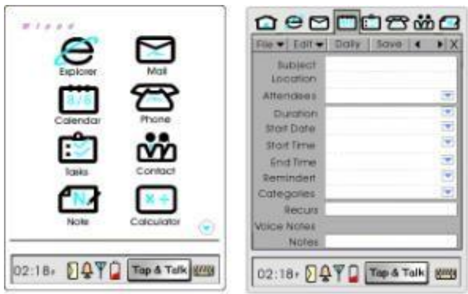
Towards Different Perspectives in Automatic Human-Computer Conversational Systems

#Go_ChatBots!!

Rui Yan, Peking University

http://www.ruiyan.me

Brief History

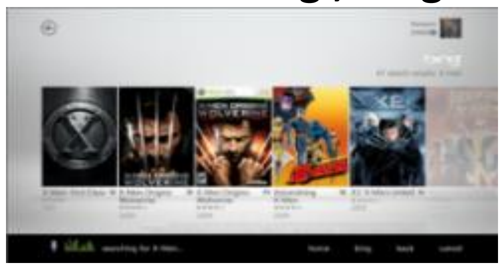


Task-specific argument extraction

(e.g., Nuance, SpeechWorks)

User: "I want to fly from Boston to New York next week."

Early 1990s



TV Voice Search

e.g., Bing on Xbox

Early 2000s

2017

Virtual Personal Assistants

Chatbots

Intent Determination

(Nuance's Emily™, AT&T HMIHY) *User: "Uh…we want to move…we want to change our phone line from this house to another house"*

Keyword Spotting

(e.g., AT&T)

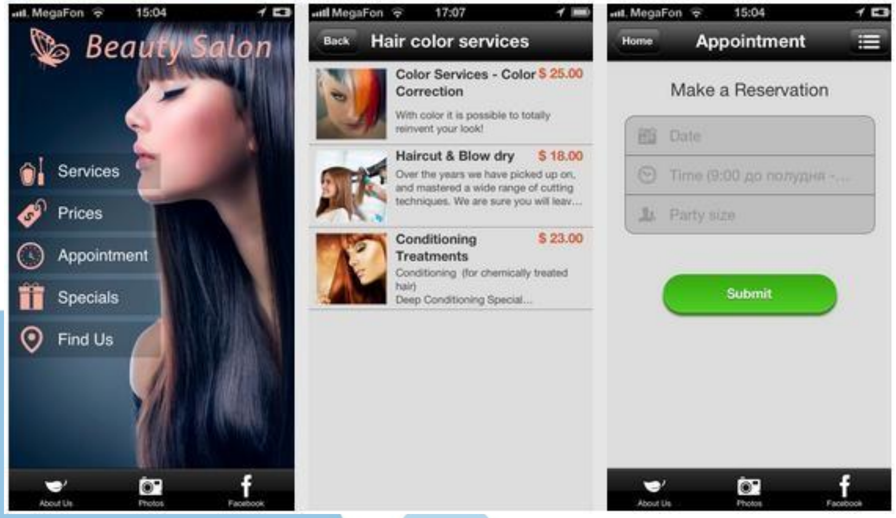
System: *"Please say collect, calling card, person, third number, or operator"*

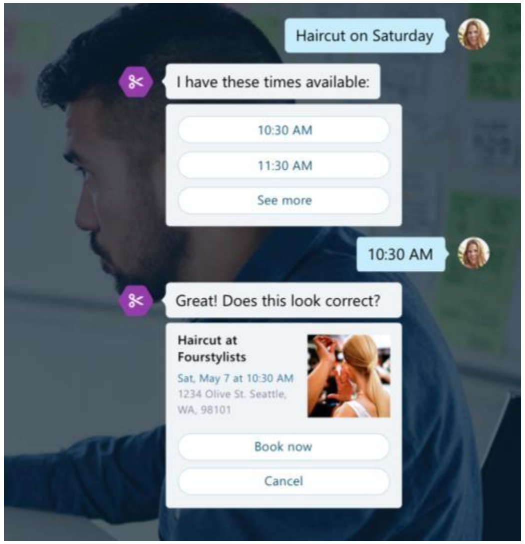
Why Do We Need

- Get things done
  - E.g. set up alarm/reminder, take note
- Easy access to structured data, services and apps
  - E.g. find docs/photos/restaurants
- Assist your daily schedule and routine
  - E.g. commute alerts to/from work
- Be more productive in managing your work and personal life

APP->Bots

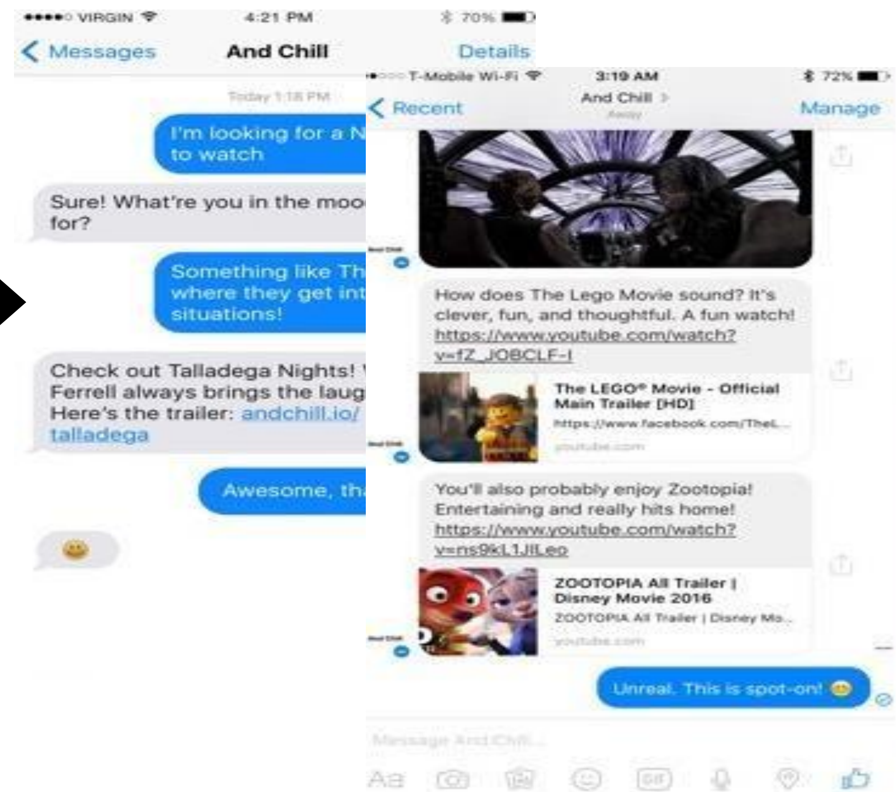
- A bot is responsible for a “single” domain, similar to an app





Users can initiate dialogues instead of following the GUI design

GUI -> Conversational UI



GUI vs. CUI

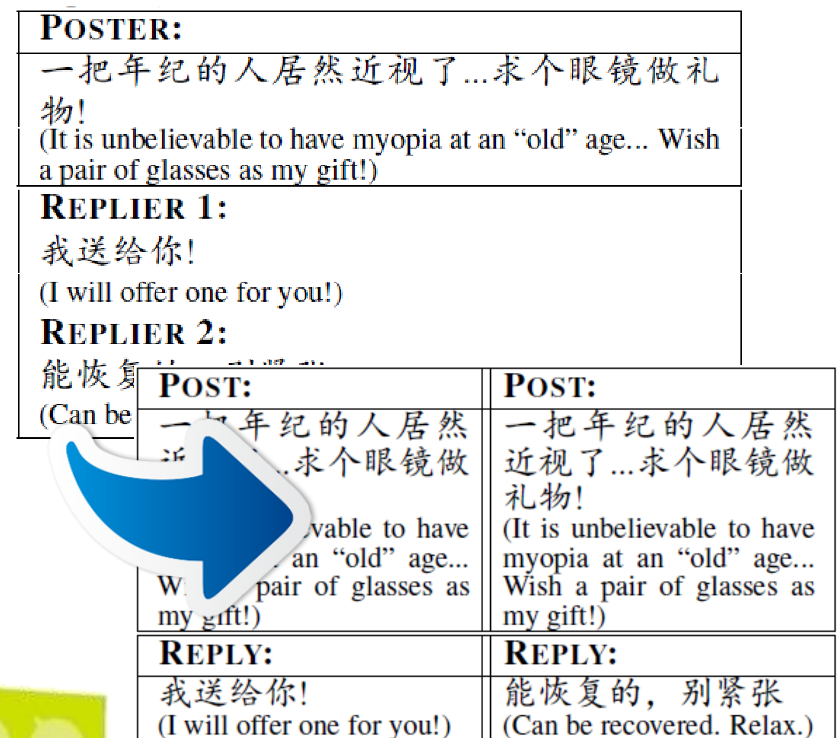
	Website/APP's GUI	Msg's CUI
Situation	Navigation, no specific goal	Searching, with specific goal
Information Quantity	More	Less
Information Precision	Low	High
Display	Structured	Non-structured
Interface	Graphics	Language
Manipulation	Click	mainly use texts or speech as input
Learning	Need time to learn and adapt	No need to learn
Entrance	App download	Incorporated in any msg-based interface
Flexibility	Low, like machine manipulation	High, like converse with a human

Era of A.I.

- **Conversation systems with A.I. prevail** (>///<)
  - Virtual personal assistant
    - Apple Siri/Microsoft Cortana/Google Now
  - ChatBot
    - Baidu Duer, Microsoft (Xiaobing, Rinna, Tay)
  - Yet to come: Facebook, Microsoft, more startups…

Conversation

- **What is conversation**
  - Given **q**, respond with **r**
- **Why is it possible?**
  - It is all about timing
  - Data-driven v.s. big data
    - 10 million is enough?
- **Why is it challenging?**
  - Needless to mention
  - Relevance
  - Interestingness
  - A lot of issues…



Background Knowledge

- **Machine learning ≈ looking for a function**

Speech Recognition

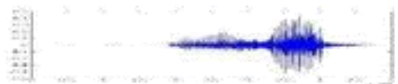
$f($  $)=$ "你好 (Hello) "

Image Recognition

$f($  $)=$ cat

Go Playing

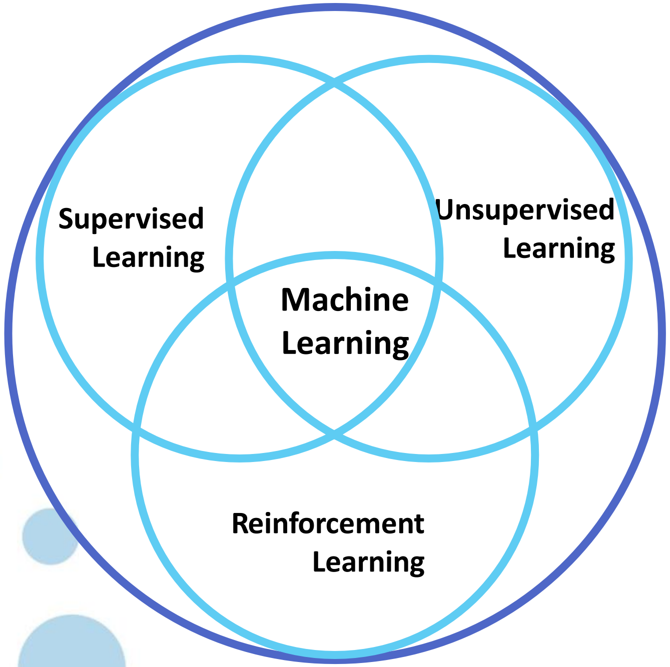
$f($  $)=$ 5-5 (next move)

Chat Bot

$f($"Where is Westin?" $)=$ "The address is..."

Given a large amount of data, the machine learns what the function f should be.

Machine Learning

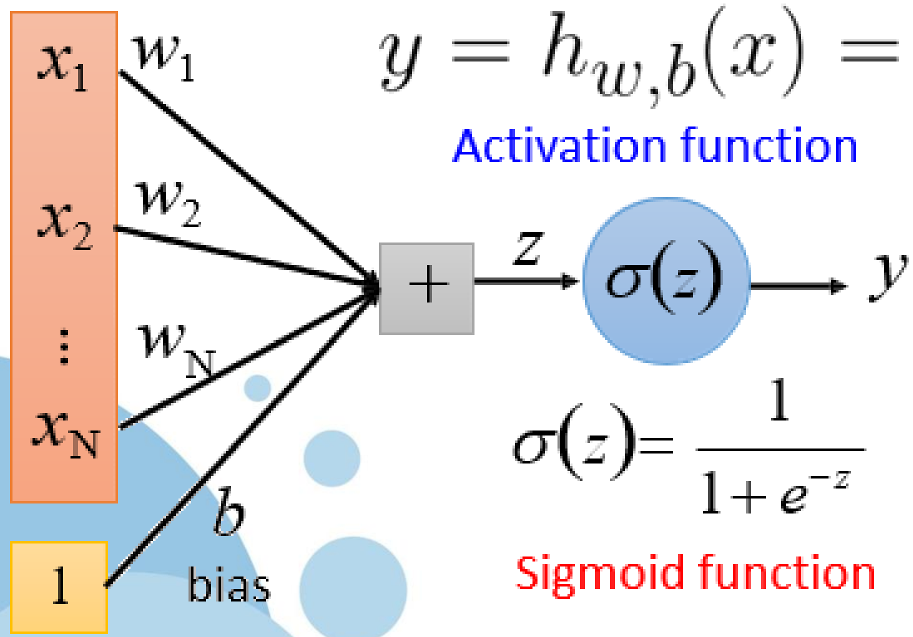


Deep learning is a type of machine learning approaches, called “neural networks”.

A Single Neuron

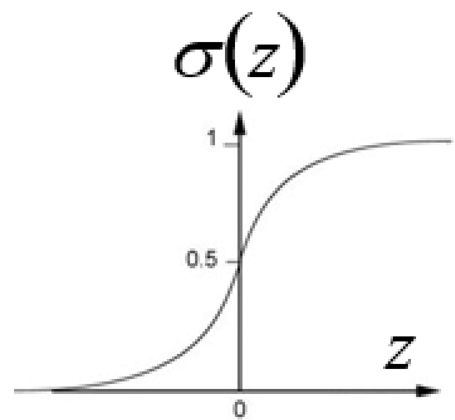
$$y = h_{w,b}(x) = \sigma(w^T x + b)$$

Activation function



$$\sigma(z) = \frac{1}{1+e^{-z}}$$

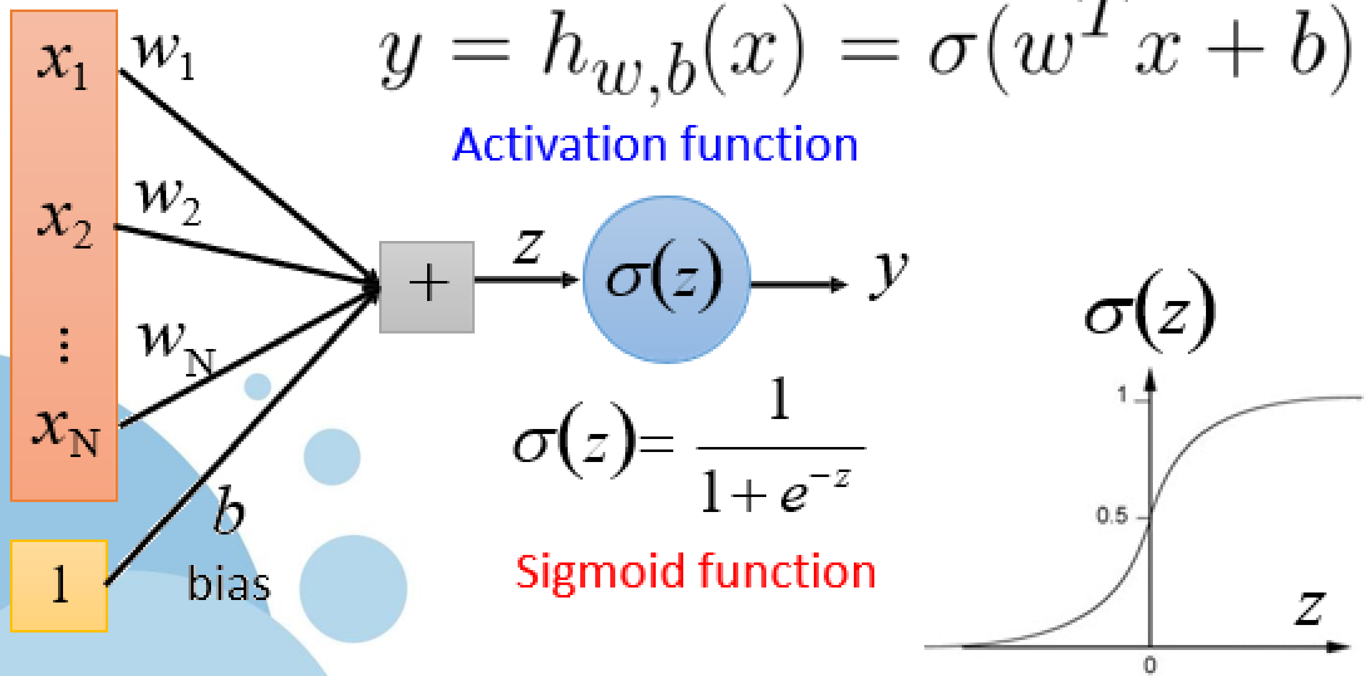
Sigmoid function



w, b are the parameters of this neuron

A Single Neuron

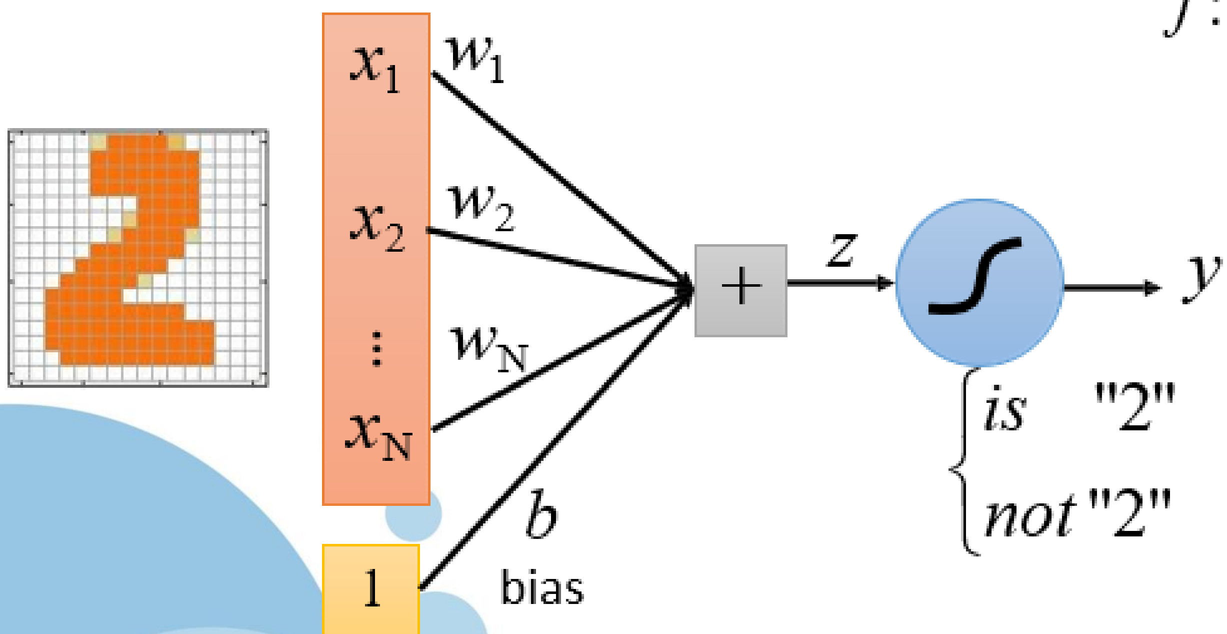
$$y = h_{w,b}(x) = \sigma(w^T x + b)$$



w, b are the parameters of this neuron

How Does It Work?

