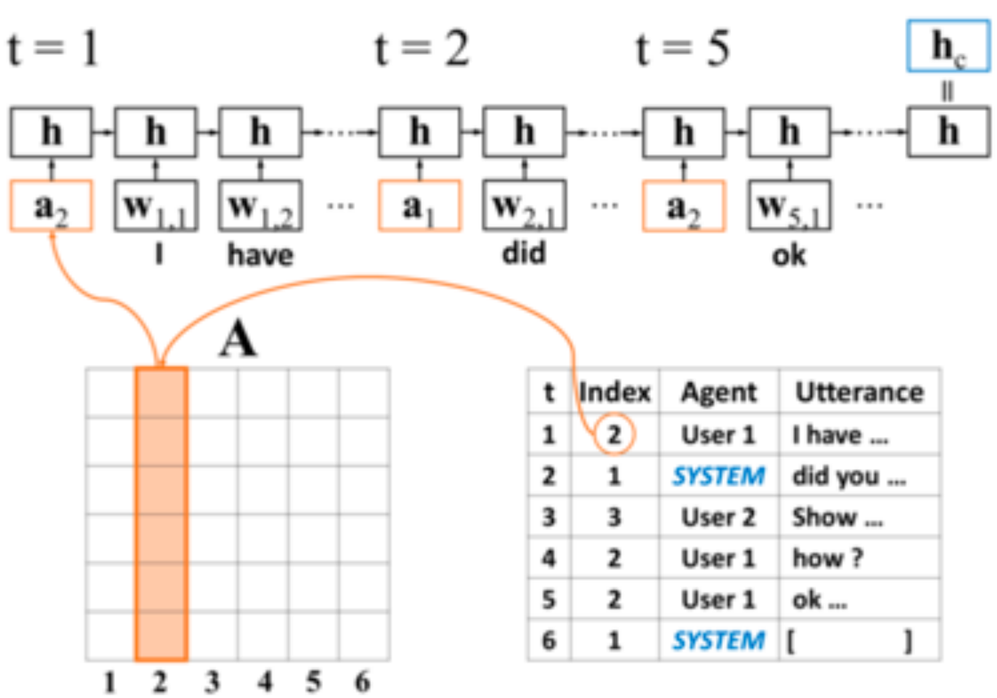
1.Multi-Party Dialogue Issue

2.Associated with Response Selection

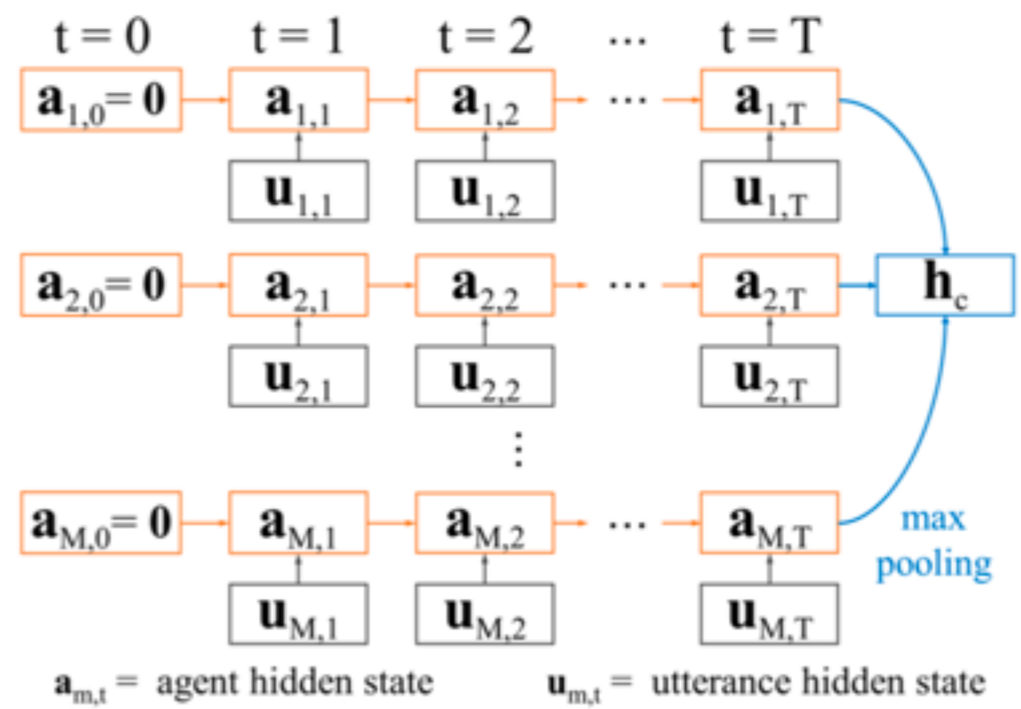
Addressee and Response Selection for Multi-Party Conversation

———EMNLP 2017

1.Jointly encoding the Utterance information and the User Information during dialog State Tracking
2. Two types of models based on whether the user representation is updated during encoding



Static Model



Dynamic Model

User	Addressee	Utterance
User 1	-	I have a problem when I install ...
SYSTEM	-	did you set initial params ?
User 2	-	Show the error message, and ...
User 1	*SYSTEM*	how ?
User 1	User 2	ok just a moment !
SYSTEM	[To Whom?]	[What?]
	1. User 1	1. see this URL : http://xxxx
	2. User 2	2. It 's already in os

Prediction

$$Pr(y(a_p) = 1|\mathbf{x}) = \sigma\left([\mathbf{a}_{res} \, ; \, \mathbf{h}_c]^{\mathrm{T}} \, \mathbf{W}_a \;\, \mathbf{a}_p\right)$$

$$Pr(y(\boldsymbol{r}_q) = 1|\mathbf{x}) = \sigma\left([\mathbf{a}_{res} \, ; \, \mathbf{h}_c]^{\mathrm{T}} \, \mathbf{W}_r \, \mathbf{h}_q\right)$$

Objective

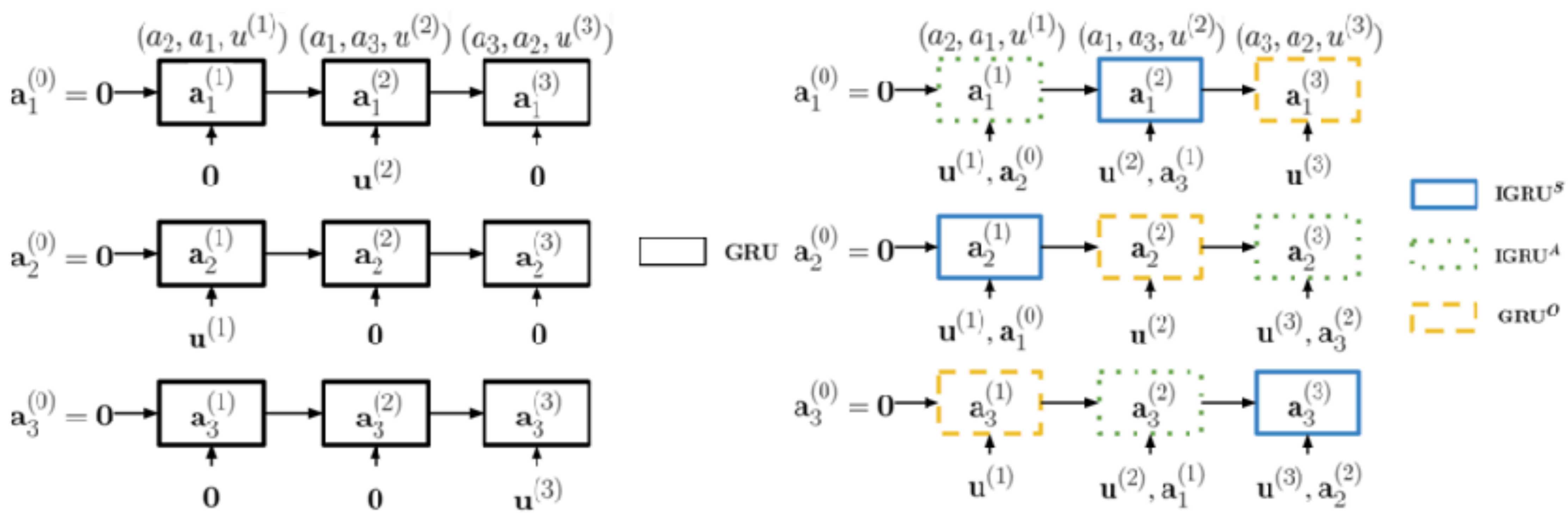
$$\mathcal{L}(\boldsymbol{\theta}) = \alpha \, \mathcal{L}_a(\boldsymbol{\theta}) + (1-\alpha) \, \mathcal{L}_r(\boldsymbol{\theta}) + \frac{\lambda}{2}||\boldsymbol{\theta}||^2$$

$$\mathcal{L}_a(\boldsymbol{\theta}) = -\sum_n [\log \; Pr(y(a^+) = 1|\boldsymbol{x}) + \log\left(1 - Pr(y(a^-) = 1|\boldsymbol{x}\right)]$$

$$\mathcal{L}_r(\boldsymbol{\theta}) = -\sum_n [\log \; Pr(y(\boldsymbol{r}^+) = 1|\boldsymbol{x}) + \log\left(1 - Pr(y(\boldsymbol{r}^-) = 1|\boldsymbol{x}\right)]$$

Addressee and Response Selection in Multi-Party Conversations with Speaker Interaction RNNs

————AAAI 2018

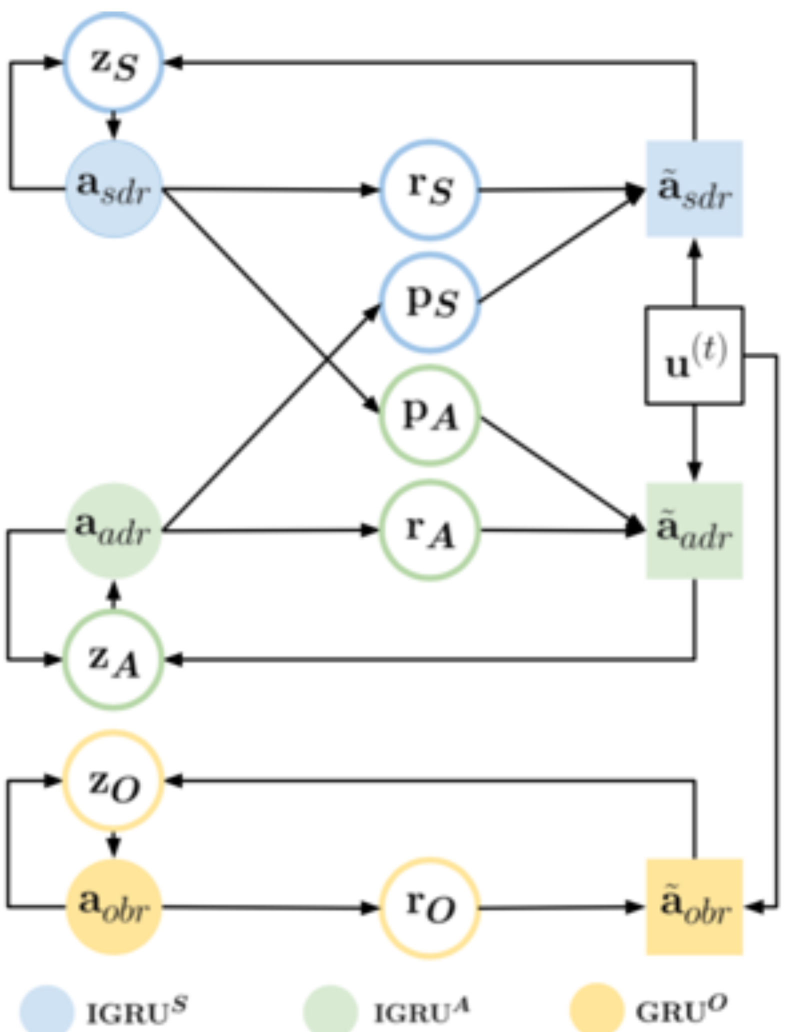


1.Users' role information is incorporated

2.The representation of each user is updated based on their role information at each utterance step