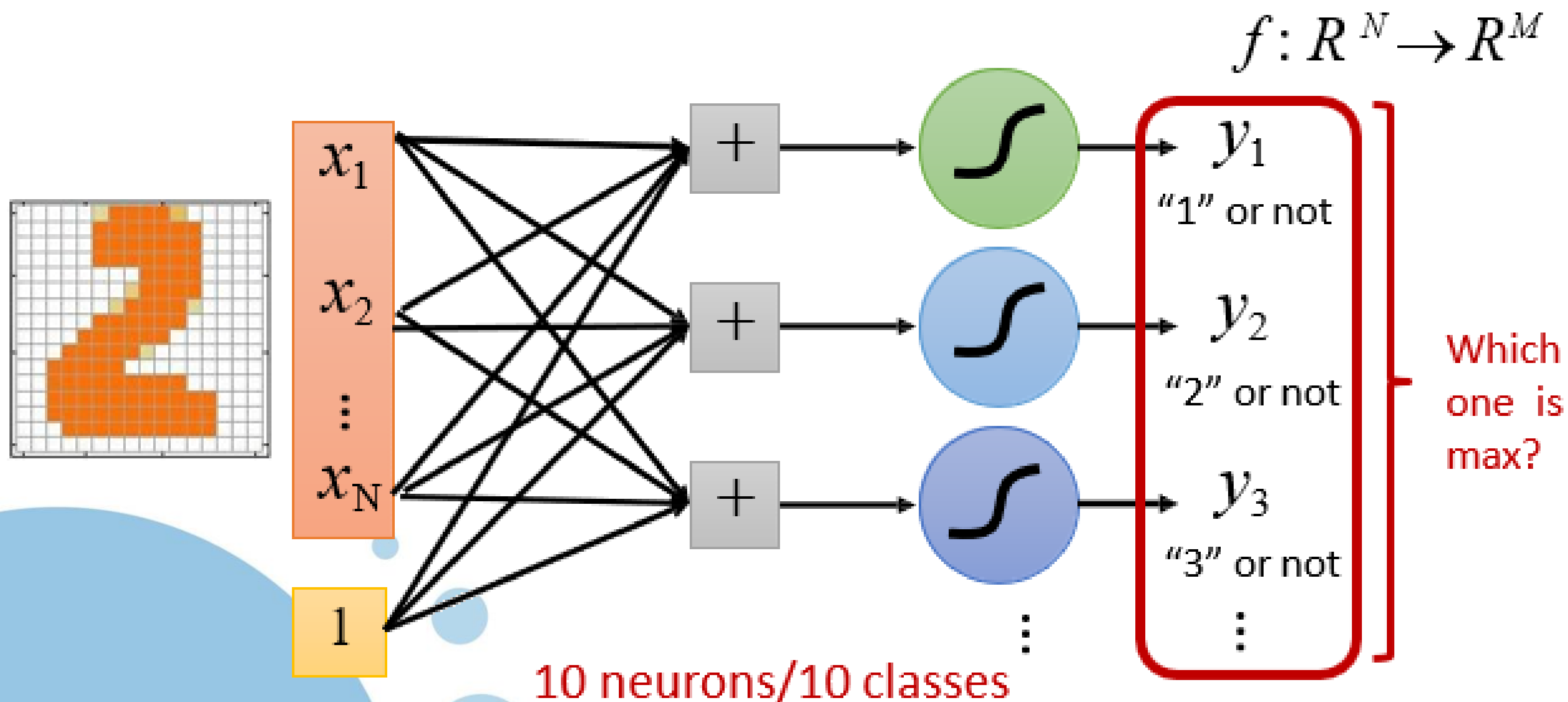
$$f: R^N \rightarrow R^M$$



$$\begin{cases} is \quad "2" & y \geq 0.5 \\ not\ "2" & y < 0.5 \end{cases}$$

A single neuron can only handle binary classification

A Layer of Neurons

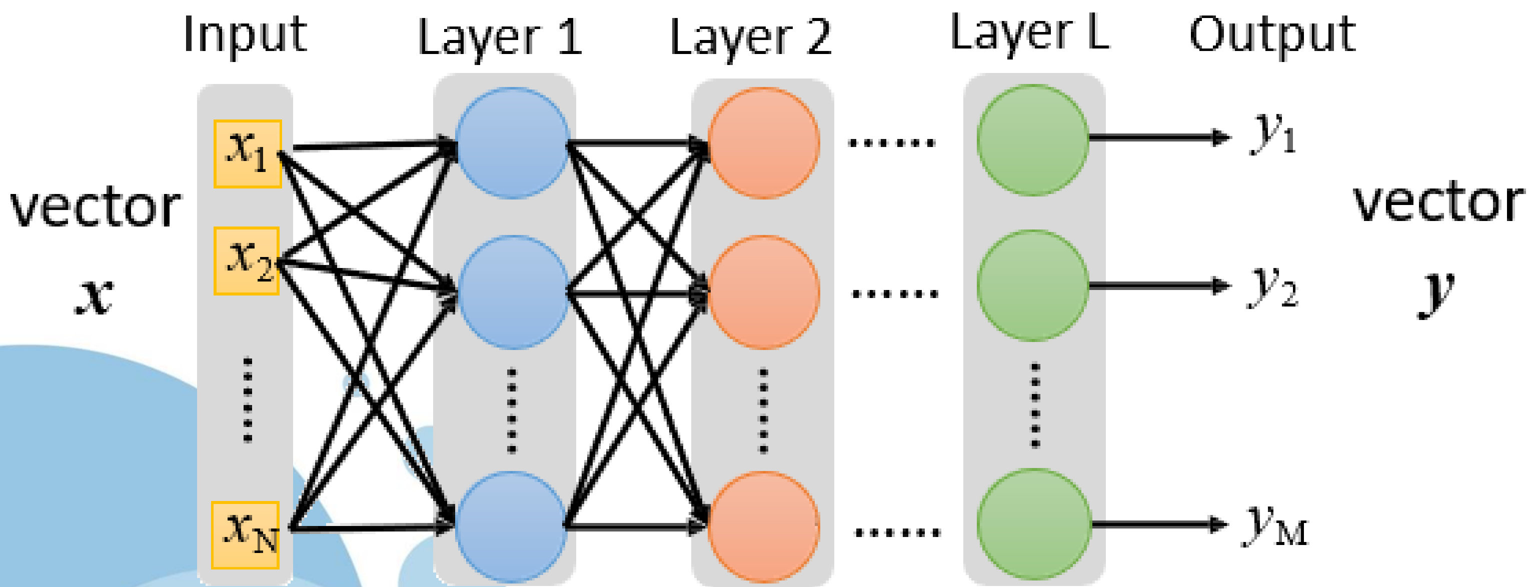


A layer of neurons can handle multiple possible output, and the result depends on the max one

Deep Neural Networks

Fully connected feedforward network

$$f : R^N \to R^M$$



Deep NN: multiple hidden layers

Categorizations

- **Domain**
  - Open-domain
  - Vertical domains
- **How to obtain a reply?**
  - Retrieval-based methods
  - Generation-based methods
  - Combination of retrieval- and generation-based methods
- **Scenarios**
  - Single-turn conversation
  - Multi-turn conversation
- **Style**
  - Passive conversation
  - Proactive conversation

Categorizations

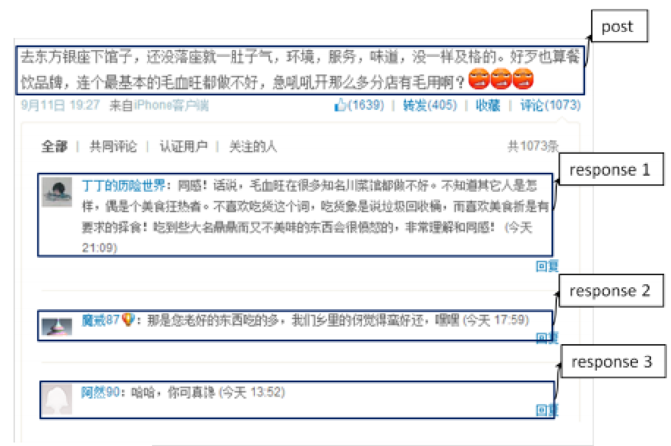
- **Domain**
  - Open-domain
  - Vertical domains
- **How to obtain a reply?**
  - Retrieval-based methods
  - Generation-based methods
  - Combination of retrieval- and generation-based methods
- **Scenarios**
  - Single-turn conversation
  - Multi-turn conversation
- **Style**
  - Passive conversation
  - Proactive conversation

RETRIEVAL-BASED CONVERSATION SYSTEM

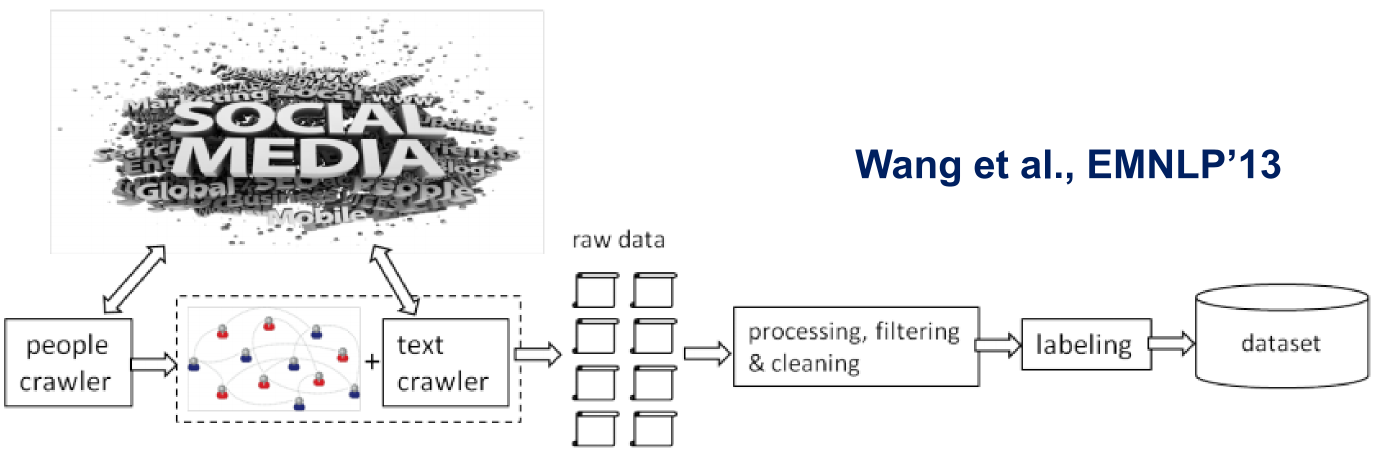
Dataset

- **Web provides opportunities with big data**
  - Social media, cQA, BBS forums

Post	
User A:	*The first day at Hawaii. Watching sunset at the balcony with a big glass of wine in hand.*
Responses	
User B:	*Enjoy it & don't forget to share your photos!*
User C:	*Please take me with you next time!*
User D:	*How long are you going to stay there?*
User E:	*When will be your talk?*
User F:	*Haha, I am doing the same thing right now. Which hotel are you staying in?*
User G:	*Stop showing-off, buddy. We are still coding crazily right now in the lab.*
User H:	*Lucky you! Our flight to Honolulu is delayed and I am stuck in the airport. Chewing French fries in MacDonald's right now.*



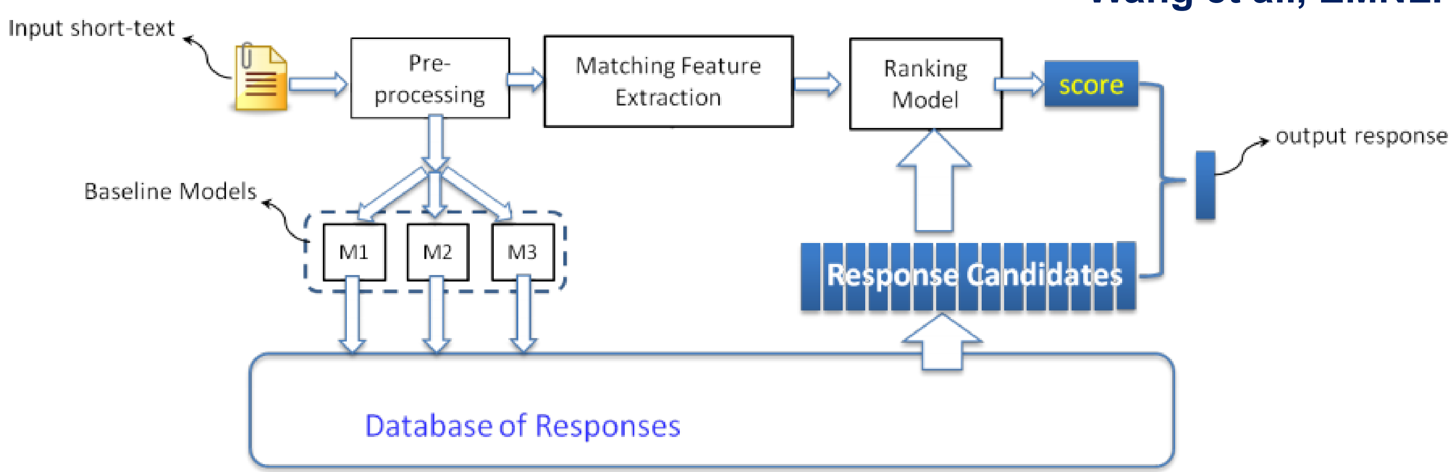
Wang et al., EMNLP'13



Retrieval Process

- **Retrieval model: 2-stage retrieval**
  - Fast retrieval: fast matching
    - Post-response semantic matching (mapping to low-dimension vectors)
    - Post-response similarity (vsm)
    - Post-post similarity
  - Linear match: $$\text{score}(x, y) = \sum_{i \in \Omega} w_i \Phi_i(x, y)$$

Wang et al., EMNLP'13



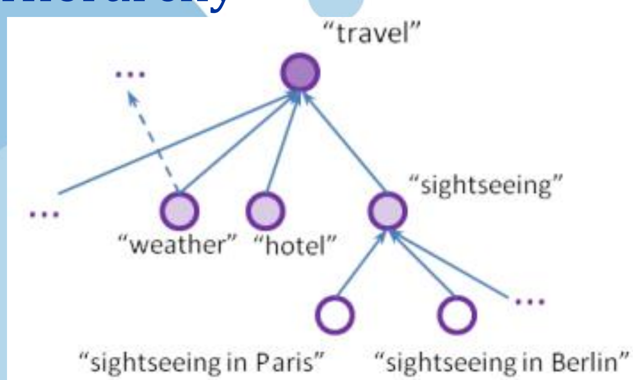
Deep Match

- **Essence**
  - Matching: inner-product of two representing feature vectors

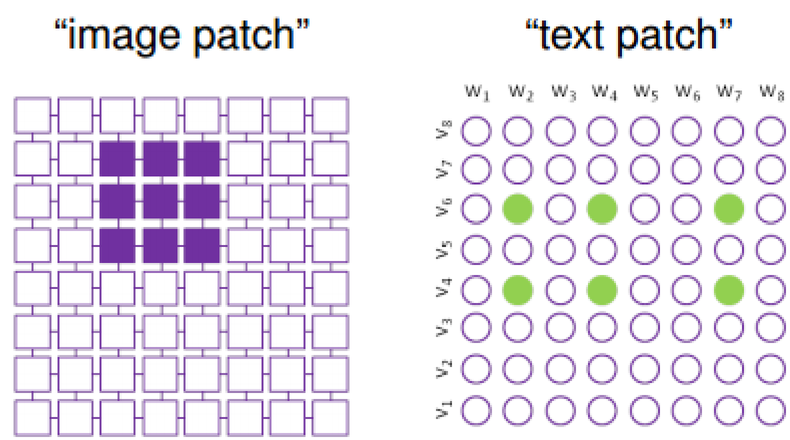
$$\text{match}(x, y) = < \Phi_{\mathcal{Y}}(x), \Phi_{\mathcal{X}}(y) >_{\mathcal{H}}$$

$$\text{match}(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{A} \mathbf{y} = \sum_{m=1}^{D_x} \sum_{n=1}^{D_y} A_{nm} x_m y_n$$