User	Addressee	Utterance
User 1	-	I have a problem when I install ...
SYSTEM	-	did you set initial params ?
User 2	-	Show the error message, and ...
User 1	*SYSTEM*	how ?
User 1	User 2	ok just a moment !
SYSTEM	[To Whom?]	[What?]
	1. User 1	1. see this URL : http://xxxx
	2. User 2	2. It 's already in os

Updating Cell



Prediction

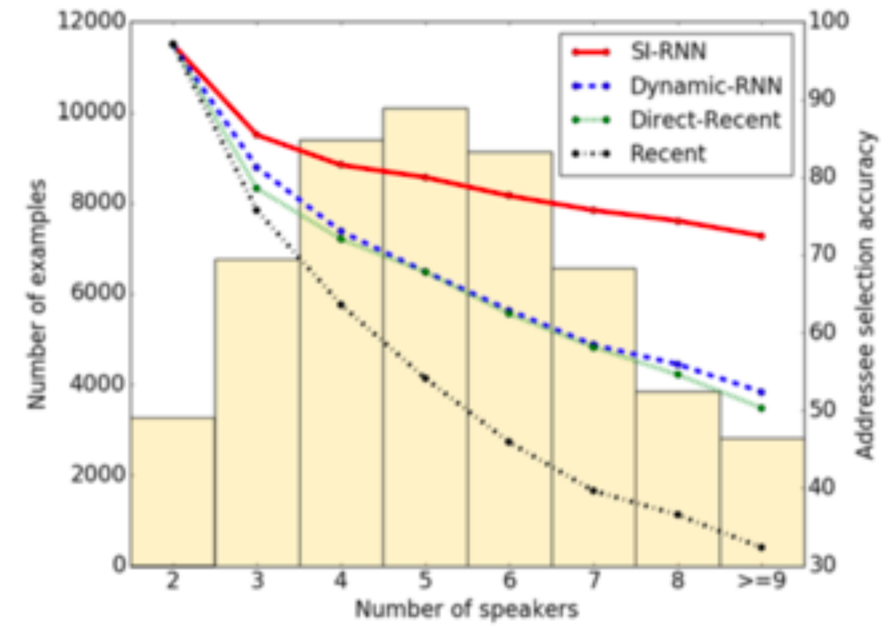
$$\mathbb{P}(a_p|\mathcal{C}, r) = \sigma([\mathbf{a}_{res}; \mathbf{h}_{\mathcal{C}}; \mathbf{r}]^\top \mathbf{W}_{ar} \mathbf{a}_p)$$

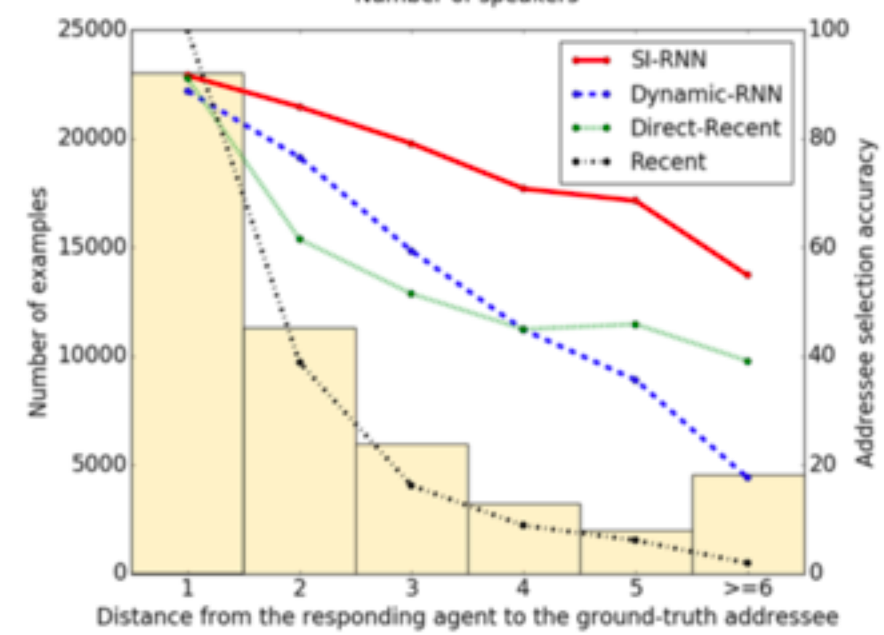
$$\mathbb{P}(r_q|\mathcal{C}, a_{adr}) = \sigma([\mathbf{a}_{res}; \mathbf{h}_{\mathcal{C}}; \mathbf{a}_{adr}]^\top \mathbf{W}_{ra} \mathbf{r}_q)$$

$$\begin{aligned}\hat{a}, \hat{r} &= \underset{a_p, r_q \in \mathcal{A}(\mathcal{C}) \times \mathcal{R}}{\arg\max} \ \mathbb{P}(r_q, a_p|\mathcal{C}) \\ &= \underset{a_p, r_q \in \mathcal{A}(\mathcal{C}) \times \mathcal{R}}{\arg\max} \ \mathbb{P}(r_q|\mathcal{C}) \cdot \mathbb{P}(a_p|\mathcal{C}, r_q) \\ &\qquad\qquad + \mathbb{P}(a_p|\mathcal{C}) \cdot \mathbb{P}(r_q|\mathcal{C}, a_p)\end{aligned}$$

Performance

	T	RES-CAND = 2				RES-CAND = 10			
		DEV	TEST			DEV	TEST		
		ADR-RES	ADR-RES	ADR	RES	ADR-RES	ADR-RES	ADR	RES
Chance	-	0.62	0.62	1.24	50.00	0.12	0.12	1.24	10.00
Recent+TF-IDF	15	37.11	37.13	55.62	67.89	14.91	15.44	55.62	29.19
Direct-Recent+TF-IDF	15	45.83	45.76	67.72	67.89	18.94	19.50	67.72	29.40
Static-RNN	5	47.08	46.99	60.39	75.07	21.96	21.98	60.26	33.27
(Ouchi and Tsuboi 2016)	10	48.52	48.67	60.97	77.75	22.78	23.31	60.66	35.91
	15	49.03	49.27	61.95	78.14	23.73	23.49	60.98	36.58
Static-Hier-RNN	5	49.19	49.38	62.20	76.70	23.68	23.75	62.24	34.51
(Zhou et al. 2016)	10	51.37	51.76	64.61	78.28	25.46	25.83	64.86	36.94
(Serban et al. 2016)	15	52.78	53.04	65.84	79.08	26.31	26.62	65.89	37.85
Dynamic-RNN	5	49.38	49.80	63.19	76.07	23.44	23.72	63.28	33.62
(Ouchi and Tsuboi 2016)	10	52.76	53.85	66.94	78.16	25.44	25.95	66.70	36.14
	15	54.45	54.88	68.54	78.64	26.73	27.19	68.41	36.93
	5	60.57	60.69	74.08	78.14	30.65	30.71	72.59	36.45
SI-RNN (Ours)	10	65.34	65.63	78.76	80.34	34.18	34.09	77.13	39.20
	15	**67.01**	**67.30**	**80.47**	**80.91**	**35.50**	**35.76**	**78.53**	**40.83**
SI-RNN w/ shared IGRUs	15	59.50	59.47	74.20	78.08	28.31	28.45	73.35	36.00
SI-RNN w/o joint selection	15	63.13	63.40	77.56	80.38	32.24	32.53	77.61	39.73





Dialogue Plus

1.User Modeling

- Addressee Identification
- Speaker Identification

2.Dialogue Based Application

- Recommendation
- Image Retrieval

3.Dialogue Content Mining

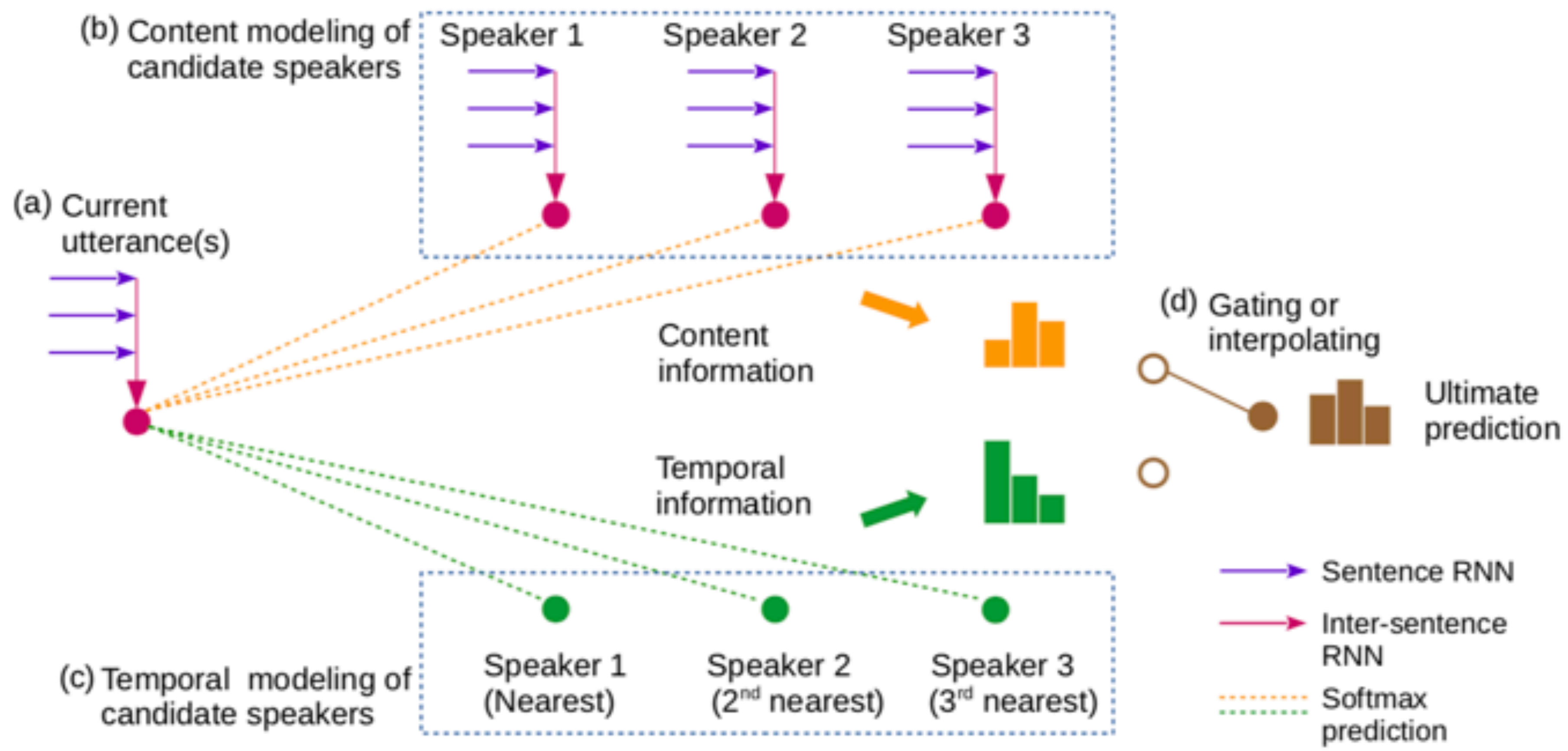
- Dialogue Act Classification
- Structure Mining
- Interest Mining
- Inference&Understanding

Towards Neural Speaker Modeling in Multi-Party Conversation: The Task, Dataset, and Models

———AAAI 2018 Workshop

1.Multi-Party Dialogue Issue

2.Given the context and the query utterance, the objective is to predict the speaker



Prediction:

$$\widetilde{p}_i = \exp\{\boldsymbol{s}_i^\top \boldsymbol{u}\}$$

$$p(s_i) = \frac{\widetilde{p}_i}{\sum_i \widetilde{p}_j}$$

$$\boldsymbol{p}^{(\text{hybrid})} = (1-g)\cdot\boldsymbol{p}^{(\text{temporal})} + g\cdot\boldsymbol{p}^{(\text{content})}$$

Statistics

Data partition	# of samples
Train	174,487
Validation	21,071
Test	20,501

Performance

Model	Macro F_1	Weighted F_1	Micro F_1	Acc.	MRR.
Random guess	19.93	34.19	27.53	27.53	N/A
Majority guess	21.26	62.96	74.01	74.01	N/A
Hybrid random/majority guess	25.26	61.99	69.29	69.29	N/A
Temporal information	26.07	63.60	73.99	73.99	84.85
Content information	42.61	65.04	61.82	58.58	74.86
+ static attention	42.50	65.28	61.79	58.99	74.89
+ sentence-by-sentence attention	42.56	65.96	62.86	59.81	75.58
Hybrid: Interpolating after training	**44.25**	**71.35**	**76.10**	**75.84**	**85.73**
Hybrid: Interpolating while training	41.30	70.10	75.57	75.31	85.20
Hybrid: Self-adaptive gating	39.45	69.55	74.11	74.09	84.85

Dialogue Plus

1.User Modeling

Addressee Identification

Speaker Identification

2.Dialogue Based Application

- Recommendation

Image Retrieval

3.Dialogue Content Mining

Dialogue Act Classification

Structure Mining

Interest Mining

Inference&Understanding

Towards Deep Conversational Recommendations

——NIPS 2018

Contributions:

1.Jointly learning to generate response & recommend & classify sentiment

2.A dataset for conversational recommendation

Dataset Construction:

1.Pairing up AMT workers and give each of them a role. (movie seeker and recommender.