- **From Linear to Deep**
  - Localness
  - Hierarchy

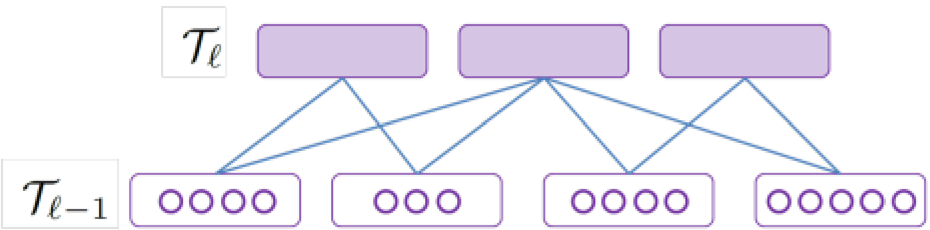


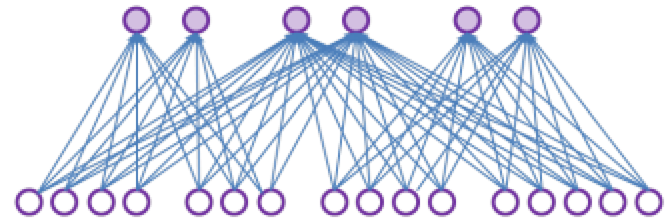
Lu et al., NIPS'13



Neural Network Structure

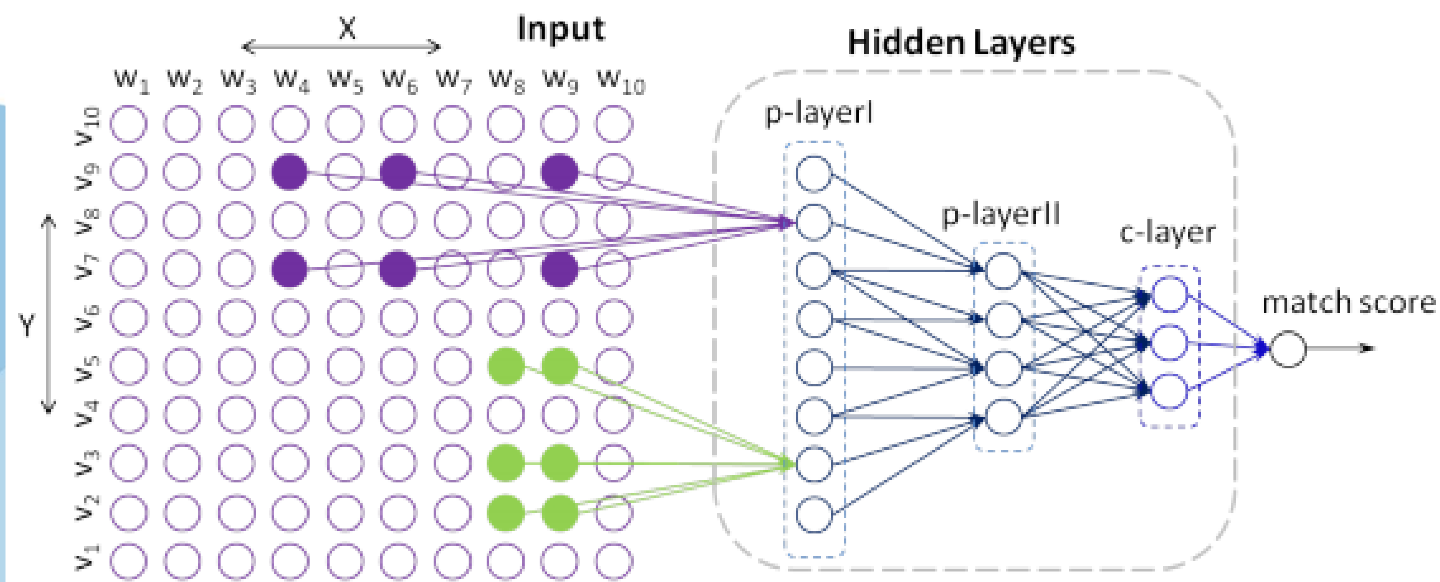
- **Connect with topic patterns**





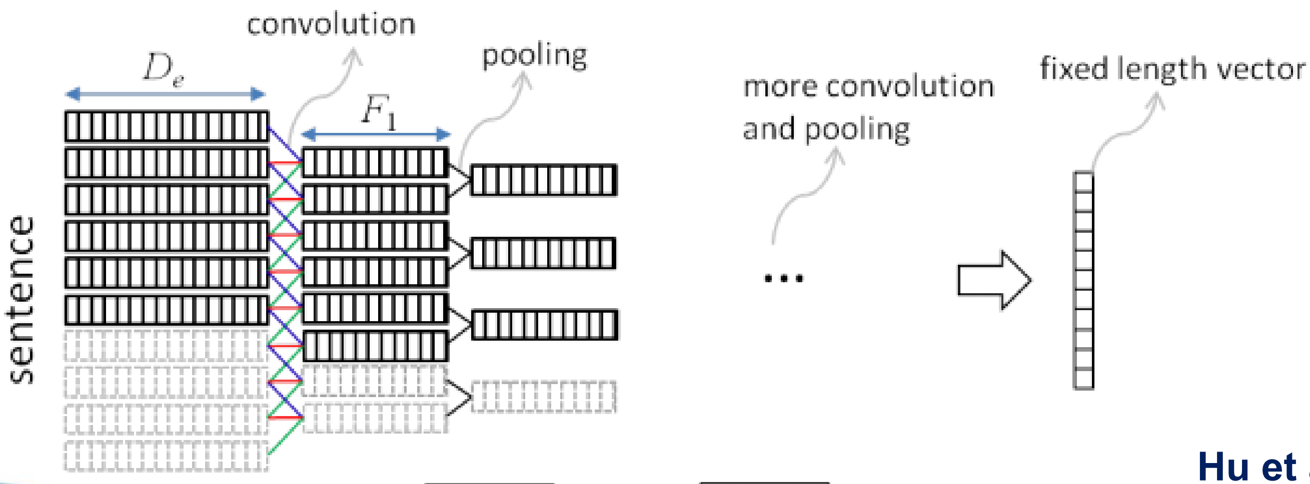
- **Matching architecture**

Lu et al., NIPS'13

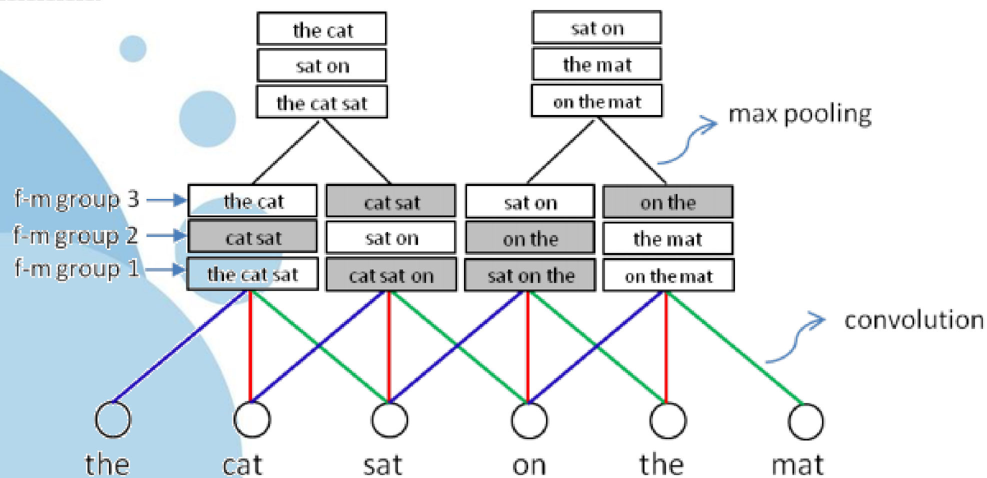


Convolutions

- **Convolutional sentence model**



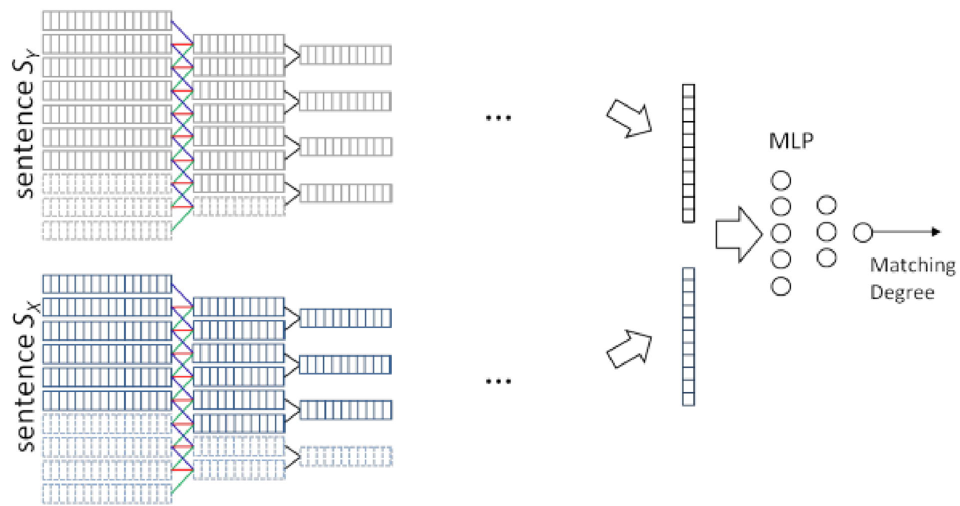
Hu et al., NIPS'14



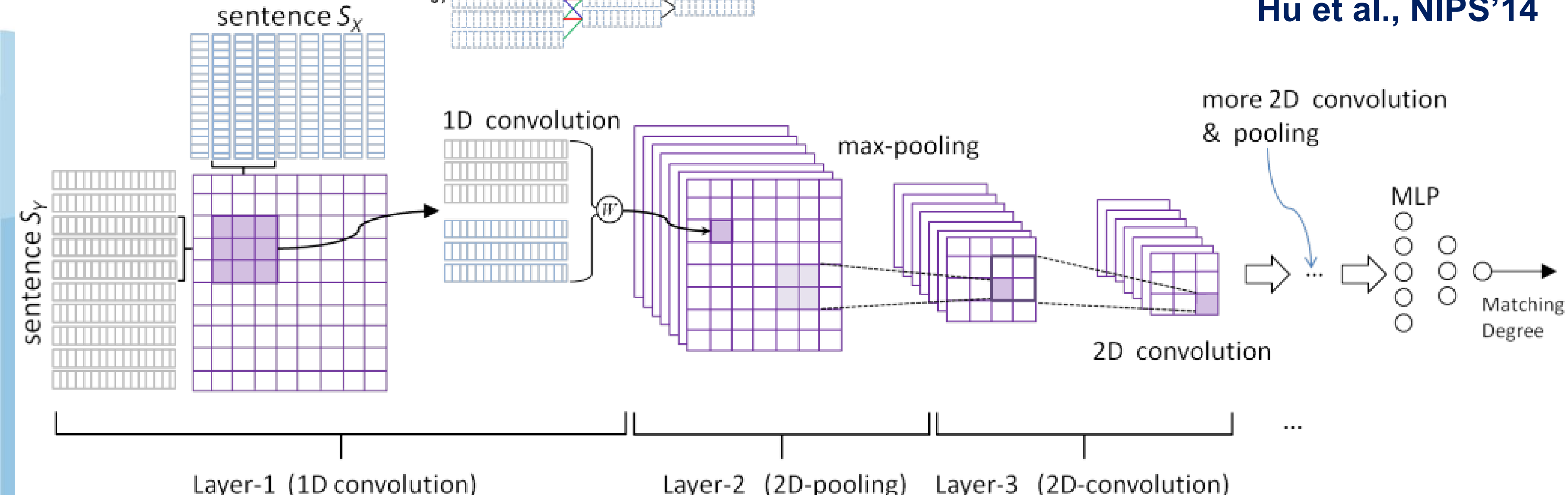
- **Illustration of convolutional sentence model**

Convolutional Match

- **ARC-I and ARC-II matching model**

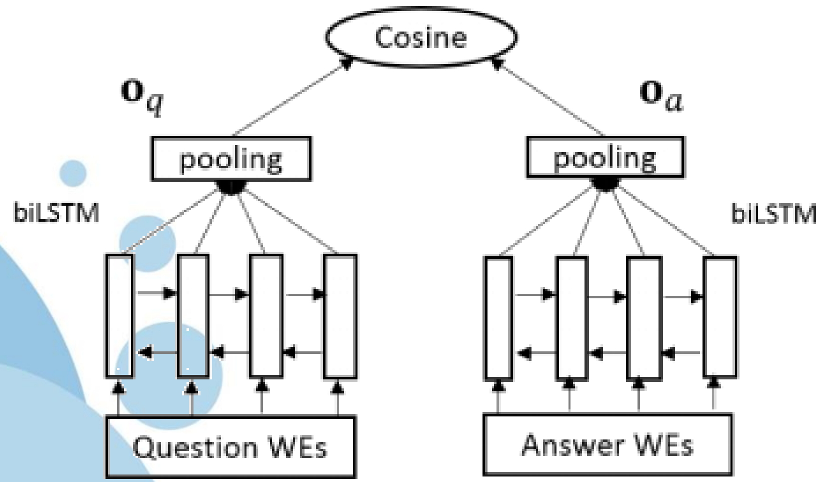


Hu et al., NIPS'14



Recurrent Modeling

- **Question-Answer Matching**
  - Standard LSTM
    - Concatenation of the last vectors on both directions of the biLSTM
    - Average pooling over all the output vectors of the biLSTM
    - Max pooling over all the output vectors

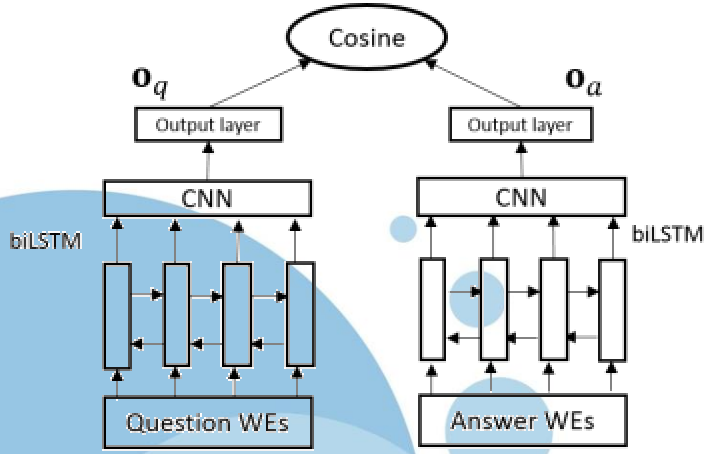


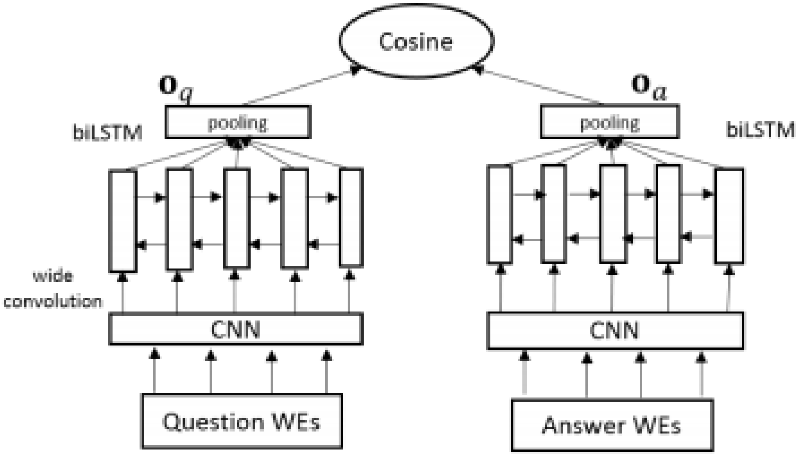
Tan et al., ACL'16

CNN+RNN Match

- **Question-Answer Matching**
  - Convolutional LSTM
    - LSTM first, then convolution
    - Convolution first, then LSTM
    - Based on results: more or less the same

Tan et al., ACL'16





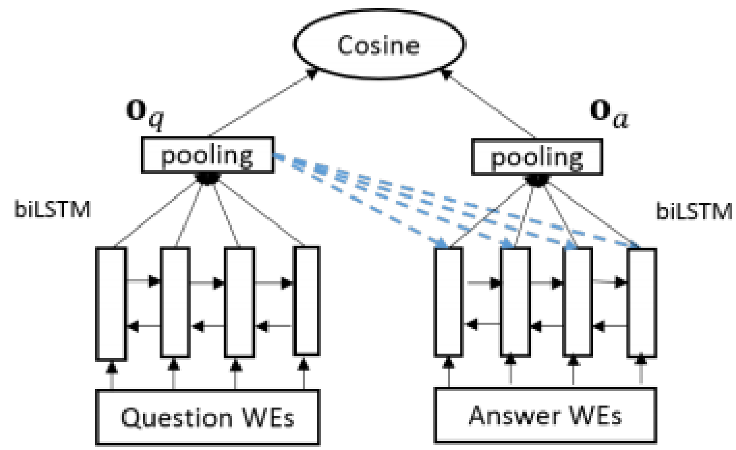
E	Conv-pooling LSTM (c=4000,K=1)	66.2	64.6	62.2
F	Conv-pooling LSTM (c=200,K=50)	66.4	67.4	63.5
G	Conv-pooling LSTM (c=400,K=50)	67.8	67.5	64.4
H	Conv-based LSTM ($\|h\|$=200,K=50)	66.0	66.1	63.0
I	Conv-based LSTM ($\|h\|$=400,K=50)	67.1	67.6	64.4

CNN+RNN Match + Attention

- **Question-Answer Matching**
  - Attentive matching

Tan et al., ACL'16

$$
\begin{aligned}
\mathbf{m}_{a,q}(t) &= \mathbf{W}_{am}\mathbf{h}_a(t) + \mathbf{W}_{qm}\mathbf{o}_q \\
s_{a,q}(t) &\propto \exp(\mathbf{w}_{ms}^T \tanh(\mathbf{m}_{a,q}(t))) \\
\widetilde{\mathbf{h}}_a(t) &= \mathbf{h}_a(t) s_{a,q}(t)
\end{aligned}
$$