2.Three questions are asked after dialogue collection for each pair.

(1) Whether the movie was mentioned by the seeker?

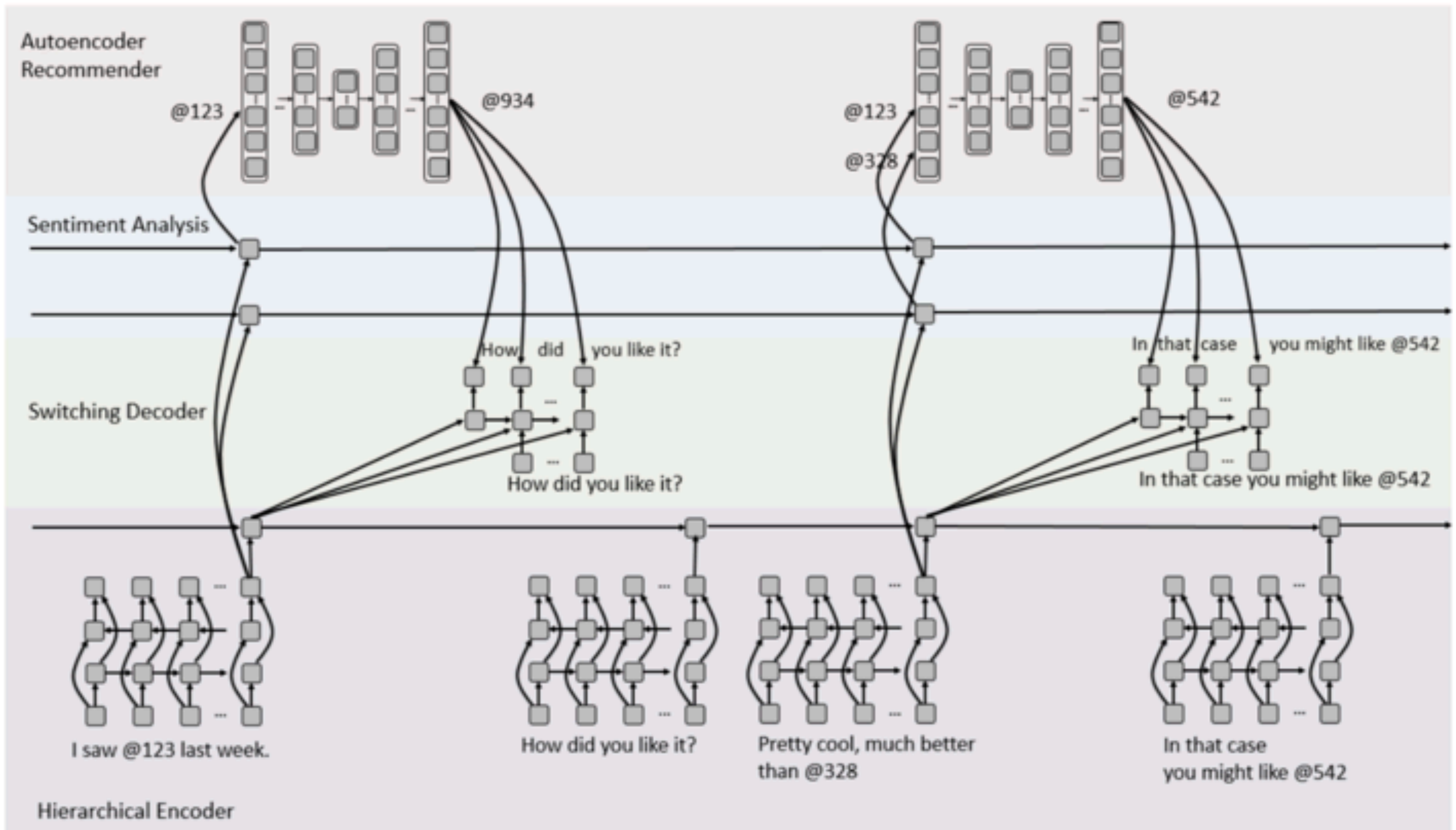
(2) Whether the seeker has seen the movie?

(3) Whether the seeker liked the movie?

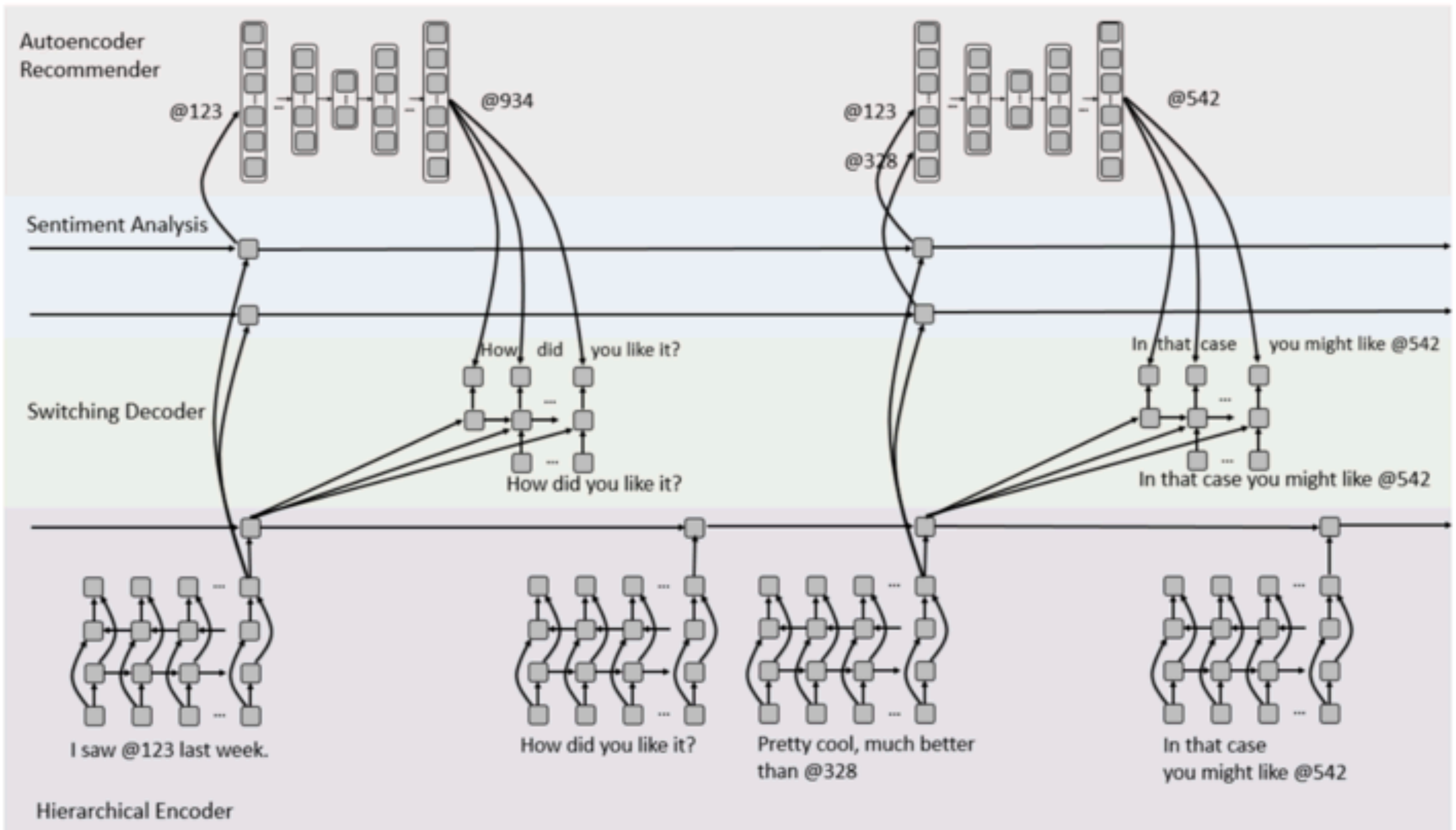
Statistics

# 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

Architecture



HERD Encoder



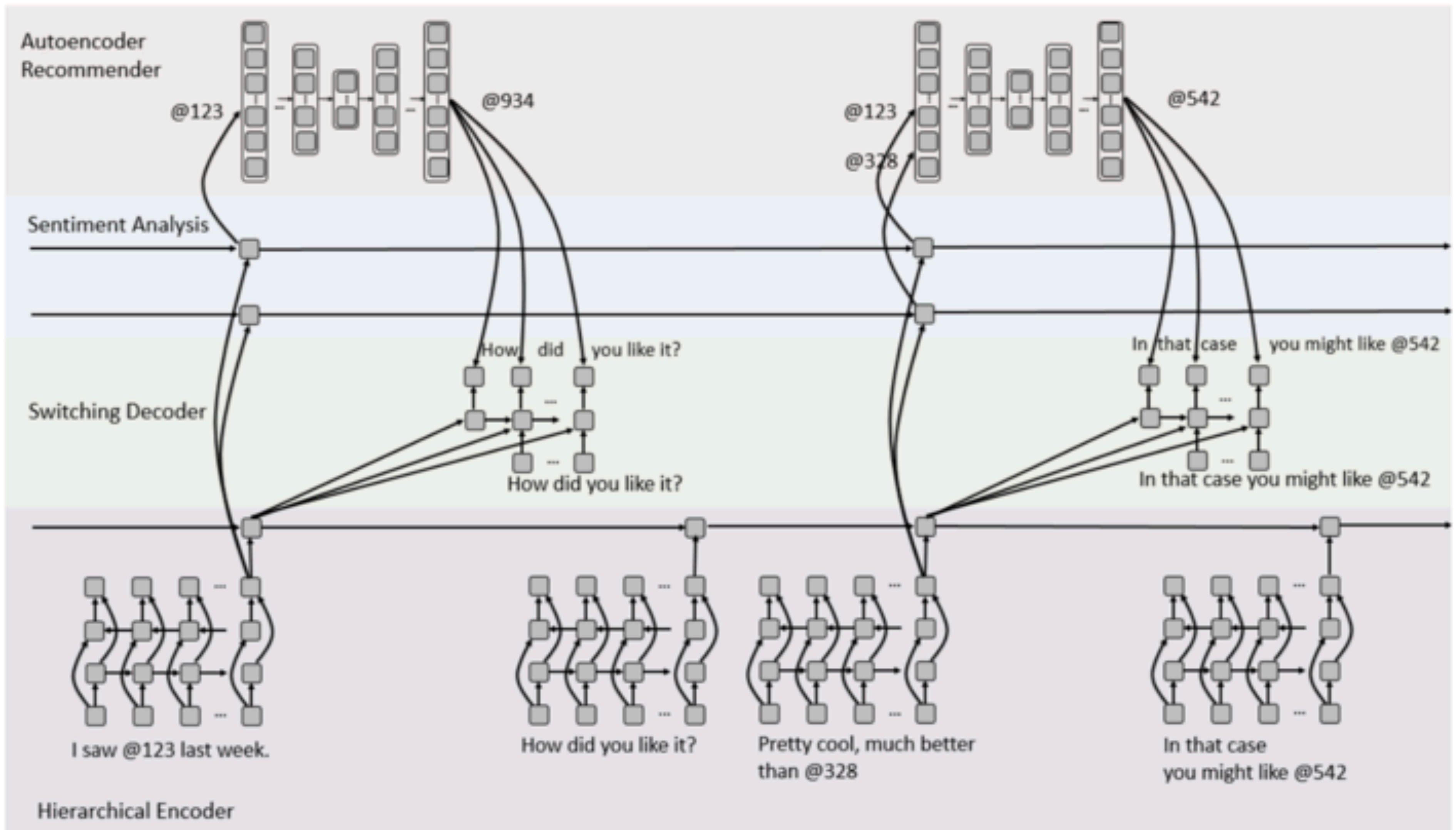
1.The HERD encoder encodes the utterances from both the recommender and the seeker.