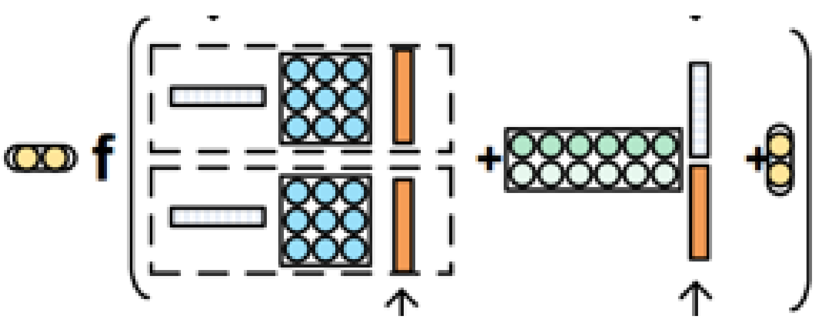


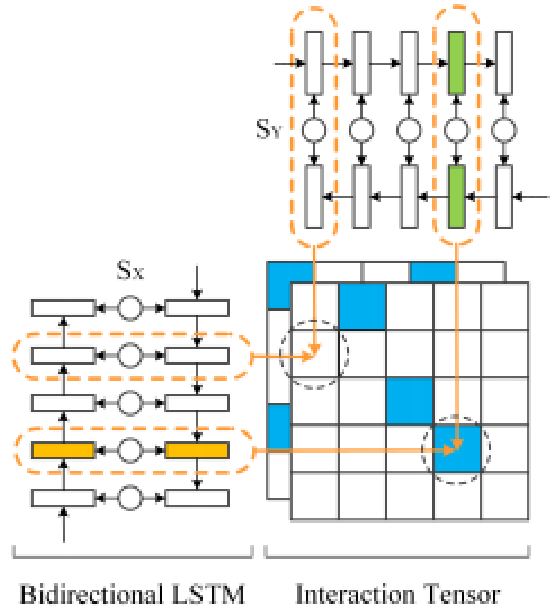
$$\tilde{\mathbf{h}}_a(t) = \mathbf{h}_a(t) s_{a,q}(t)$$

L	Attentive LSTM (avg-pooling K=1)	68.4	68.1	62.2
M	Attentive LSTM (avg-pooling K=50)	68.4	67.8	63.2
N	Attentive LSTM (max-pooling K=50)	**68.9**	**69.0**	**64.8**

Positional Matching

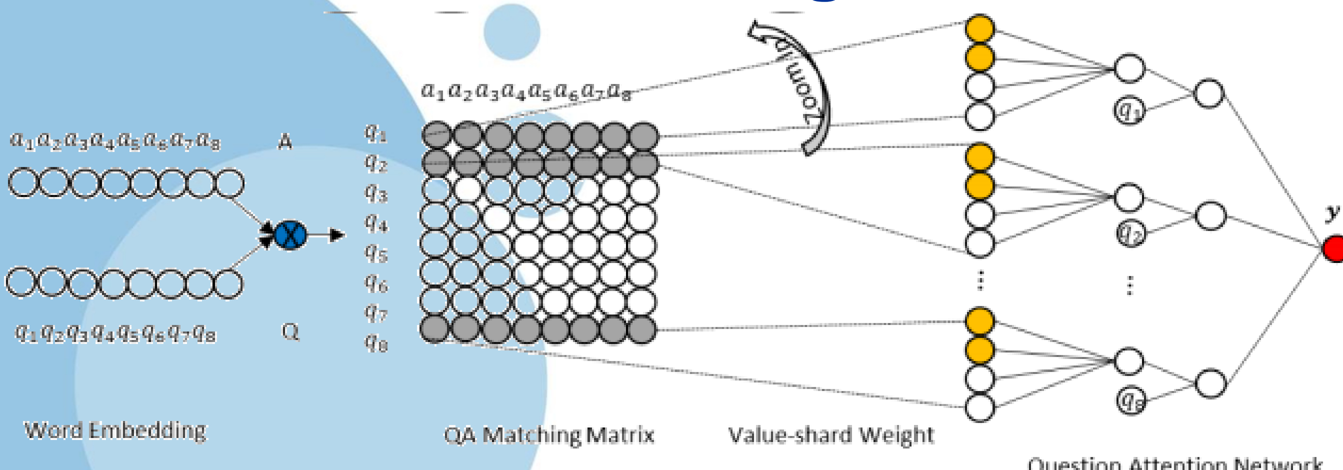
- **Positional-based match**





Wan et al., AAAI'16

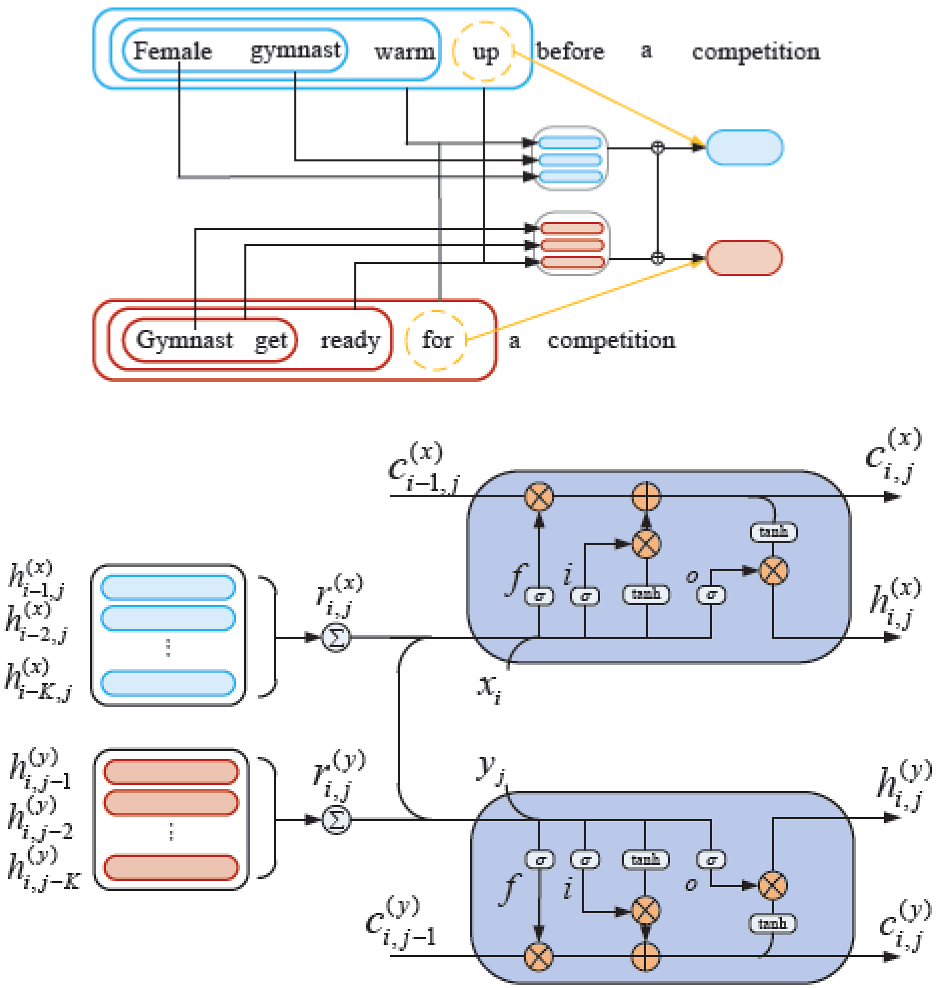
- **Attention-based ranking**



Liu et al., CIKM'16

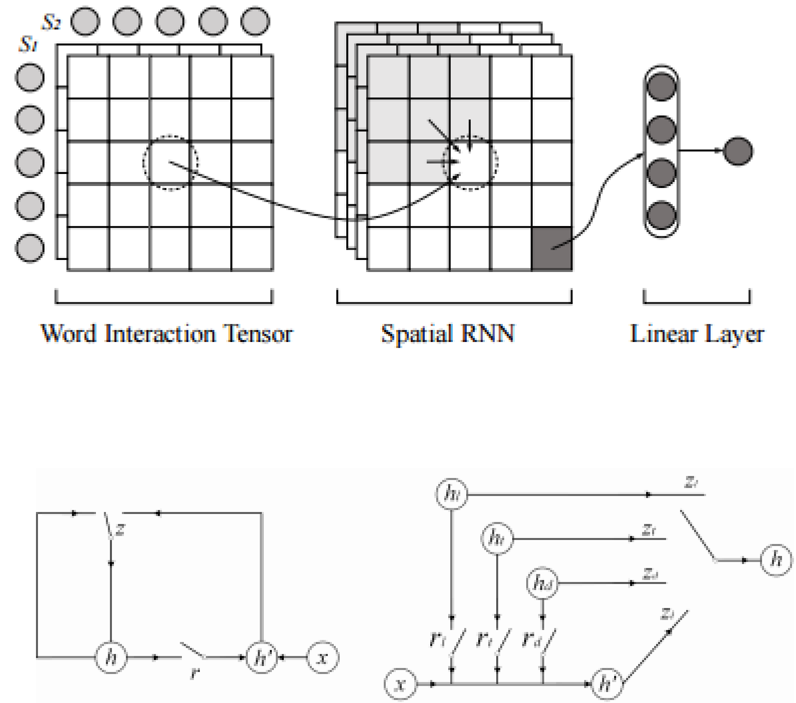
Recursive Matching

- **Fused recursive match**



Liu et al., ACL’16

- **Recursive match**

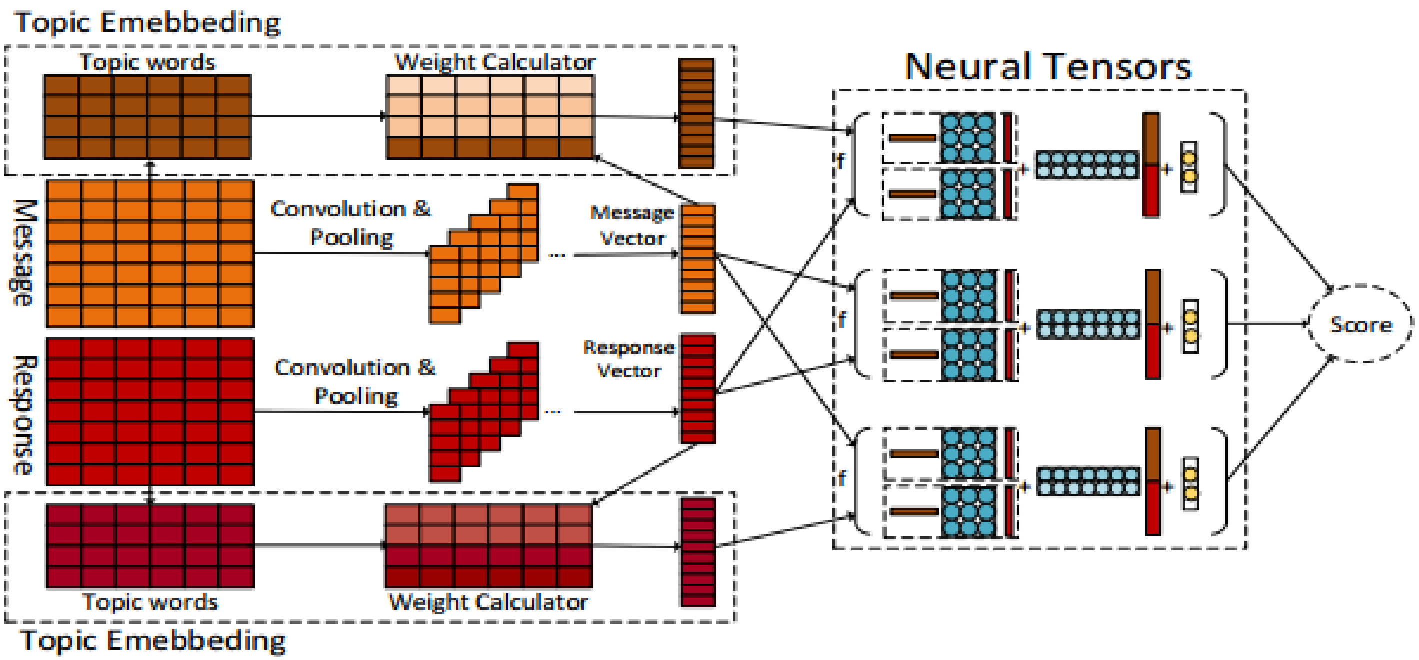


Wan et al., IJCAI’16

Matching with Topic Info

- **Additional info might help!**
  - Topic, knowledge, etc
- **Topic information**
  - Topic word generation: LDA
  - Topic-aware neural network

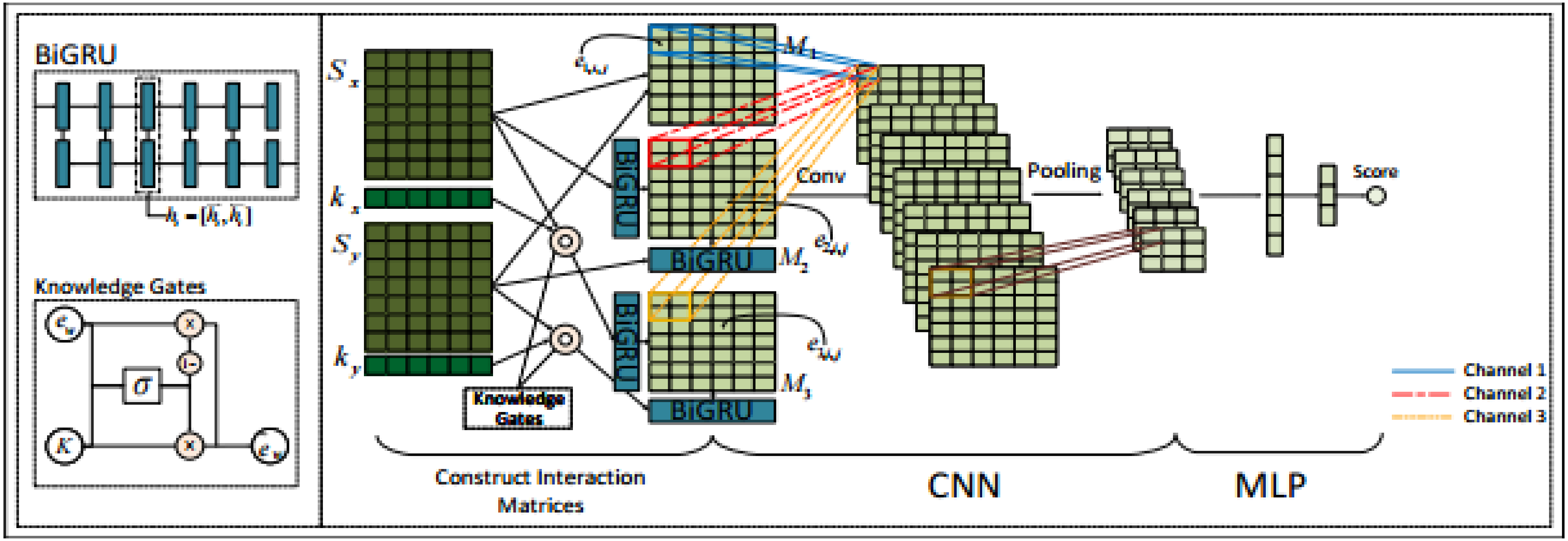
Wu et al., arXiv'16



Matching with Knowledge

- **Here comes the knowledge**
  - Prior knowledge of sentence
    - Tags, keywords, topics, entities, …
  - Fusion of knowledge gate
    - 3 channels: similarity, Bi-GRU match, Bi-GRU with knowledge match

Wu et al., arXiv'16



Multi-Turn Conversation

- **2 typical scenarios for a conversation system**
  - Single-Turn Conversation
  - Multi-Turn Conversation

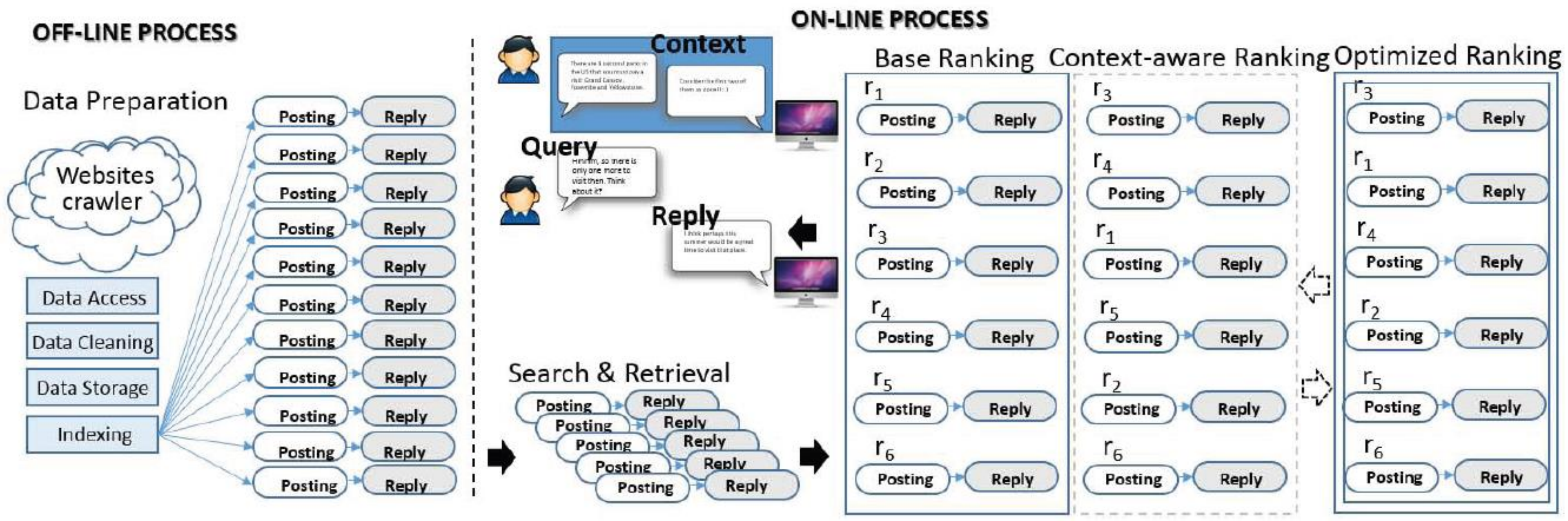
- **Practical Concerns**
  - Effectiveness
  - Efficiency