2.Adding one dimension as the movie-name indicator after the first layer

[“<s>”, "you", "would", "like", "the", "sixth", "sense", ".", "</s>"]

[0, 0, 0, 0, 1, 1, 1, 0, 0]

Sentiment Classifier



A transformation from the dialogue state to a 7-dim vector

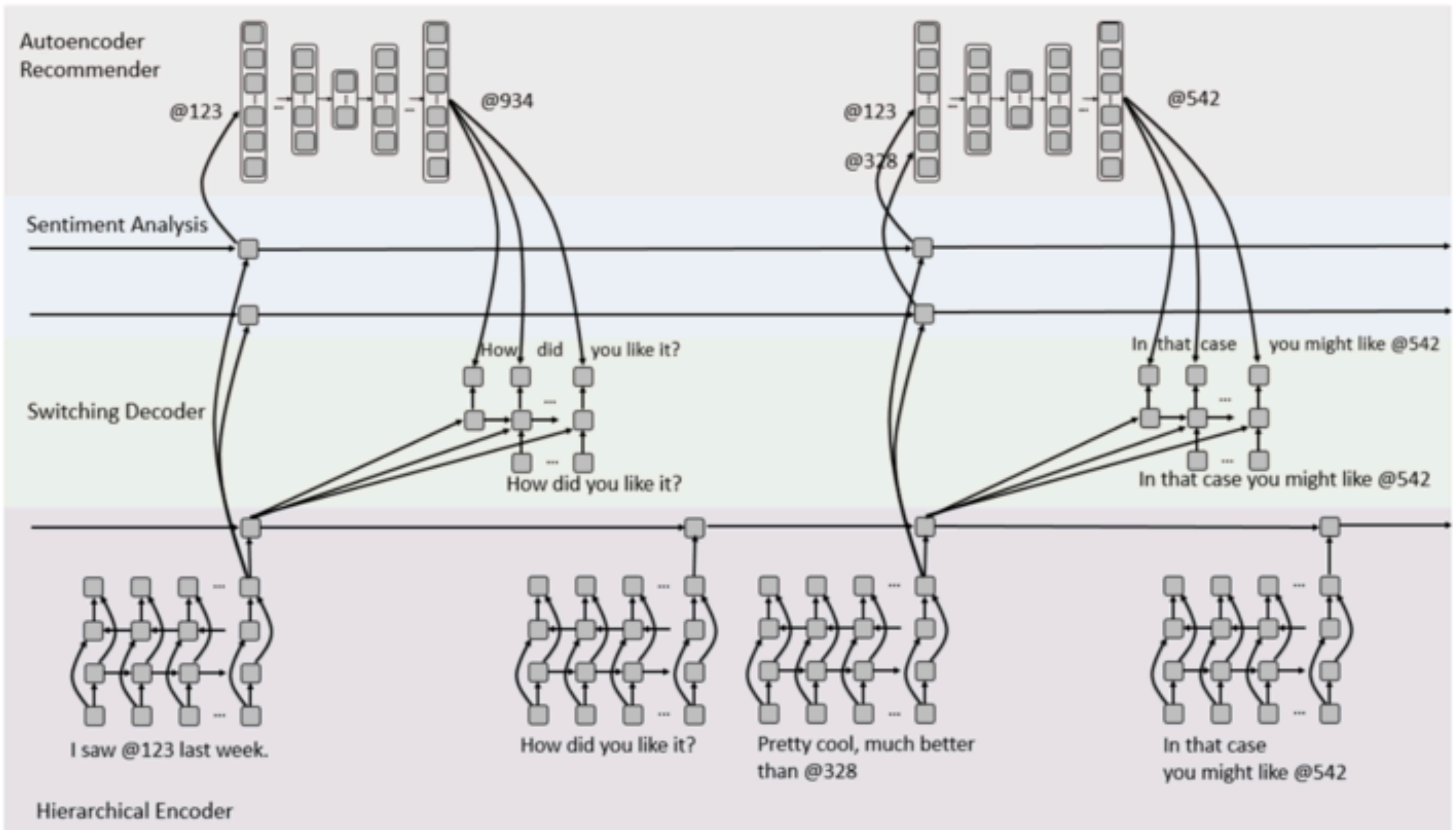
1st dim: Whether the movie was mentioned by the seeker? ————sigmoid

2nd-4th dim: Whether the seeker has seen the movie?———— softmax

5th-7th dim: Whether the seeker liked the movie? ———— softmax

$$o_i^{\text{sugg}}, o_i^{\text{seen}}, o_i^{\text{liked}}$$

Recommender

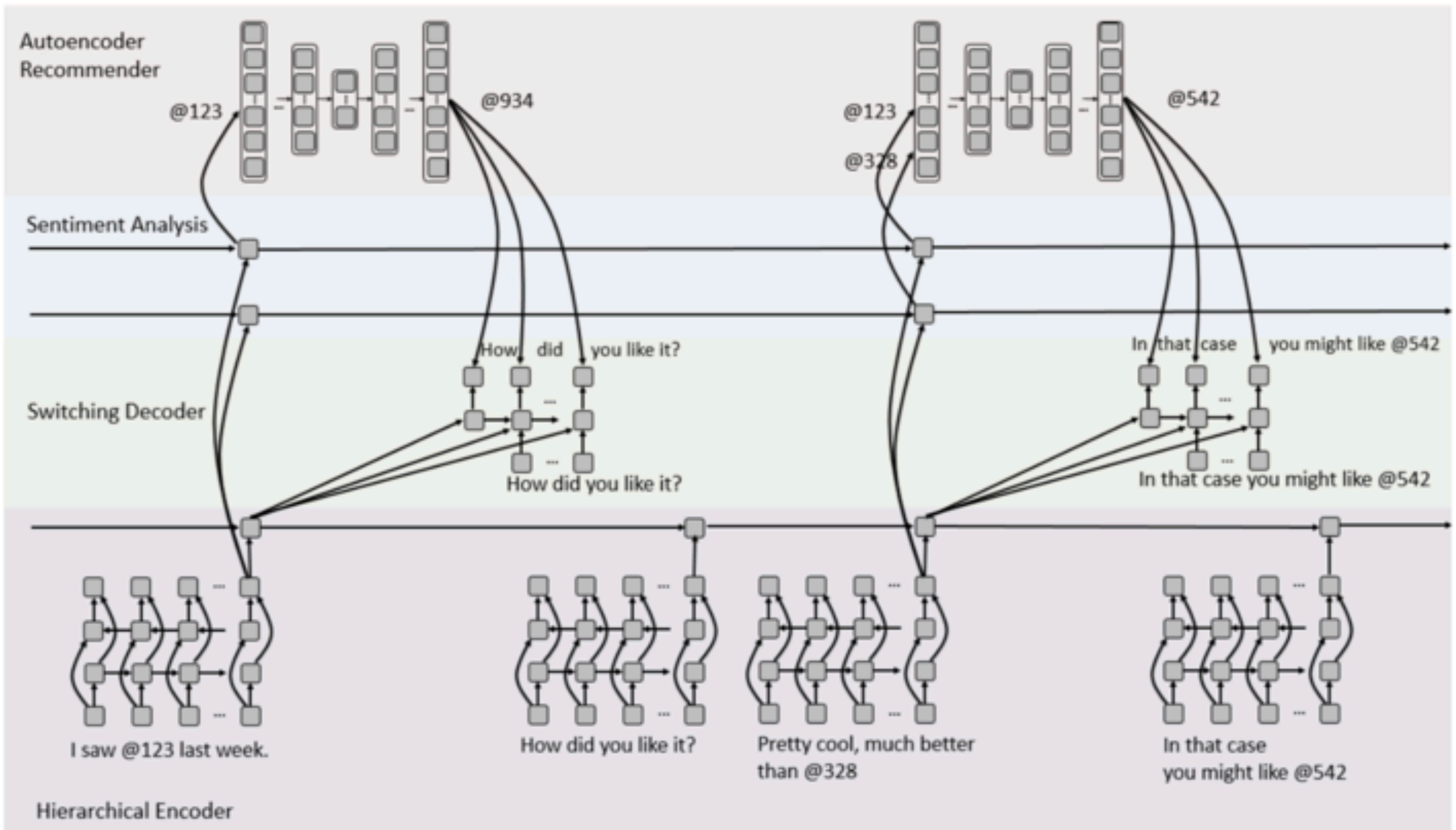


1.Pre-training denoting auto-encoder recommender on a large dataset

$$L_{\mathbf{R}}(\theta) = \sum_{u=1}^{M} \|\mathbf{r}^{(u)} - h(\mathbf{r}^{(u)};\theta)\|_{\mathcal{O}}^2 + \lambda\|\theta\|^2$$

2.Tuned on the conversational recommendation dataset with the ‘liked vector’ of the sentiment classifier as input

Dialogue Generator



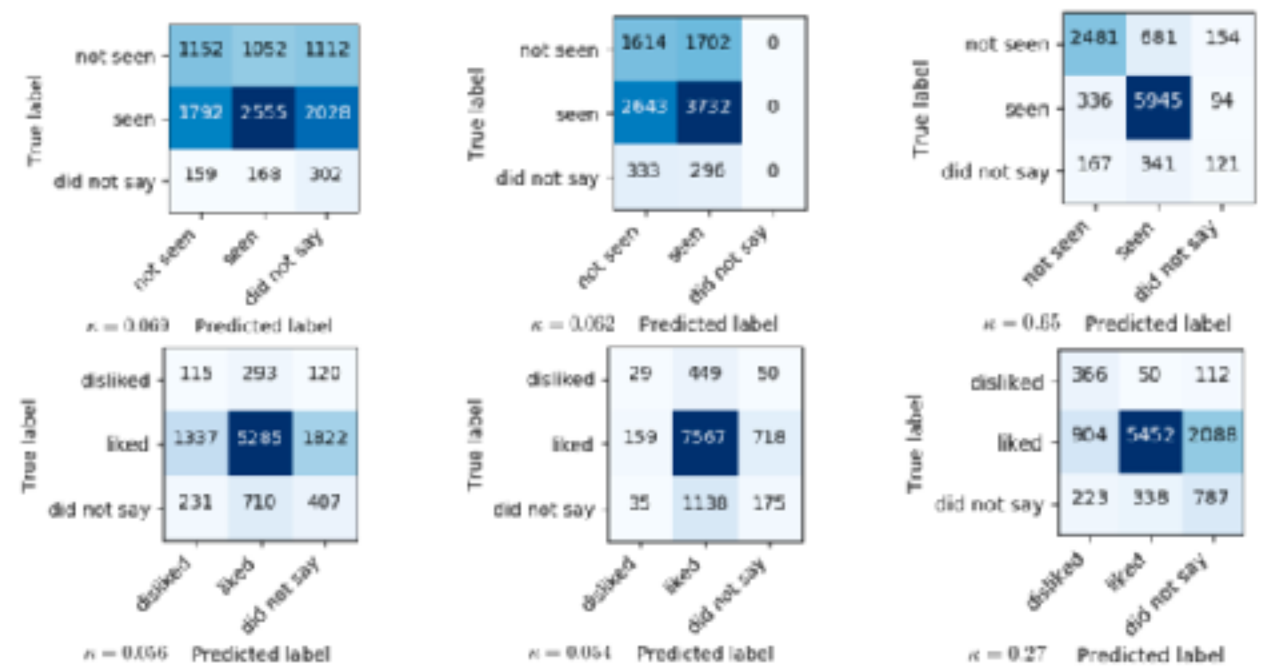
1.Both the dialogue state and the recommender state are incorporated as inputs