Yan et al., CIKM'16

Re-ranking Framework

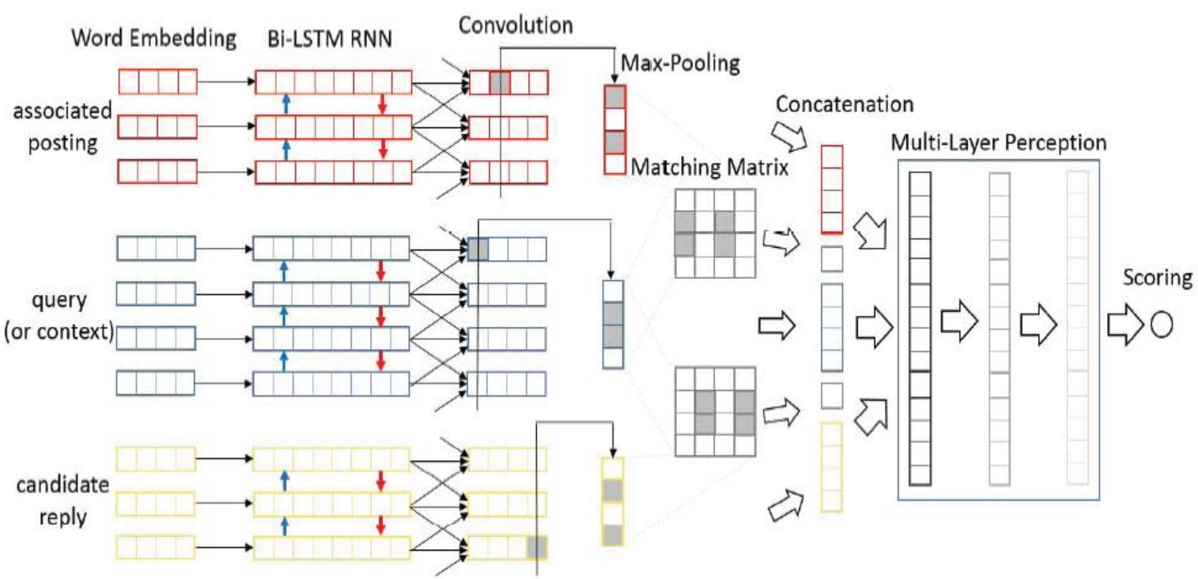
- **Off-line Process**
  - Data preparation: access, cleaning, storage, and indexing
- **Online Process**
  - Search and retrieval
  - Rankings
  - Optimization: rank combination

Yan et al., CIKM'16



Matching

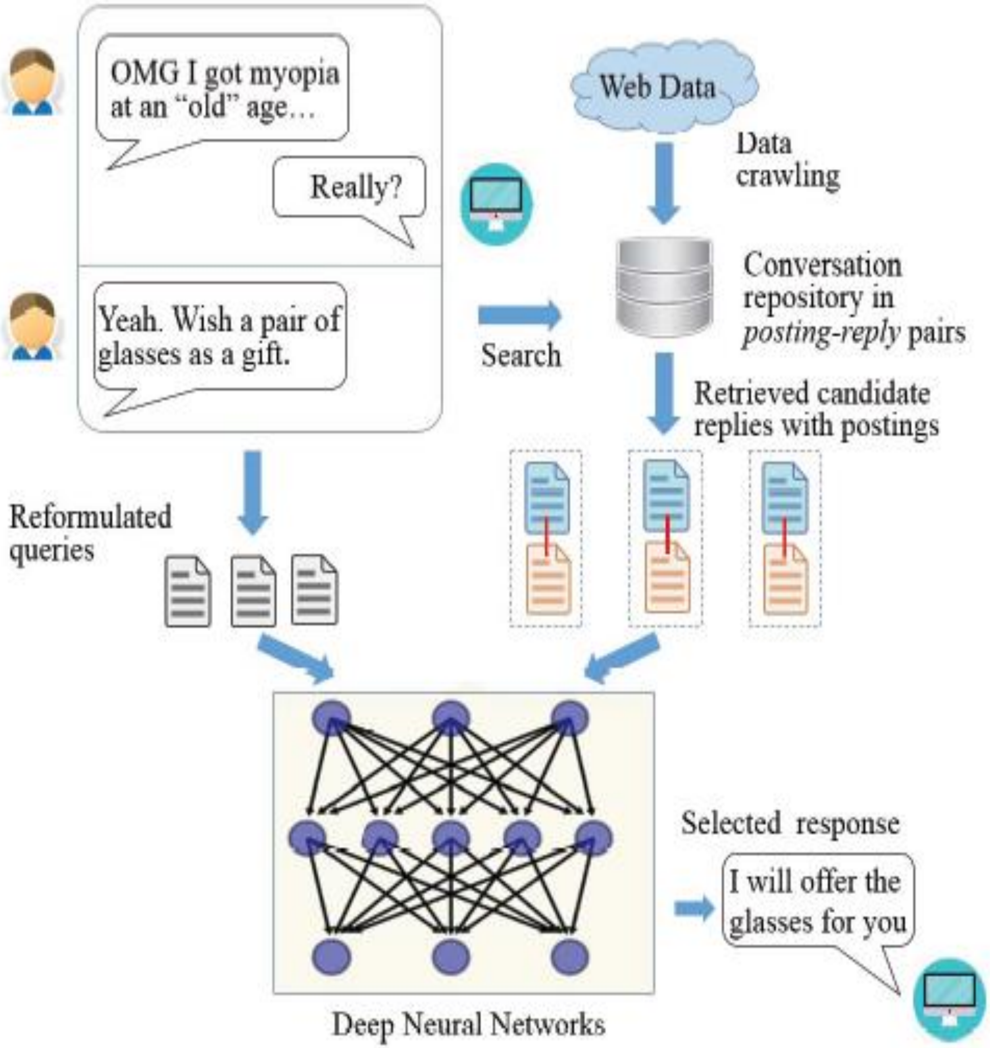
- **Rankers**
  - Shallow Ranker
    - Representations: term-level, topic-level, entity-level
    - Hand-crafted features: matching score (similarity, mutual information), translation probability, language model, term weighting, length, and fluency
  - Deep Ranker
    - Word Embeddings
    - Bi-Directional LSTM
    - Convolution
    - Pooling
    - Concatenation
    - Matching



Yan et al., CIKM'16

Another View

Yan et al., SIGIR'16



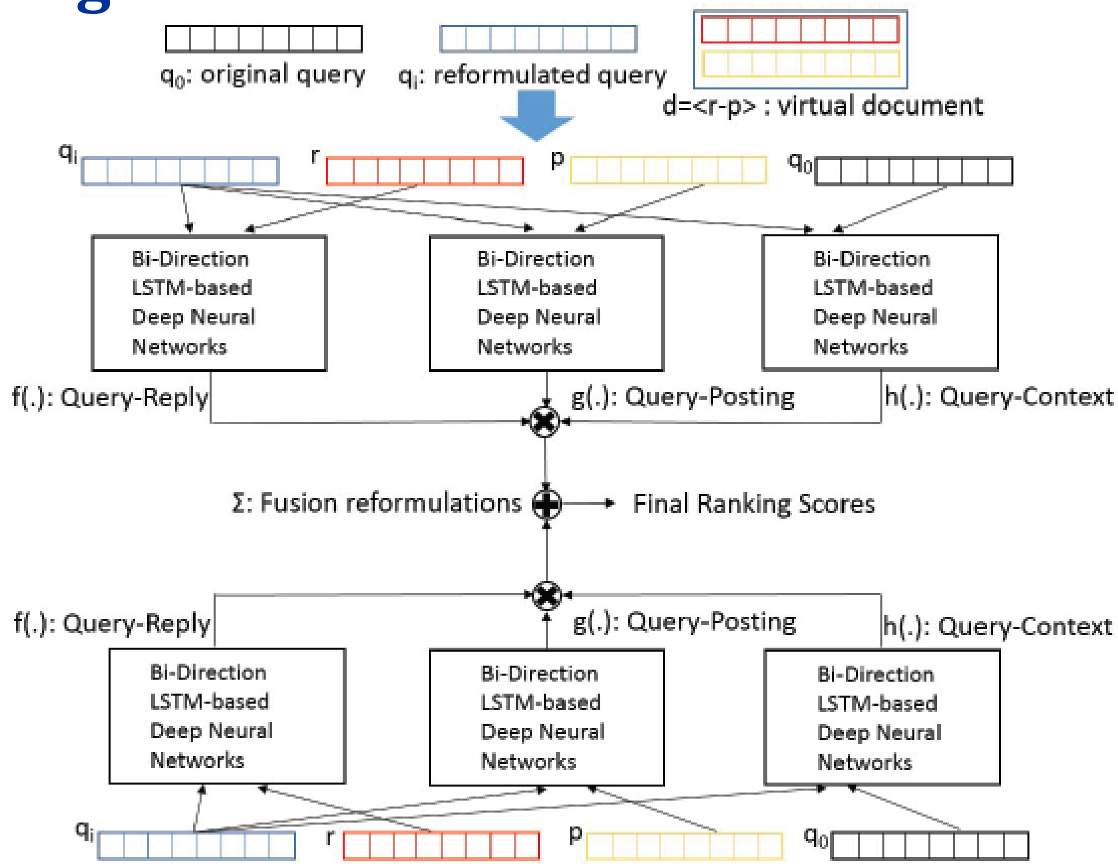
- **Data**
- **Search and retrieval**
- **Contextual reformulation**
  - Possible reformulations

Human-Computer Conversation
A_1: 天哪一把年纪的人居然近视了 (OMG I got myopia at such an "old" age) B_1: 真的吗？ (Really?) A_2:嗯哪。求个眼镜做礼物！ (Yeah. Wish a pair of glasses as a gift.)
B_2:我送你眼镜！ (I will offer the glasses for you!)

Task Formulation
User query: $q_0 = A_2$ **Context information:** $\mathbb{C}=\{c_1=A_1, c_2=B_1\}$ **Reformulated queries:** $q_1=A_2 \boxplus A_1$, $q_2=A_2 \boxplus B_1$ $q_3=A_2 \boxplus A_1 \boxplus B_1, \ldots$
Top-1 ranked response: $r^* = Reply_1$

Learning to Respond

- **Sentence pair matching**
  - f(q,r)
  - g(q,p)
  - h(q,q_0)
- **Representation**
  - Word embedding
  - Bi-Directional LSTM
  - Convolution
  - Pooling
  - Concatenation
  - Matching

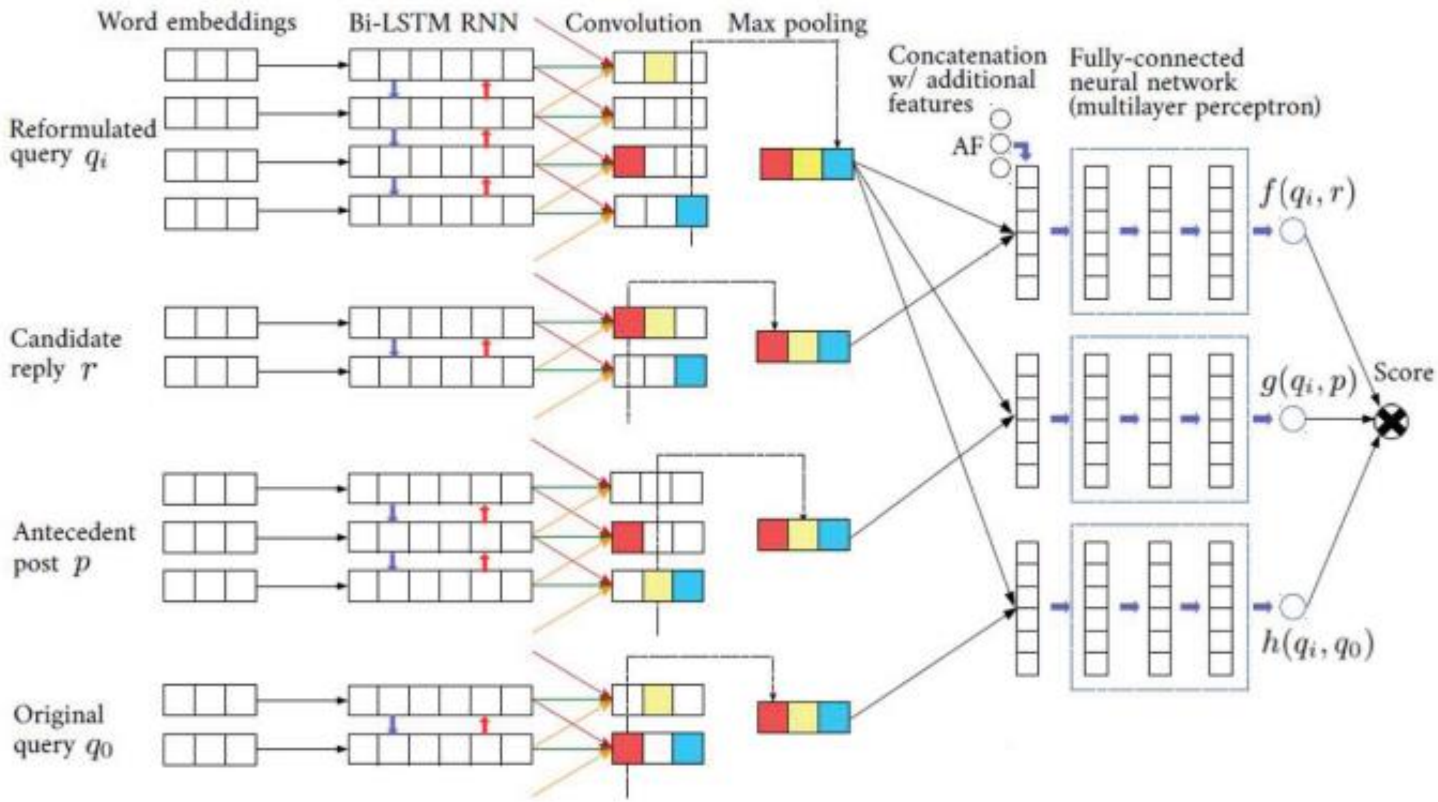


Yan et al., SIGIR'16

Deep Learning to Respond

- **Matching metric**

Yan et al., SIGIR'16



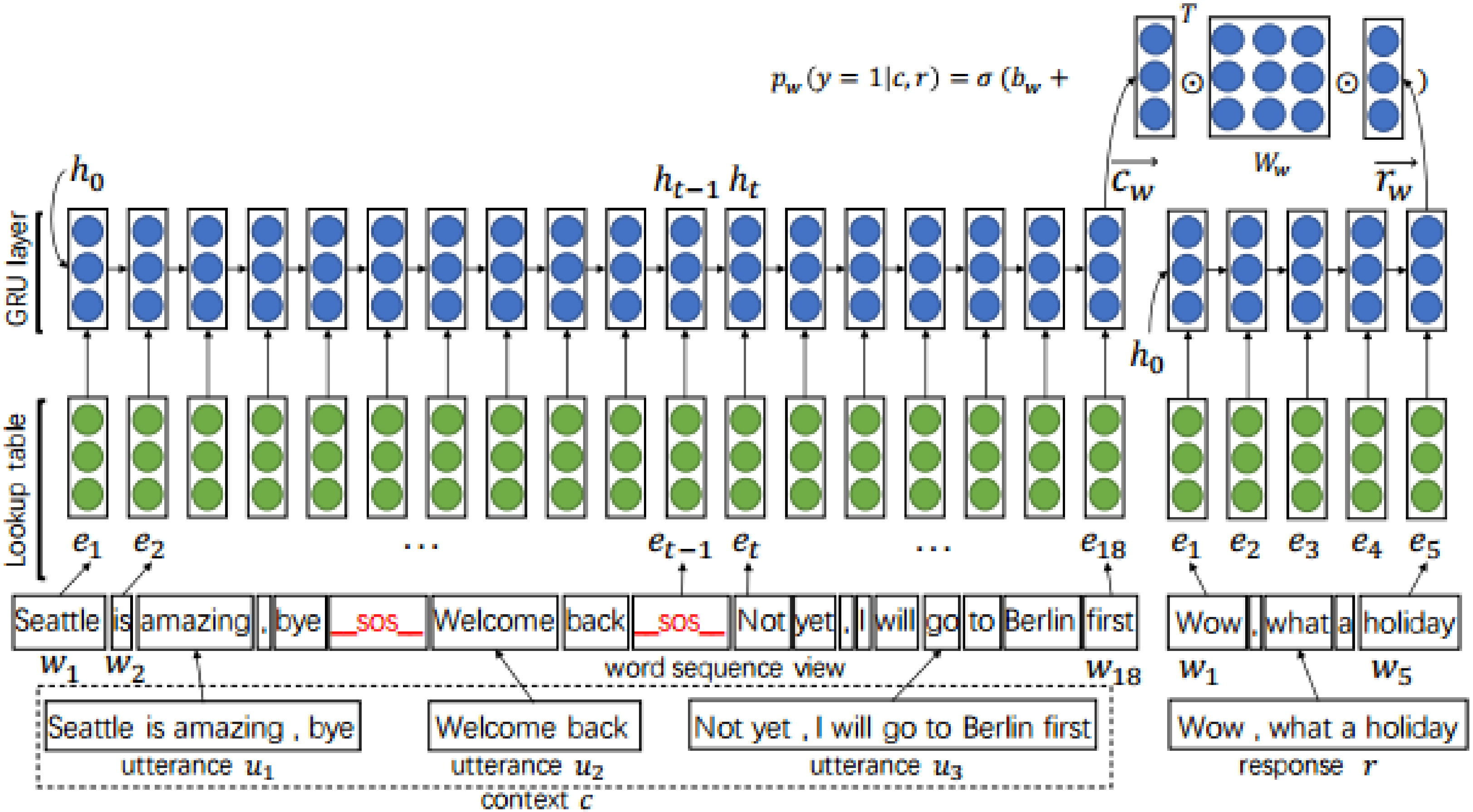
- **Sum-Product Process**

$$\mathcal{F}(q_0, r) = \sum_{i=0}^{|Q|} \left(h(q_0, q_i) \sum_{p} \left(f(q_i, r) \cdot g(q_i, p) \right) \right)$$

Word Sequence Model

- **Response selection**
  - Choose a response given contexts

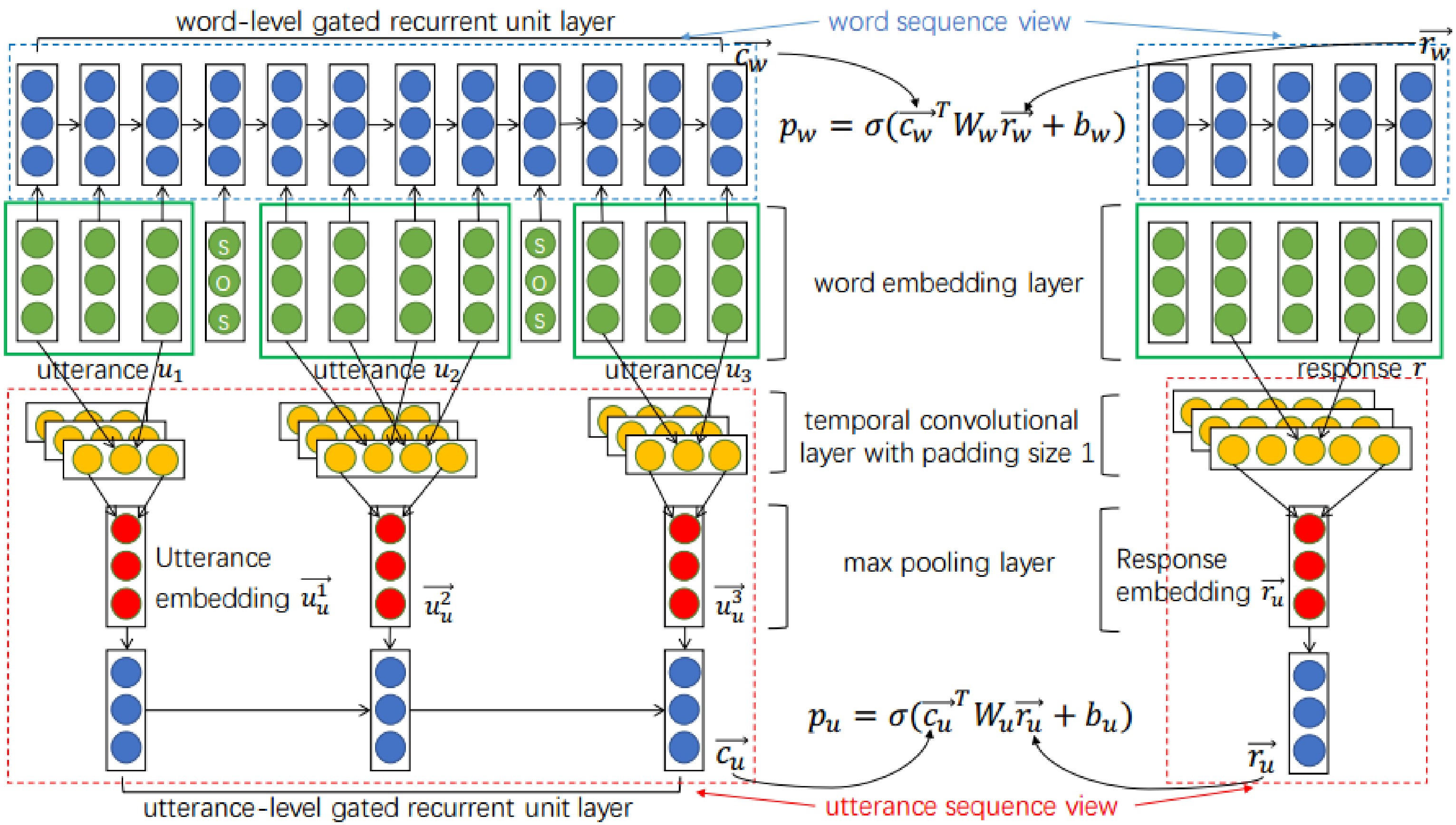
Zhou et al., EMNLP'16



Multi-view Model

- **Views: hierarchical**
  - Word sequence
  - Utterance sequence

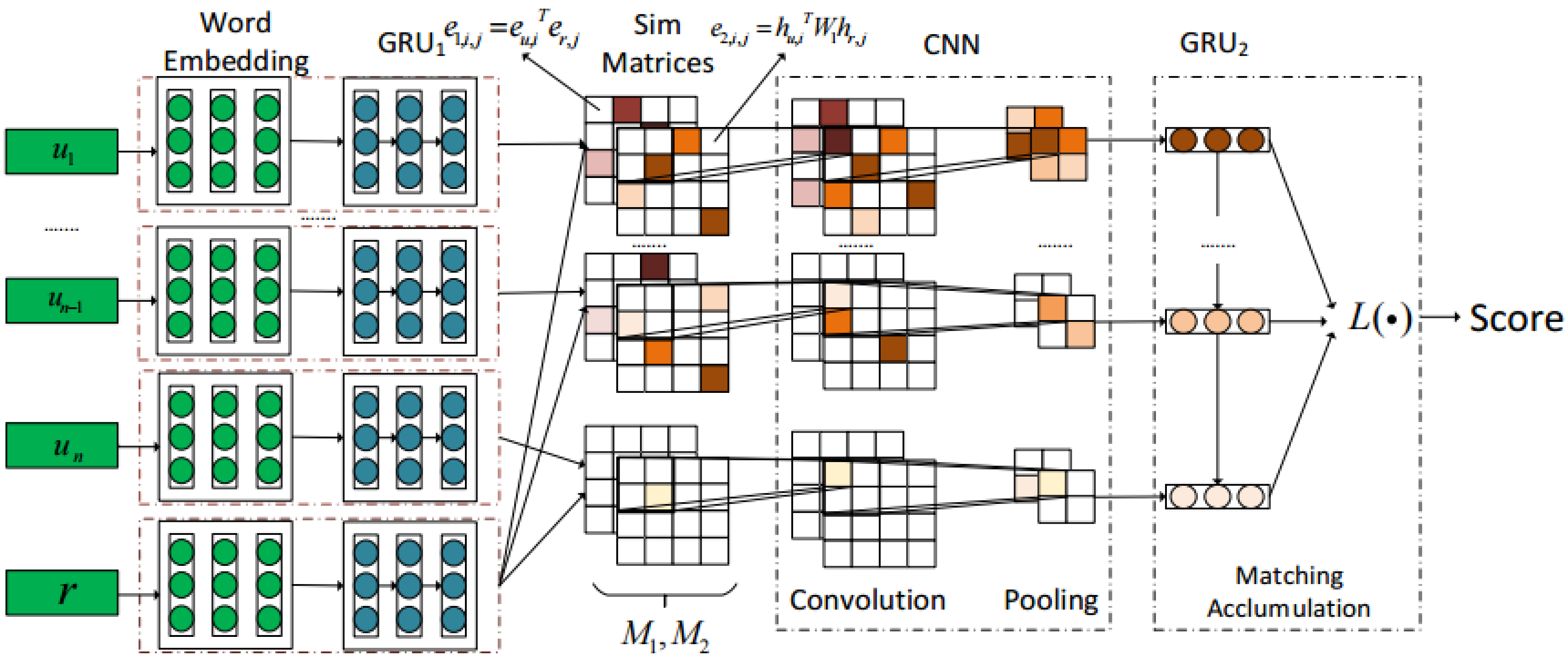
Zhou et al., EMNLP'16



Sequential Match Network

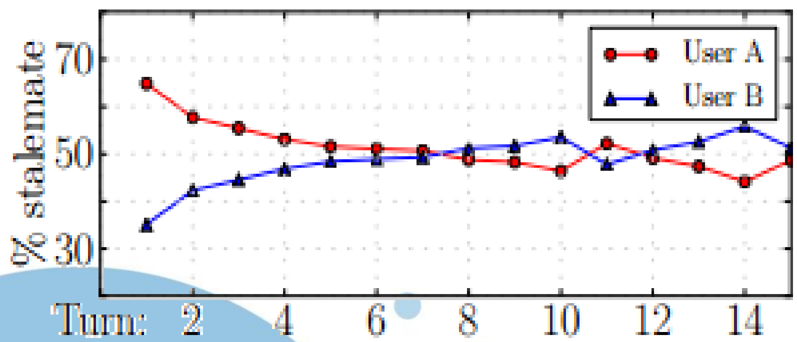
- **Context modeling with sequential utterances**
  - M1: match on the word-level
  - M2: match on the segment-level (based on position)
  - Convolution and pooling
  - Matching sequence

Wu et al., arXiv'16

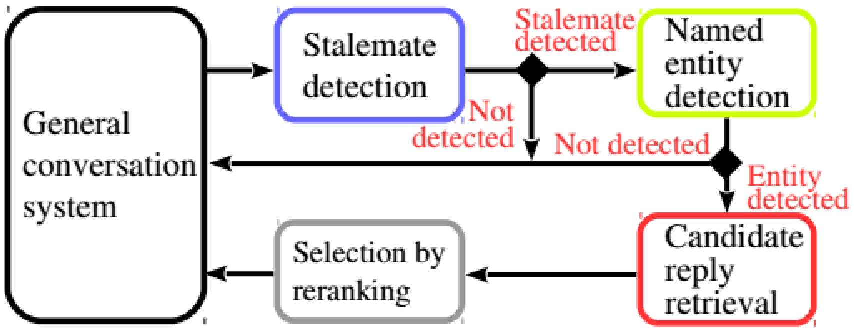


Add-On Component

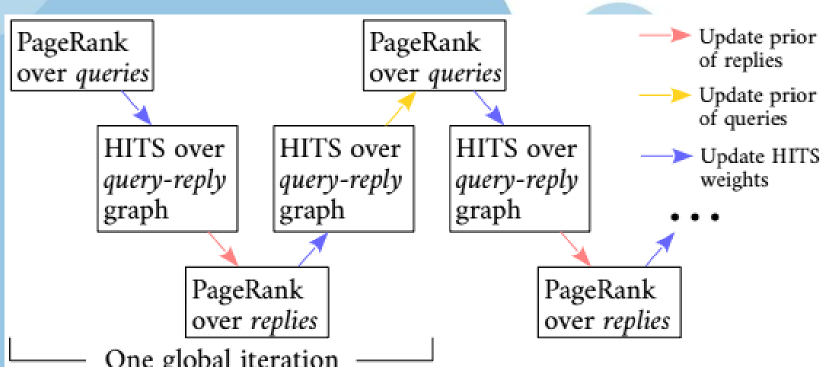
- **StalementBreaker**
- **Human-human conversation**
  - Everyone leads the conversation!

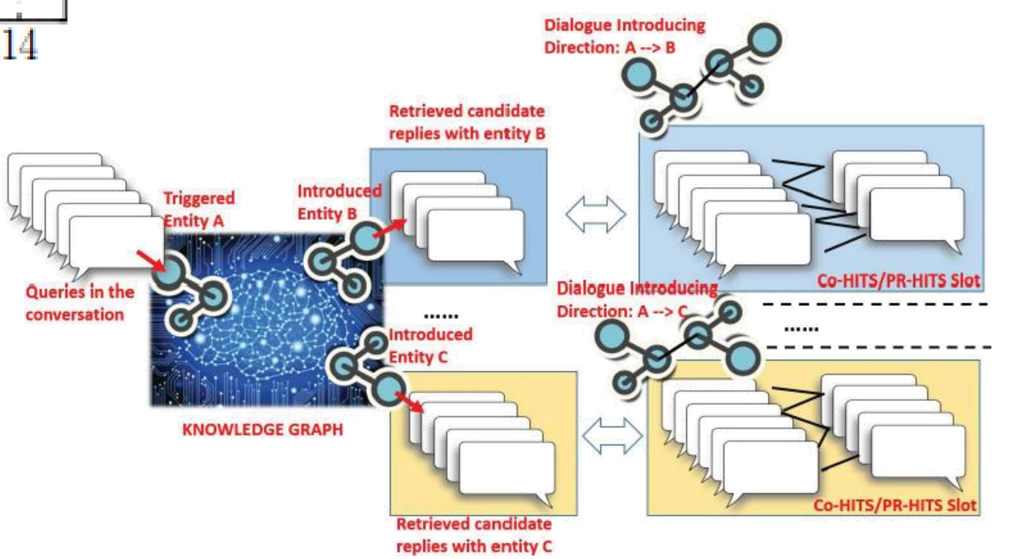


Li et al., IJCAI'16



- **Ranking algorithm**





Gossips and Impacts

- **Pilot study and state-of-the-art**

Li et al., IJCAI'16

- **News media coverage**
  - Been reported within only 3 days after got published on arXiv
  - UK DailyMail, The Stack, Headline Today, China Science, Peking University News and several others

Re-thinking

- Till now we ALREADY have a well-defined paradigm for conversational systems
  - Given a human utterance as a query, the system returns a response
  - The most standard situation: a single q to a single r
  - Some extensions: many q to a single r for multi-turn conversations [Yan et al., SIGIR'16]
- Query suggestions is important in IR

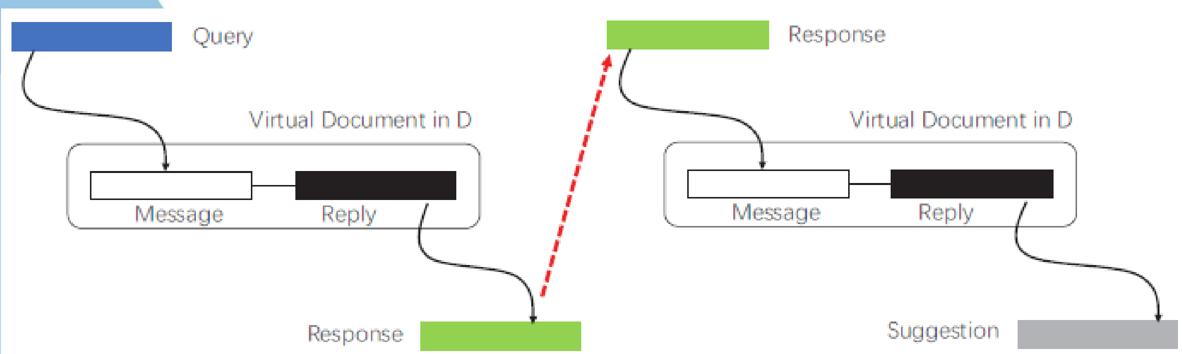
A New Idea

- What if
  - we borrow the query suggestion solution in conversation systems?
  - E.g. “response ranking” and “next utterance suggestion” simultaneously
- Potential benefits
  - From passive conversation mode to proactive conversation mode
  - Brings information outside users’ scope
  - Improve conversational experiences
- Typical situations
  - Predict something that users might say next
  - New contents to talk about: people are open-minded in chit-chat

A New Task

- Problem formulation
  - Given a query q
  - Retrieve a candidate response r
  - Suggest a next utterance s
  - A triple of {q,r,s}
- Given the candidate responses r and suggestions s, we learn to couple them together so as to rank a pair of (r, s) given q.
- A ranking function

$$(r, s)^* = \underset{r,s}{\operatorname{argmax}}\, \mathcal{F}((r, s)|q)$$



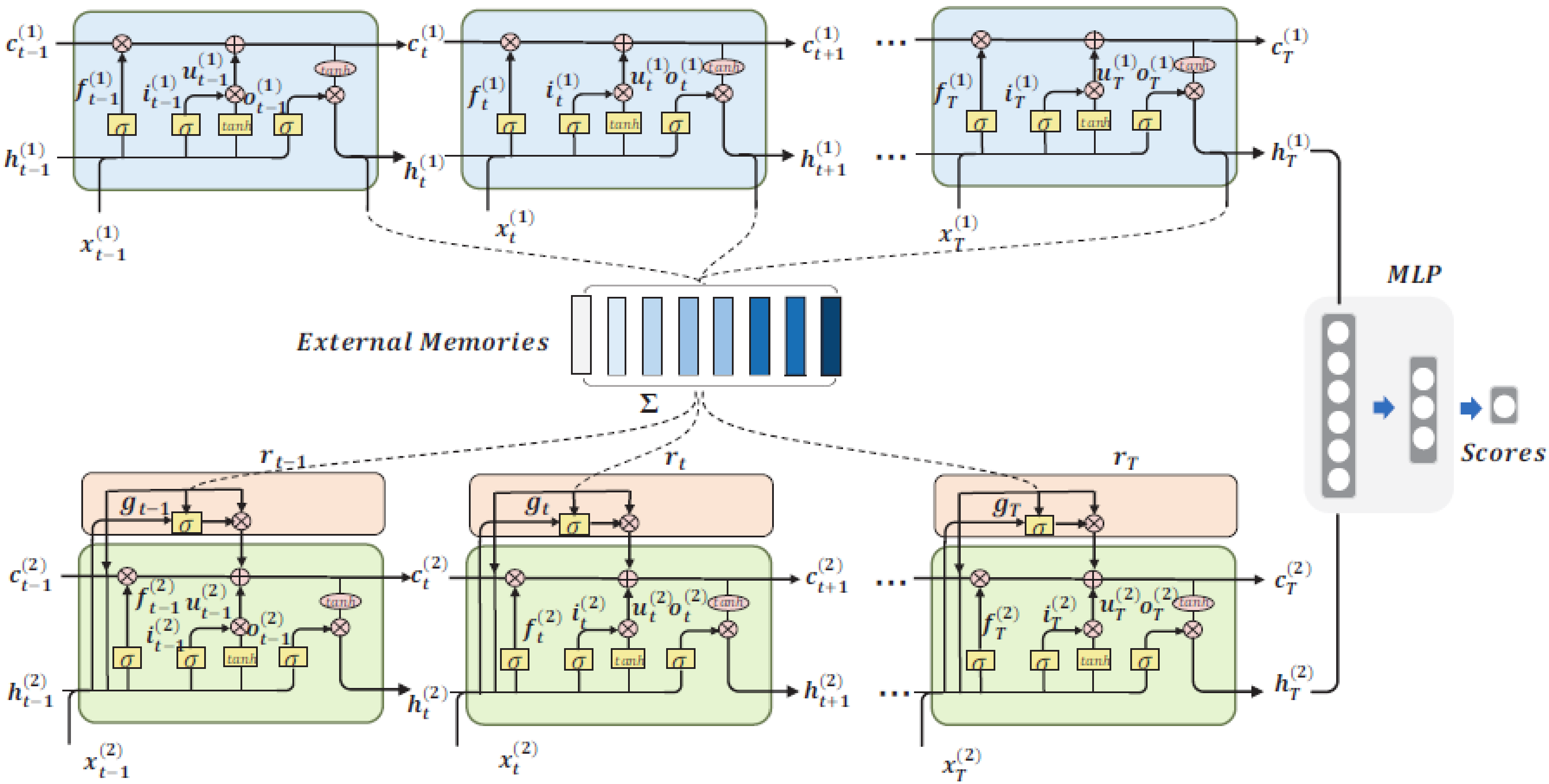
(a) 1st-hop Retrieval

(b) 2nd-hop Retrieval

Dual-LSTM Chain Model

- **Model framework**

Yan et al., SIGIR'17



Results

- **Appropriateness**

Model	p@1	MAP	nDCG	MRR	
Okapi BM25	0.272	0.253	0.302	0.169	I
	0.259	0.226	0.284	0.156	II
	0.138	0.126	0.187	0.091	III
ARC-II	0.394	0.294	0.421	0.232	I
	0.387	0.291	0.415	0.217	II
	0.255	0.201	0.278	0.142	III
LSTM-RNN	0.338	0.283	0.371	0.228	I
	0.351	0.300	0.366	0.237	II
	0.206	0.195	0.233	0.128	III
MV-LSTM	0.435	0.322	0.409	0.308	I
	0.410	0.313	0.414	0.301	II
	0.269	0.251	0.267	0.168	III
Chain-LSTM	0.416	0.328	0.429	0.301	I
	0.422	0.316	0.410	0.307	II
	0.261	0.246	0.298	0.183	III
Dual-LSTM	0.431	0.339	0.441	0.312	I
	0.442	0.326	0.437	0.319	II
	0.425	**0.315**	**0.419**	**0.303**	III