2.For each step in the generation, there is a switch gate controlling whether to generate or recommend

3.The recommender state is fixed during dialogue generation

Performance

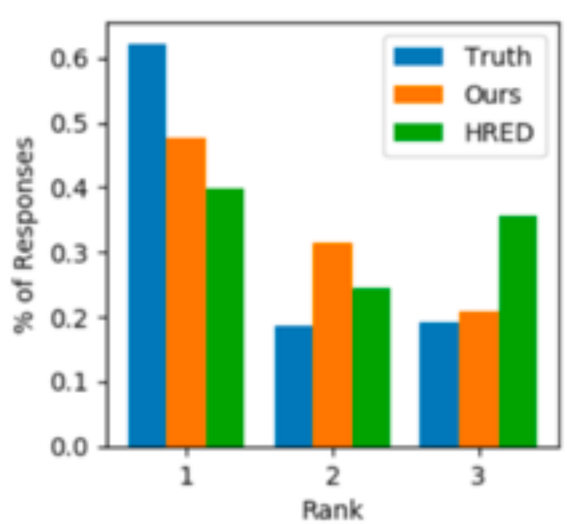
Sentiment Classification



Recommendation

Training procedure	Experiments on MovieLens	Experiments on ReDial	
		No pre-training	Pre-trained on MovieLens
Standard Baseline	0.182 ± 0.0002 (0.820)	0.35	0.29
Denoising Autorec	$\mathbf{0.179} \pm 0.0002$ (0.805)	0.33	**0.28**

Dialogue Generation



Dialogue Plus

1.User Modeling

Addressee Identification

Speaker Identification

2.Dialogue Based Application

Recommendation

Image Retrieval

3.Dialogue Content Mining

Dialogue Act Classification

Structure Mining

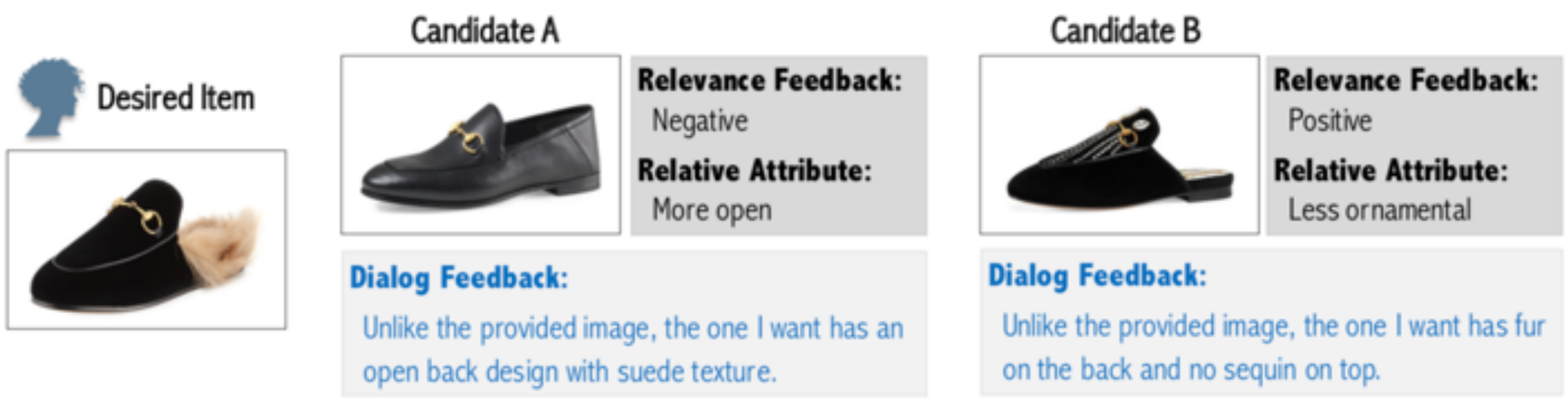
Interest Mining

Inference&Understanding

Dialog-based Interactive Image Retrieval

——NIPs 2018

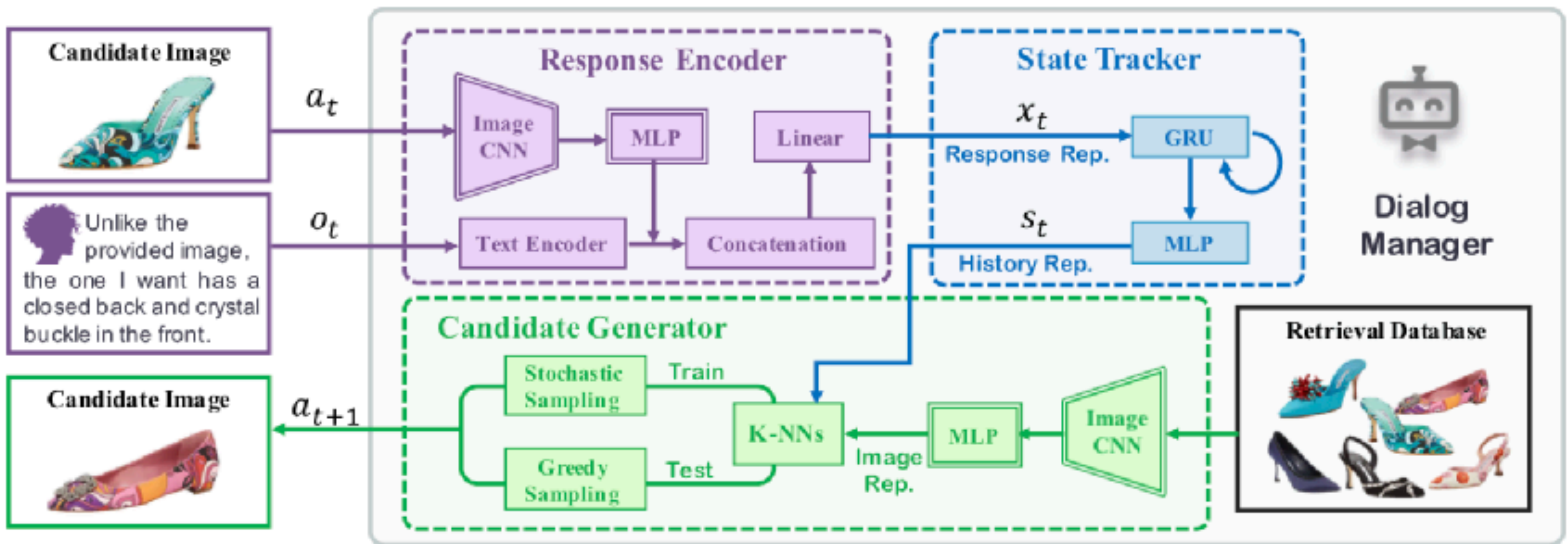
Task Description:



Objective:

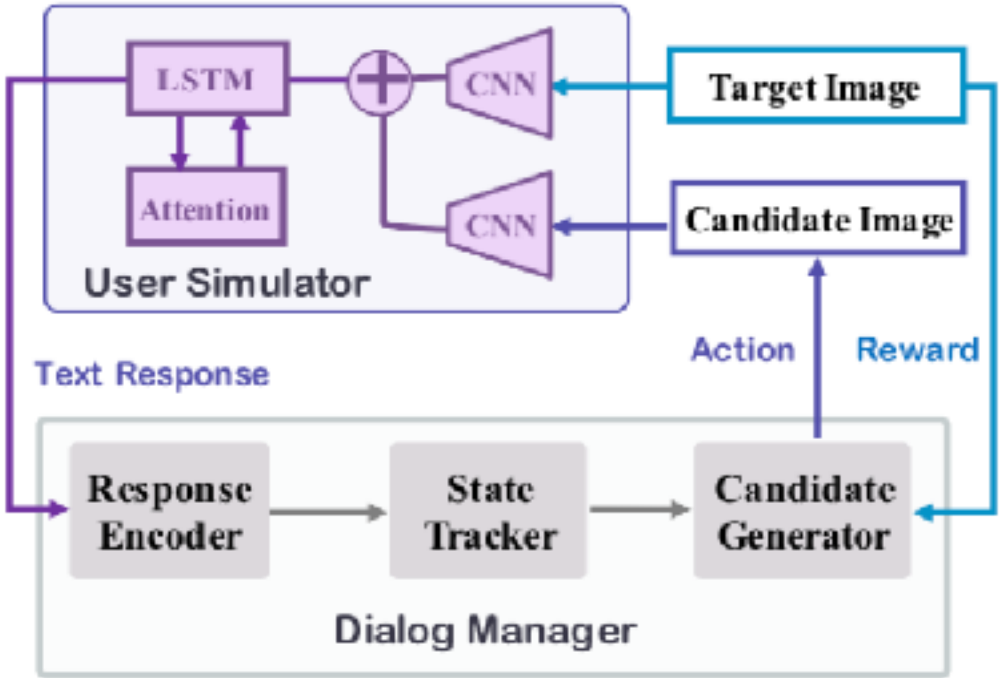
To minimize the rank of the desired item

Architecture



Dialogue Simulator

Step1:AMT workers are recruited to annotate relative image captions for image pairs
Step2: A relative-image-captioner is trained based on annotated data



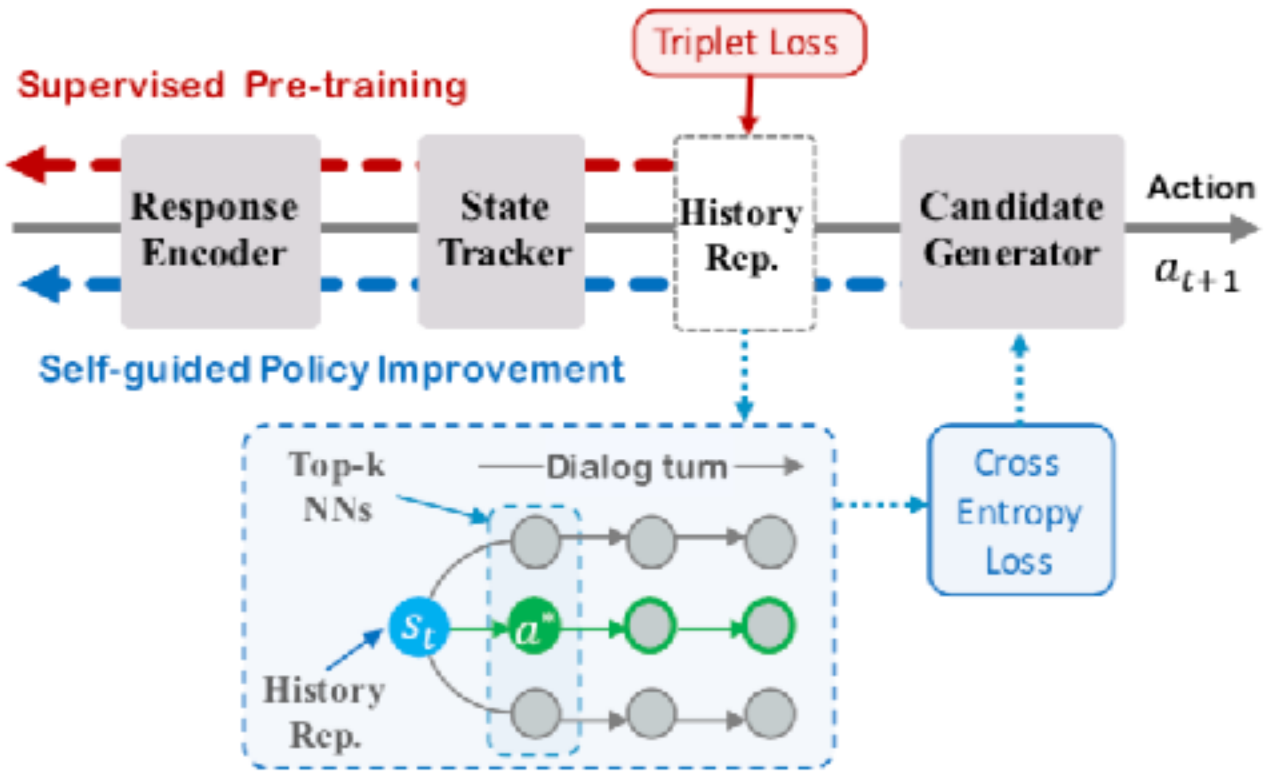
Training:

Supervised Pre-training

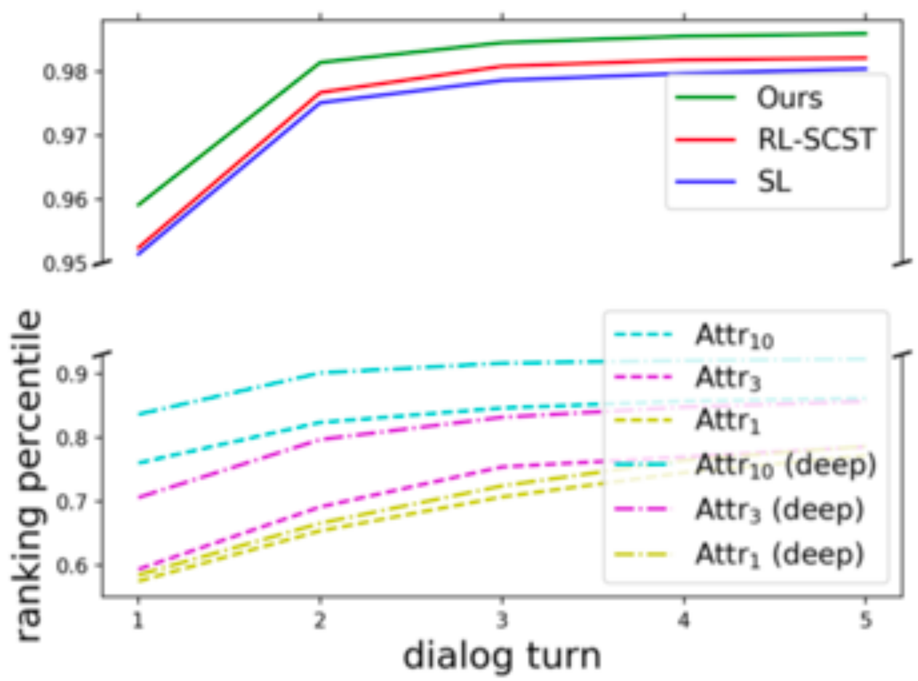
$$\mathcal{L}^{\text{sup}} = \mathbb{E}\Big[\sum_{t=1}^{T} \max(0, \|s_t - x^+\|_2 - \|s_t - x^-\|_2 + \mathrm{m})\Big]$$

Reinforcement Learning

$$\mathcal{L}^{\text{imp}} = \mathbb{E}\Big[-\sum_{t=1}^{T} \log\Big(\pi(a_t^*|h_t)\Big)\Big]$$



Performance



Dialogue Plus

1.User Modeling

Addressee Identification

Speaker Identification

2.Dialogue Based Application

Recommendation

Image Retrieval

3.Dialogue Content Mining

- Dialogue Act Classification

Structure Mining

Interest Mining

Inference&Understanding

Neural-based Context Representation Learning for Dialog Act Classification

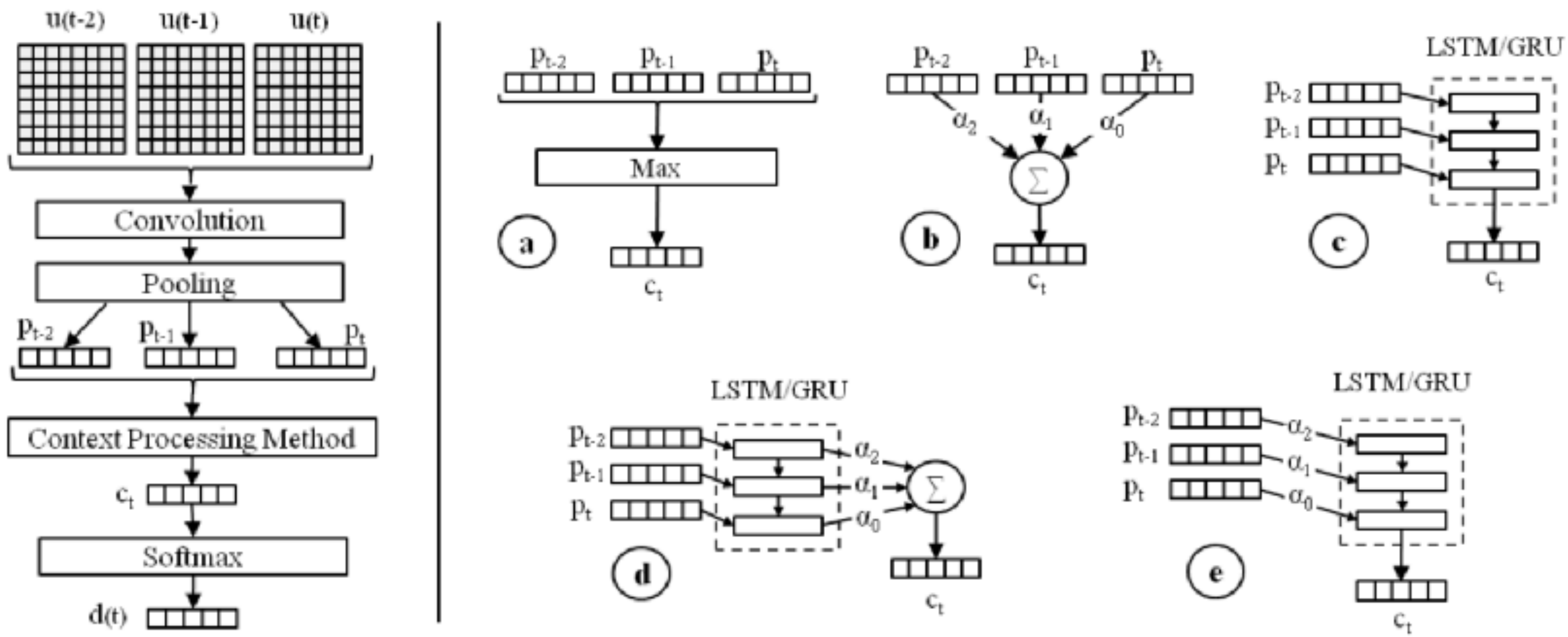
———SIGDIAL 2017

1.Sequence Labeling: Each utterance corresponds to an action label
2.Dataset:

Dataset	C	|V|	Train	Validation	Test
MRDA	5	12k	78k	16k	15k
SwDA	43	20k	193k	23k	5k

name	act_tag	example
Statement-non-opinion	sd	Me, I'm in the legal department.
Acknowledge (Backchannel)	b	Uh-huh.
Statement-opinion	sv	I think it's great
Agree/Accept	aa	That's exactly it.
Abandoned or Turn-Exit	%	So, -
Appreciation	ba	I can imagine.
Yes-No-Question	qy	Do you have to have any special training?
Non-verbal	x	[Laughter], [Throat_clearing]
Yes answers	ny	Yes.
Conventional-closing	fc	Well, it's been nice talking to you.
Uninterpretable	%	But, uh, yeah
Wh-Question	qw	Well, how old are you?
No answers	nn	No.
Response Acknowledgement	bk	Oh, okay.
Hedge	h	I don't know if I'm making any sense or not.
Declarative Yes-No-Question	qy^d	So you can afford to get a house?
Other	fo_o_fw_by_bc	Well give me a break, you know.
Backchannel in question form	bh	Is that right?
Quotation	^q	You can't be pregnant and have cats
Summarize/reformulate	bf	Oh, you mean you switched schools for the kid
Affirmative non-yes answers	na	It is.
Action-directive	ad	Why don't you go first
Collaborative Completion	^2	Who aren't contributing.
Repeat-phrase	b^m	Oh, fajitas
Open-Question	qo	How about you?
Rhetorical-Questions	qh	Who would steal a newspap
Hold before answer/agreement	^h	I'm drawing a blank.
Reject	ar	Well, no
Negative non-no answers	ng	Uh, not a whole lot.
Signal-non-understanding	br	Excuse me?
Other answers	no	I don't know
Conventional-opening	fp	How are you?
Or-Clause	qrr	or is it more of a company?
Dispreferred answers	arp_nd	Well, not so much that.
3rd-party-talk	t3	My goodness, Diane, get do
Offers, Options, Commits	oo_co_cc	I'll have to check that out
Self-talk	t1	What's the word I'm lookin
Downplayer	bd	That's all right.
Maybe/Accept-part	aap_am	Something like that
Tag-Question	^g	Right?
Declarative Wh-Question	qw^d	You are what kind of buff?
Apology	fa	I'm sorry.
Thanking	ft	Hey thanks a lot

Architecture



Dialogue Plus

1.User Modeling

- Addressee Identification
- Speaker Identification

2.Dialogue Based Application

- Recommendation
- Image Retrieval

3.Dialogue Content Mining

- Dialogue Act Classification
- Structure Mining
- Interest Mining
- Inference&Understanding

Find The Conversation Killers: A Predictive Study of Thread-ending Posts

——WWW 2018

Objective: identifying a post that is unlikely to be further replied to

Motivation :improve the engagement of users into the conversations.