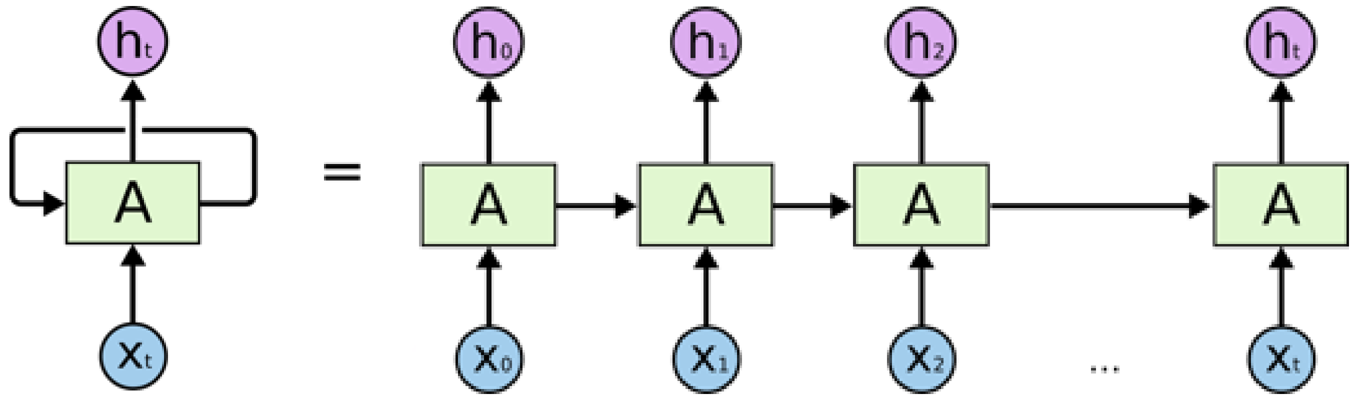
- **Component Evaluations**

Model	p@1	MAP	nDCG	MRR	
-MLP	0.422	0.326	0.432	0.304	I
	0.425	0.320	0.428	0.310	II
	0.271	0.248	0.313	0.216	III
-Cell	0.426	0.333	0.438	0.308	I
	0.430	0.331	0.422	0.317	II
	0.392	0.299	0.401	0.289	III
Dual-LSTM	0.431	0.339	0.441	0.312	I
	0.442	0.326	0.437	0.319	II
	0.425	0.315	0.419	0.303	III

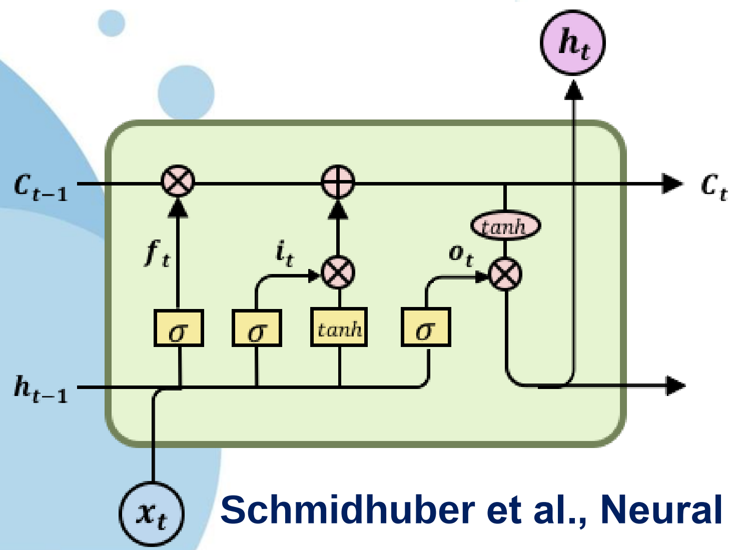
GENERATION-BASED CONVERSATION SYSTEM

RNN Family

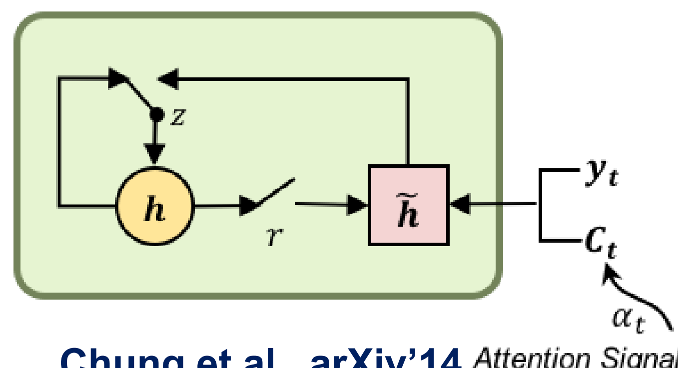
- **Recurrent Neural Networks**
  - Vanilla RNN



  - LSTM
  - GRU



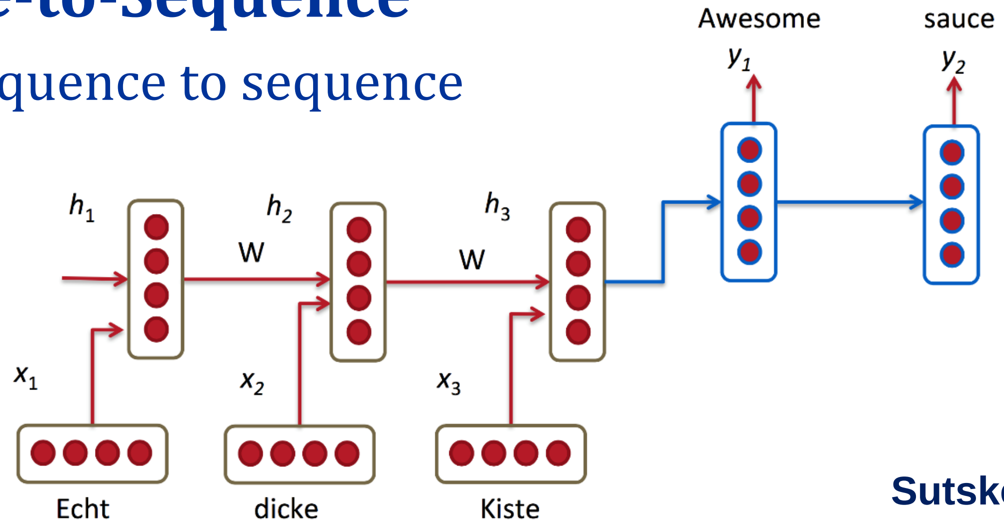
Schmidhuber et al., Neural Computing'97



Chung et al., arXiv'14

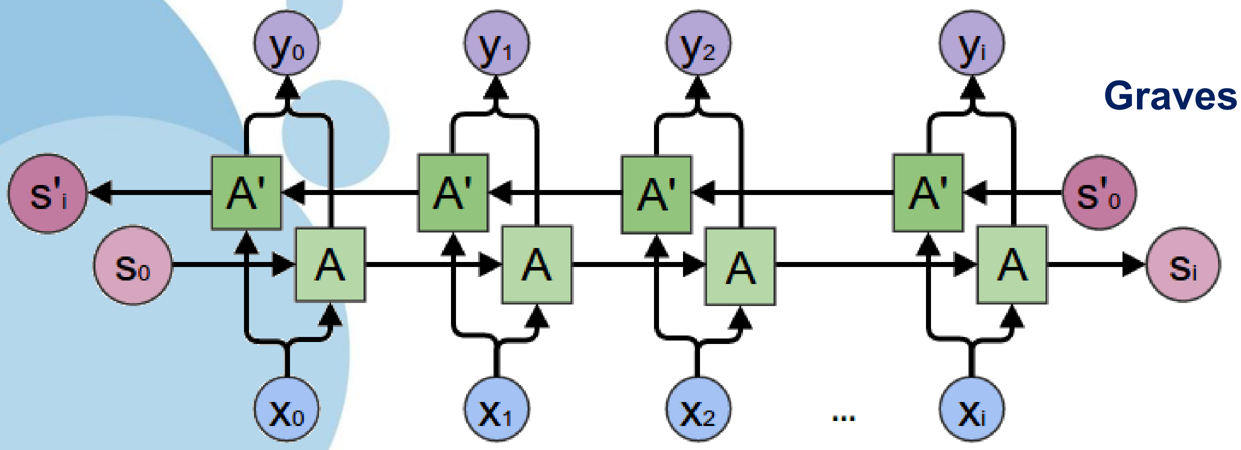
Sequence-to-Sequence

- **Sequence-to-Sequence**
  - Basic sequence to sequence



Sutskever et al., NIPS'14

  - Sequential modeling with bi-directions



Graves et al., ASSP'13

Attention Mechanism

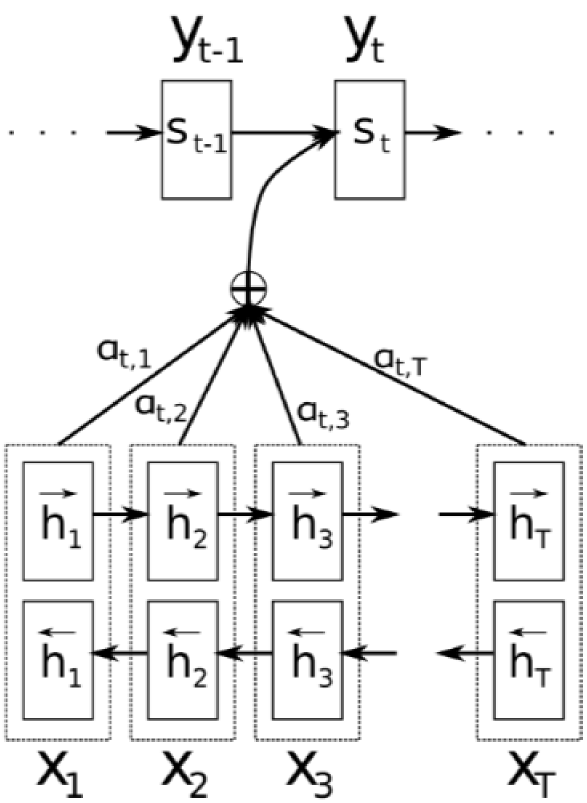
- **Attention signal**

Bahdanau et al., ICLR'14

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

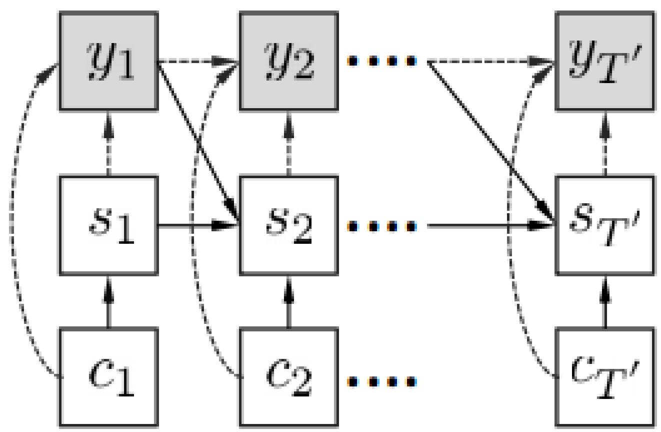
$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

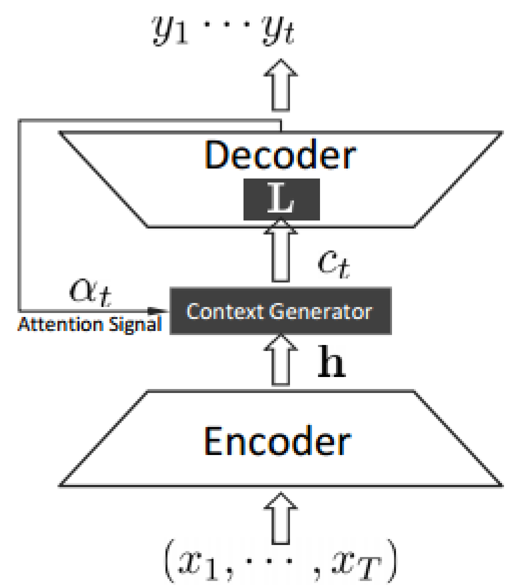
$$e_{ij} = v_a^\top \tanh(W_a s_{i-1} + U_a h_j)$$



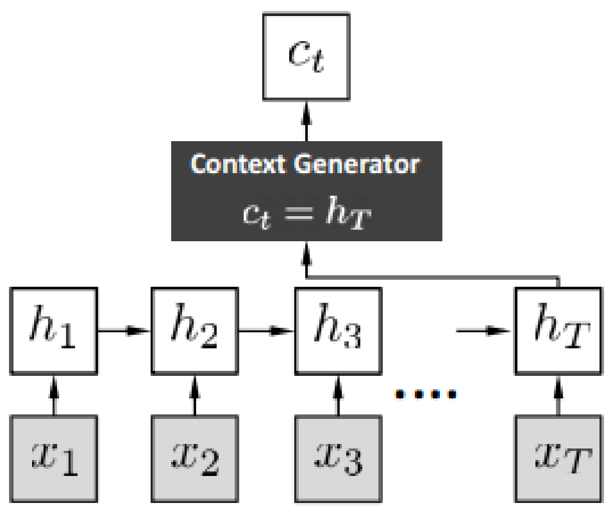
Neural Responding

- **Encoder-Decoder with Attention signal**
  - **Decoder**





  - **Encoder: global**

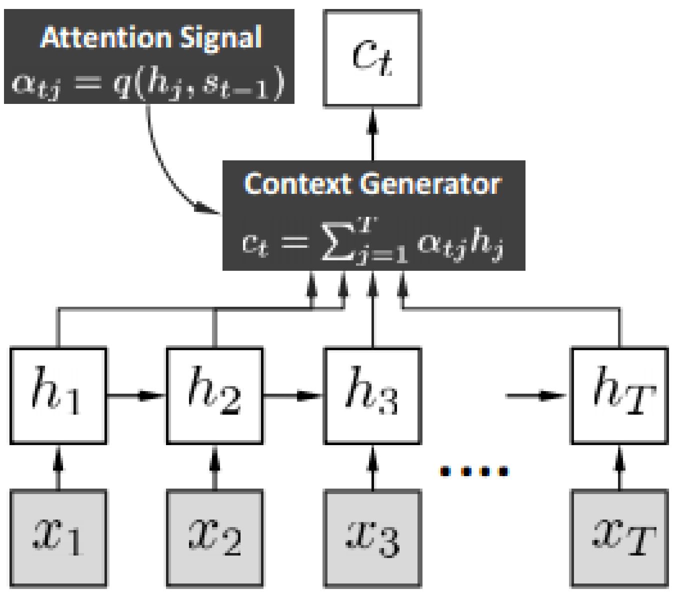


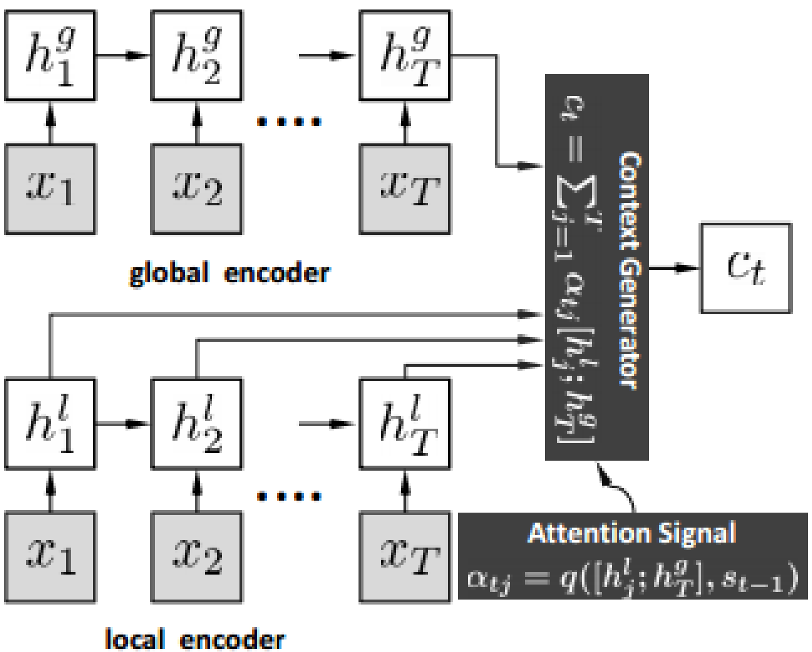
Shang et al., ACL'15

Neural Responding Machine

- **Encoder-Decoder with Attention signal**
  - More encoders: local schema
  - Combinatory schema