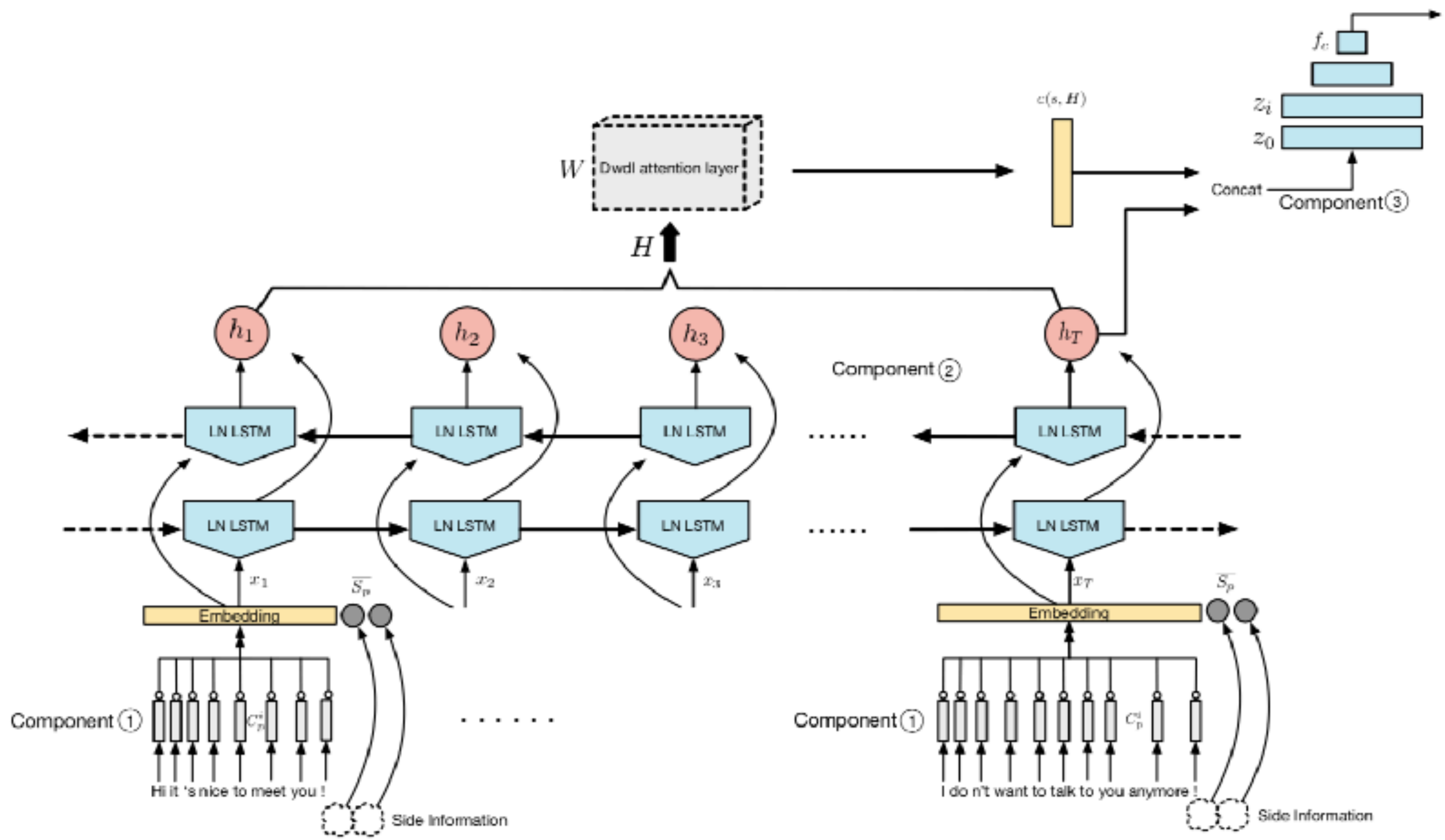
Conversartion 1

A: Oh, God! Oh God!
B: Just be cool.
A: It's a mine, isn't it?
B: Just relax.
A: How'm I gonna relax standing on a mine!?
(Following Omitted)……

Conversartion 2

A: Do you think she will give us the designs?
B: Eventually. These things are always a matter of leverage.
A: And you think O'Brien is that leverage?
B: That remains to be seen.

Architecture



Statistics

Properties	Reddit-Threads	Movie-Dialogs
Threads	83,097	100,000
Vocabulary	29,729	107,354
Max post len.	673	2689
Avg. post len.	13.02 words	43.83 words
# train threads	63,097	80,000
# val threads	10,000	10,000
# test threads	10,000	10,000

Performance

	Reddit-Threads Data Set			Movie-Dialogs Data Set		
Method	Accuracy	AUC	MAP	Accuracy	AUC	MAP
SVM-Text content(Embedding, N-grams)	$75.95_{\circ\circ}$	$81.26_{\circ\circ\circ}$	$68.84_{\circ\circ\circ}$	$74.57_{\circ\circ\circ}$	$83.12_{\circ\circ\circ}$	$64.97_{\circ\circ\circ}$
SVM-Lengths info	76.45	83.05	72.31	75.70	84.67	69.50
SVM-Background info	–	–	–	$75.43_{\circ}$	84.56	$69.09_{\circ}$
SVM-Post time	$75.55_{\circ\circ\circ}$	$81.36_{\circ\circ\circ}$	$69.85_{\circ\circ\circ}$	–	–	–
SVM-Replying structures	76.15	83.13	72.67	–	–	–
SVM-Sentiment	76.31	83.06	72.84	75.60	84.59	69.59
SVM+All features	76.39	83.30	72.60	75.84	84.67	69.63
BiLSTM+Text content (only the target post)	$60.80^{\star\star\star}_{\circ\circ\circ}$	$64.36^{\star\star\star}_{\circ\circ\circ}$	$58.20^{\star\star\star}_{\circ\circ\circ}$	$61.62^{\star\star\star}_{\circ\circ\circ}$	$61.40^{\star\star\star}_{\circ\circ\circ}$	$50.30^{\star\star\star}_{\circ\circ\circ}$
BiLSTM+Text content	$76.02_{\star\star\star}$	$83.42_{\star\star\star}$	73.33	$76.26_{\star\star\star}$	$85.22_{\star\star\star}$	$70.63_{\star\star\star}$
LNBiLSTM+Text content	$76.59_{\star\star\star}$	$84.22_{\star\star\star}$	$74.07_{\star\star\star}$	$76.75_{\star\star}$	$85.55_{\star\star\star}$	$70.85_{\star\star\star}$
Stacked LNBiLSTM+Text content	$76.42_{\star\star\star}$	$84.46_{\star\star\star}$	$74.44_{\star\star\star}$	$76.98_{\star}$	$85.87_{\star\star}$	$71.67_{\star\star}$
LNBiLSTM+All features	78.05	85.91	77.39	77.51	86.47	72.95
LNBiLSTM+All features+Standard attention	$78.05_{\circ\circ\circ}$	$85.97_{\circ\circ\circ}$	$77.70_{\circ\circ\circ}$	$77.45_{\circ\circ\circ}$	$86.32_{\circ\circ\circ}$	$72.63_{\circ\circ\circ}$
ConverNet	$78.27_{\circ\circ\circ}$	$86.22^{\star\star}_{\circ\circ\circ}$	$78.21^{\star}_{\circ\circ\circ}$	$78.04^{\star\star}_{\circ\circ\circ}$	$86.82^{\star\star\star}_{\circ\circ\circ}$	$73.76^{\star\star\star}_{\circ\circ\circ}$

Dialogue Plus

1.User Modeling

Addressee Identification

Speaker Identification

2.Dialogue Based Application

Recommendation

Image Retrieval

3.Dialogue Content Mining

Dialogue Act Classification

Structure Mining

Interest Mining

Inference&Understanding

Estimating User Interest from Open-Domain Dialogue

——SIGDIAL 2018

Interest Estimation ——Topic Prediction

Table 1: Topic Categories

Travel	Movies	Celebrities
Music	Reading	Anime / Manga
Games	Computers	Home Appliances
Beauty	Fashion	Sports / Exercise
Health	School	Outdoor Activities
Housing	Housekeeping	Marriage / Love
Animals	Family	Cooking / Meal
Vehicles	History	Politics / Economy

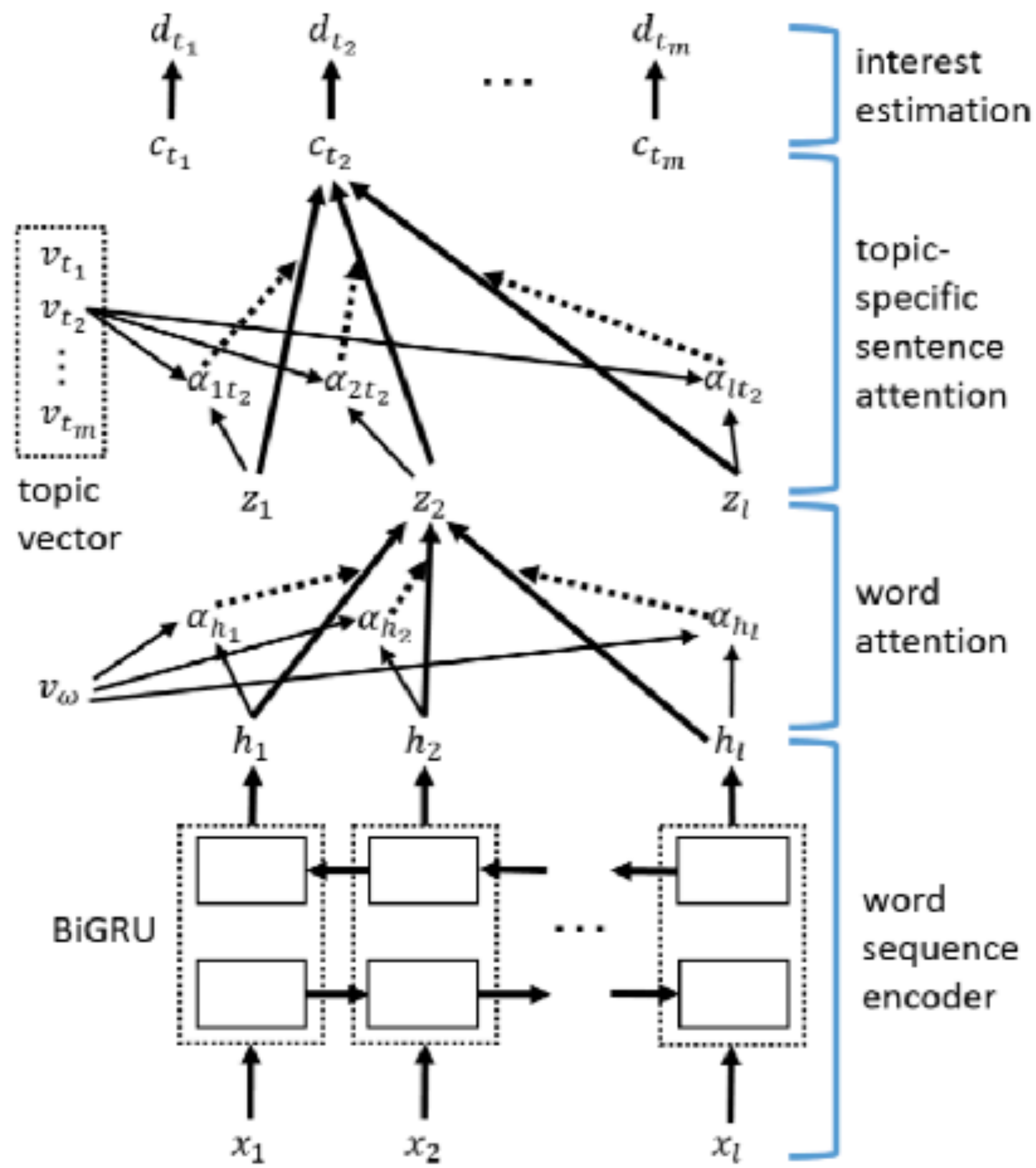
Table 3: Data Statistics

Number of users (data points)	163
Number of dialogues	408
Number of utterances	49029
Avg. number of strong interest topics	11.48
Avg. number of light interest topics	7.30
Avg. number of neutral topics	5.21

Table 2: Example dialogue (translated by authors)

A	対話を開始します。よろしくお願いします。 Let's start a conversation. Nice to meet you.
B	はい、よろしくお願いします。 Hi, nice to meet you.
A	何かご趣味はありますか？ What are your hobbies?
B	最近はペット中心の生活になっているのでペットが趣味になりますね。 Currently, I am living a pet-centered lifestyle. So, raising pets is my hobby.
A	何を飼ってらっしゃるのですか？ Which pets do you have?
B	猫を飼っています。3匹いるのでにぎやかですよ。 I have three cats and they are lively.
A	3匹ですか、いいですね！雑種ですか？ Three cats. That sounds great! Are they mixed breed?
B	はい、全部雑種です。手がかからなくて楽ですね。何か動物は飼っていますか？ Yes, they are all mixed breed cats. They are low-maintenance and easy to keep. Do you have any animals?

Architecture

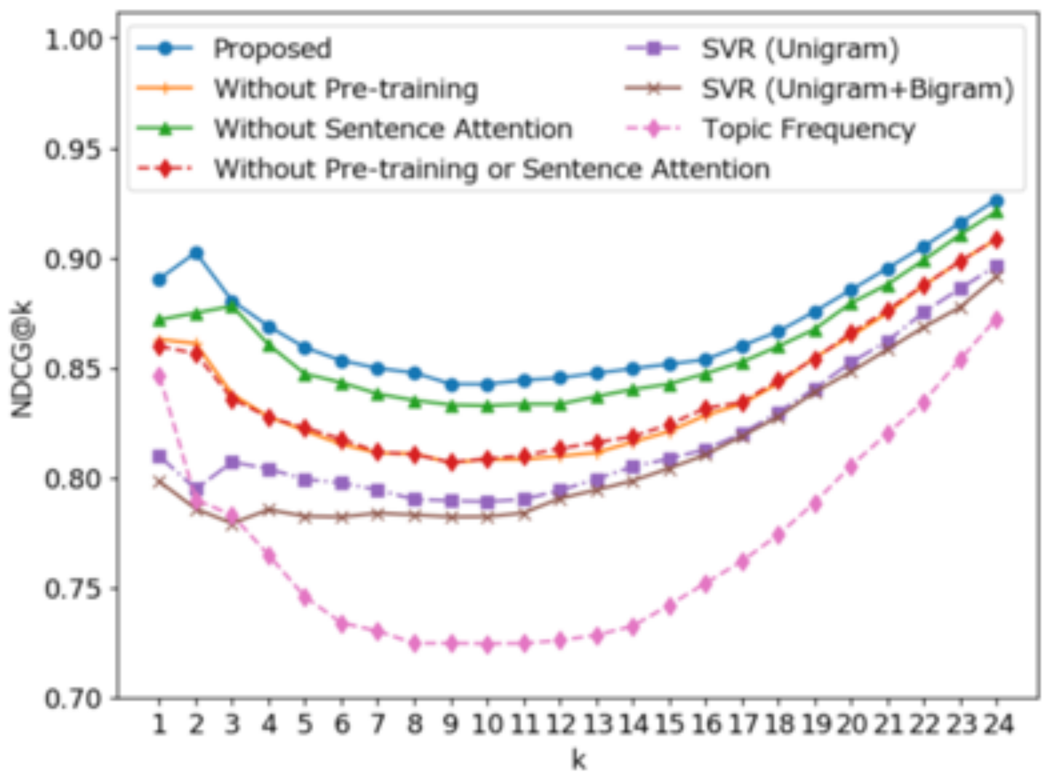


$$L = \frac{1}{n}\sum_{i}^{n}(y_i - d_{t_i})^2$$

Performance

Table 4: Mean Squared Error

Proposed	**0.533**
Without Pre-Training	0.580
Without Sentence Attention	0.561
Without Pre-Training or Sentence Attention	0.568
SVR (unigram)	0.597
SVR (unigram + bigram)	0.611



Dialogue Plus

1.User Modeling

Addressee Identification

Speaker Identification

2.Dialogue Based Application

Recommendation

Image Retrieval

3.Dialogue Content Mining

Dialogue Act Classification

Structure Mining

Interest Mining

- Inference&Understanding

Conversational Decision-Making Model for Predicting the King's Decision in the Annals of the Joseon Dynasty ——ACL 2018

Task

Statistics

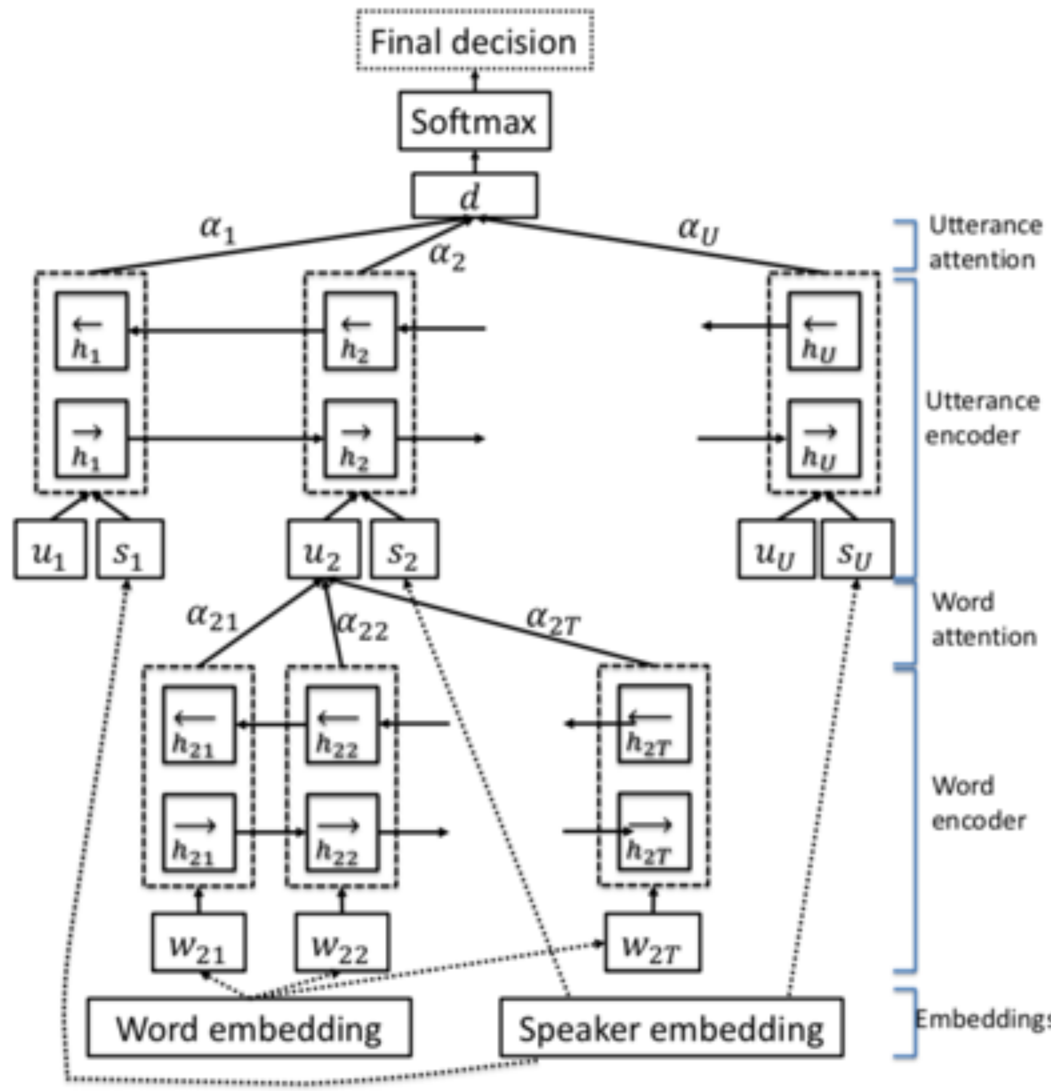
Kings	Articles	Utterances	Participants
15	13,216	95,615	4,502

(a) Basic statistics of the corpus

Order	1,996	*Accept*	1,457
Approve	2,245	*Reject*	818
Disapprove	468	*Discuss*	6,214

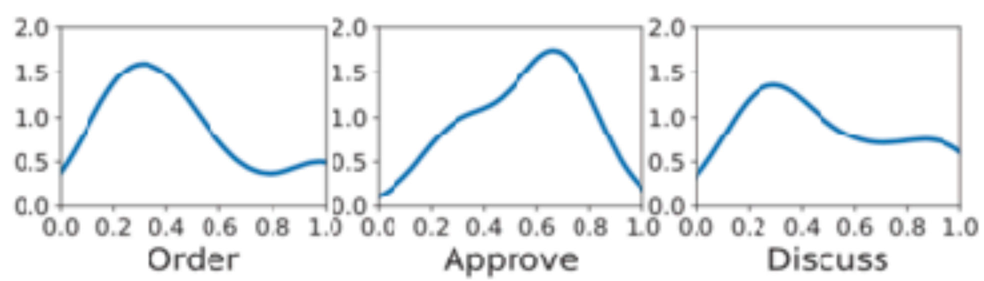
(b) Distribution of articles over decisions

Architecture

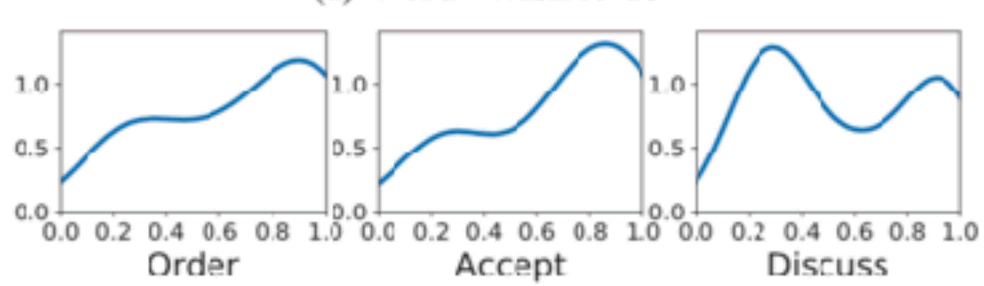


Performance

Method	*Micro* F_1	*Macro Prec*	*Macro Rec*	*Macro* F_1	*W-avg* F_1
Majority of classes	0.472	0.079	0.167	0.107	0.303
Naive Bayes	0.479	0.173	0.176	0.126	0.321
SVM linear	0.381	0.249	0.246	0.246	0.383
SVM RBF	0.487	0.236	0.186	0.142	0.337
Naive Bayes with speaker	0.466	0.268	0.177	0.135	0.323
SVM linear with speaker	0.423	0.292	0.259	0.243	0.403
SVM RBF with speaker	0.472	0.079	0.167	0.107	0.303
fastText w/o word vector	0.487	0.158	0.193	0.150	0.349
fastText	0.499	0.315	0.225	0.215	0.402
CDMM w/o speaker	0.481	0.176	0.214	0.178	0.379
CDMM with speaker (random init)	**0.504**	0.258	0.227	0.208	0.401
CDMM with speaker (pre-trained)	0.476	**0.329**	**0.307**	**0.313**	**0.456**

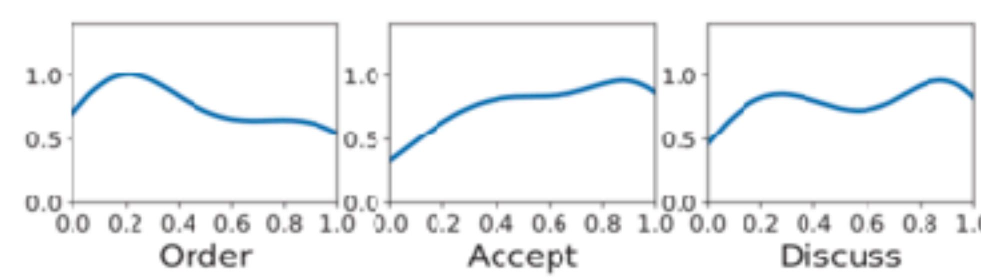


(a) Word "Wish to do"

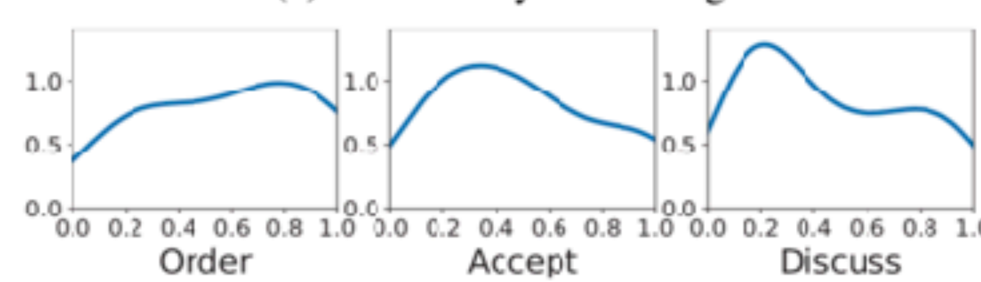


(b) Word "Okay"

Figure 3: Attention weight distribution of words for each class



(a) Word "Okay" from kings



(b) Word "Okay" from officials

Figure 4: Attention weight distribution of word for each class from kings and officials

Dialogue Plus

1.User Modeling

Addressee Identification

Speaker Identification

2.Dialogue Based Application

Recommendation

Image Retrieval

3.Dialogue Content Mining

Dialogue Act Classification

Structure Mining

Interest Mining

Inference&Understanding

Q&A