Shang et al., ACL'15

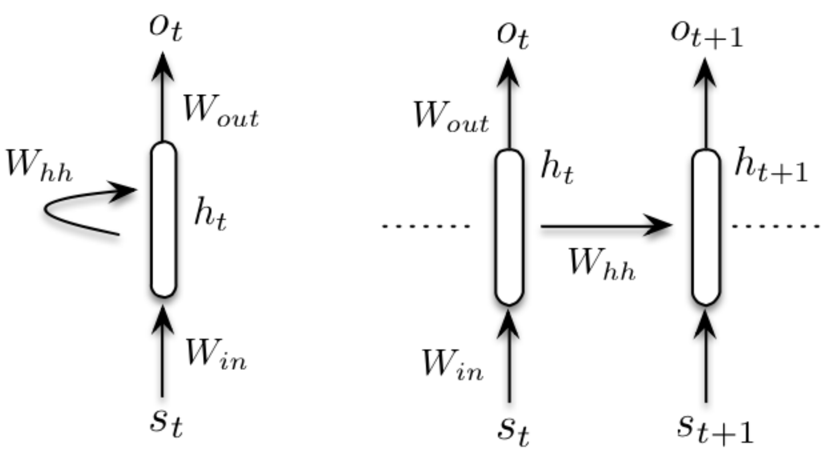


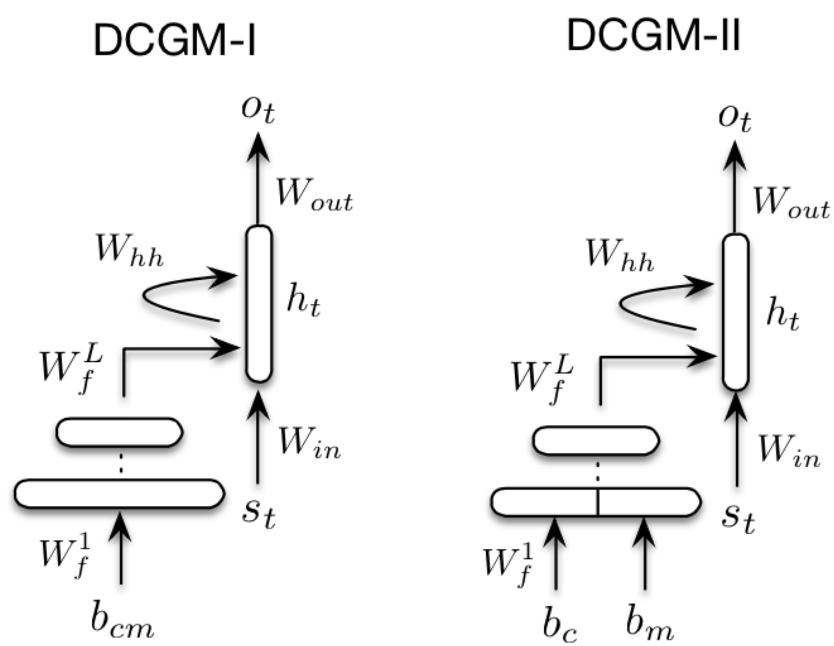


Context-Sensitive Generation

- **Encoder-Decoder with Contextual information**
  - Concatenate each utterance c, m, r into a single sentence s
- **Strengthening the context bias**
  - Bag-of-words
  - Concatenation

Sordoni et al., NAACL-HLT'15

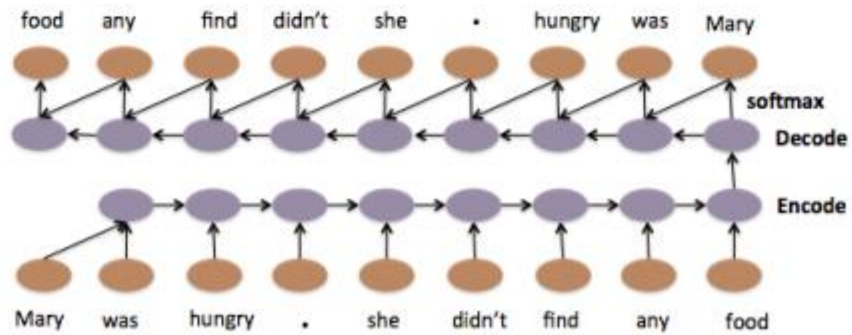


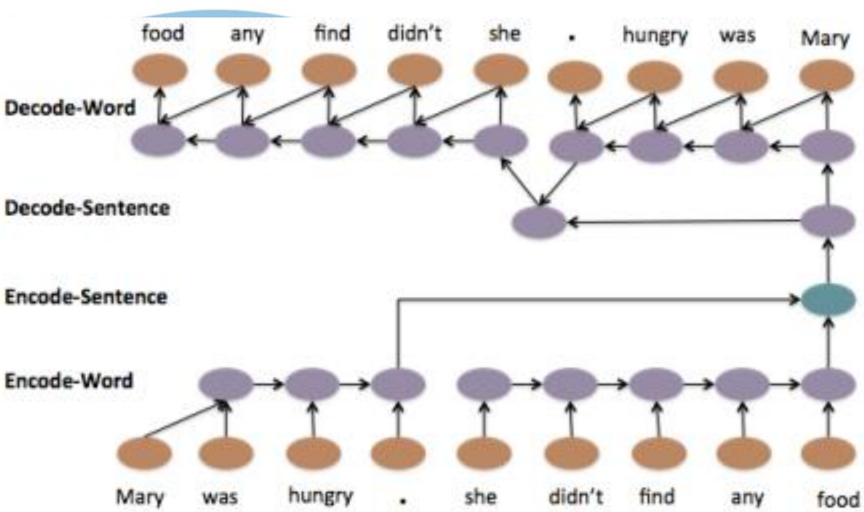


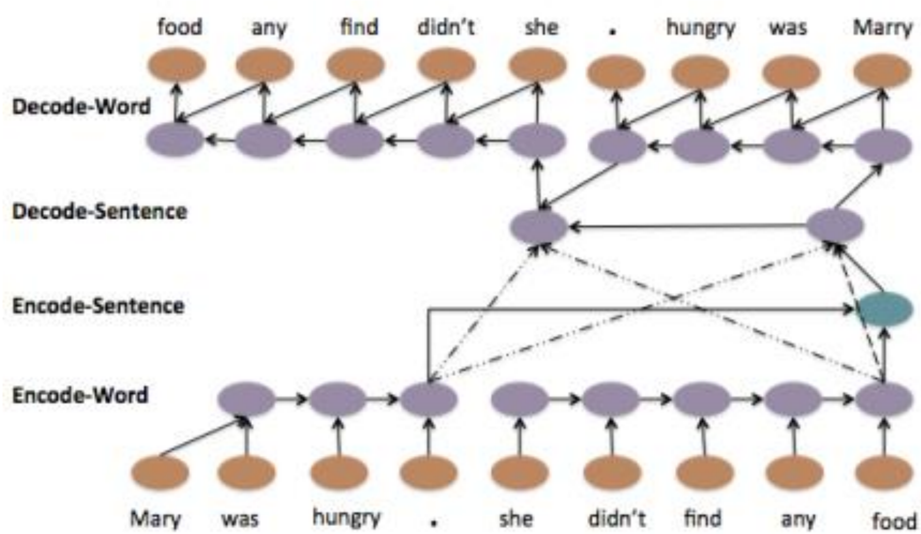
Hierarchical Language Model

- **Hierarchy**
  - Word level
  - Sentence level
- **Auto-Encoder**

Li et al., ACL’15



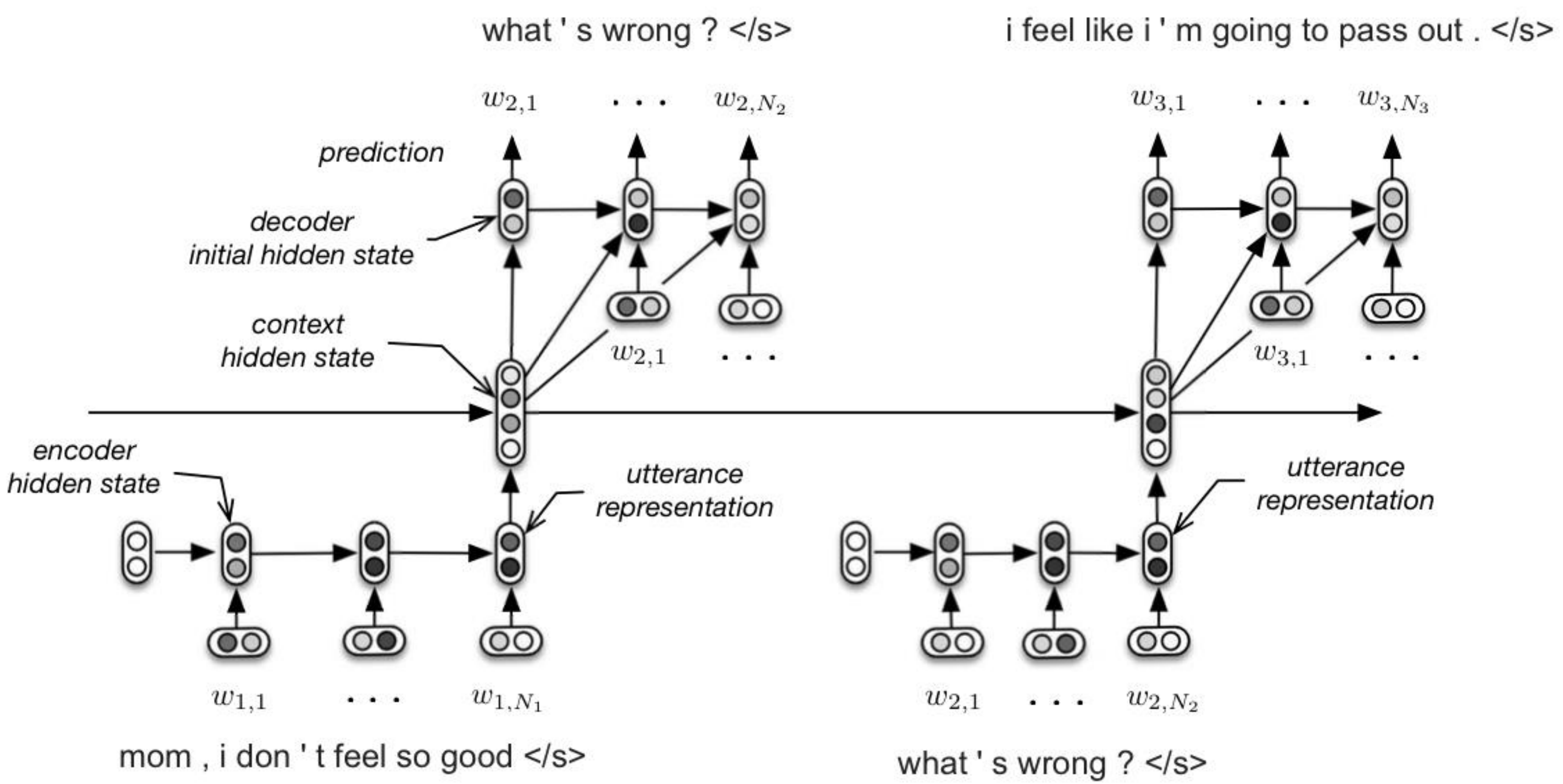




Hierarchical Encoder-decoder

- **HRED (hierarchical recurrent encoder decoder)**
  - Hierarchical architecture (two level)
    - a sequence of words for each utterance
    - a sequence of utterances

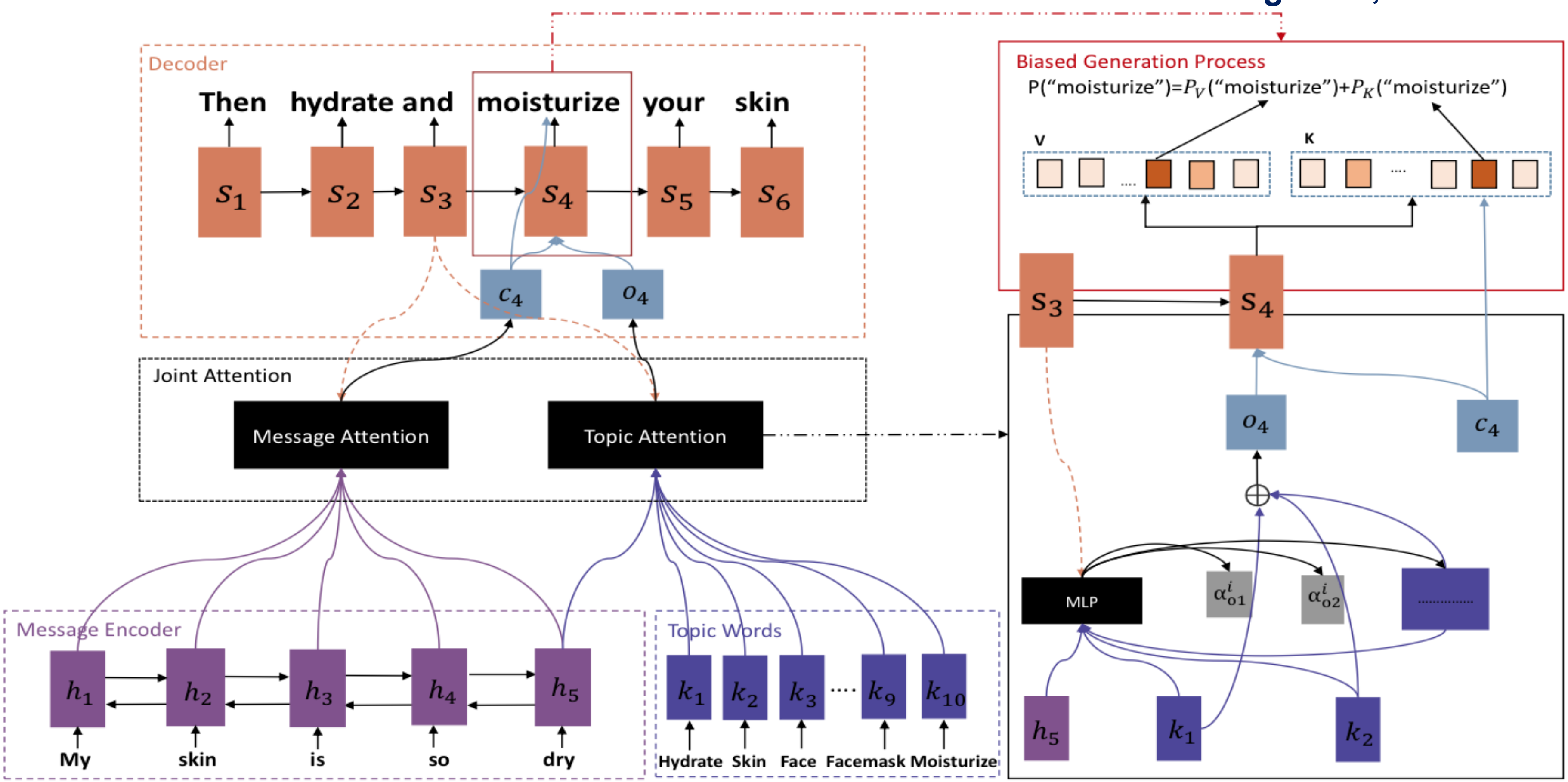
Serban et al., AAAI'16



Topic-Aware Generation

- **TA-Seq2Seq (Topic Aware Seq2Seq)**
  - Topic attention obtained from a pre-trained LDA model

Xing et al., arXiv'16

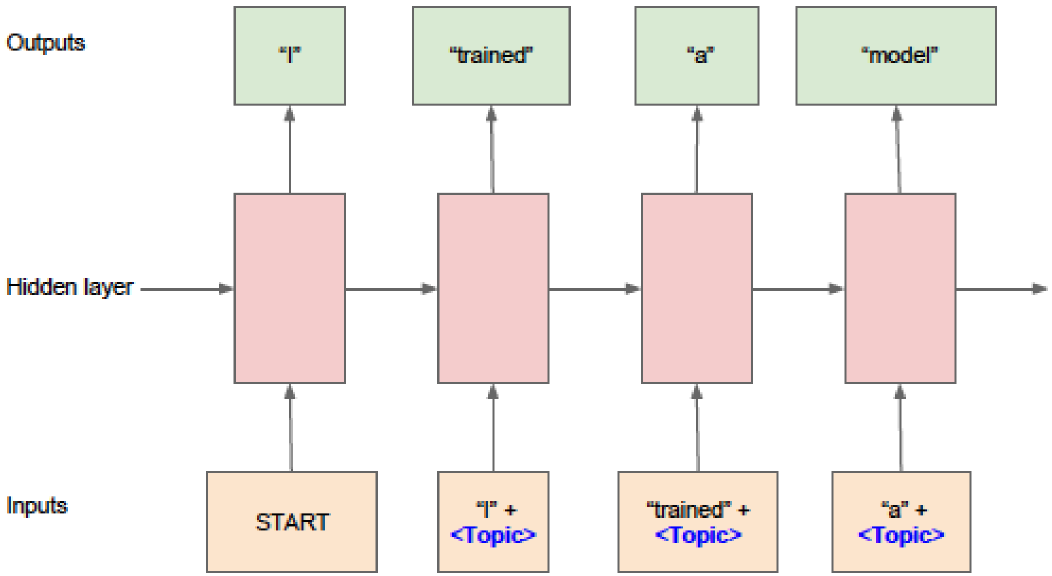


Contextual LSTM

- **Add the topic vector T**

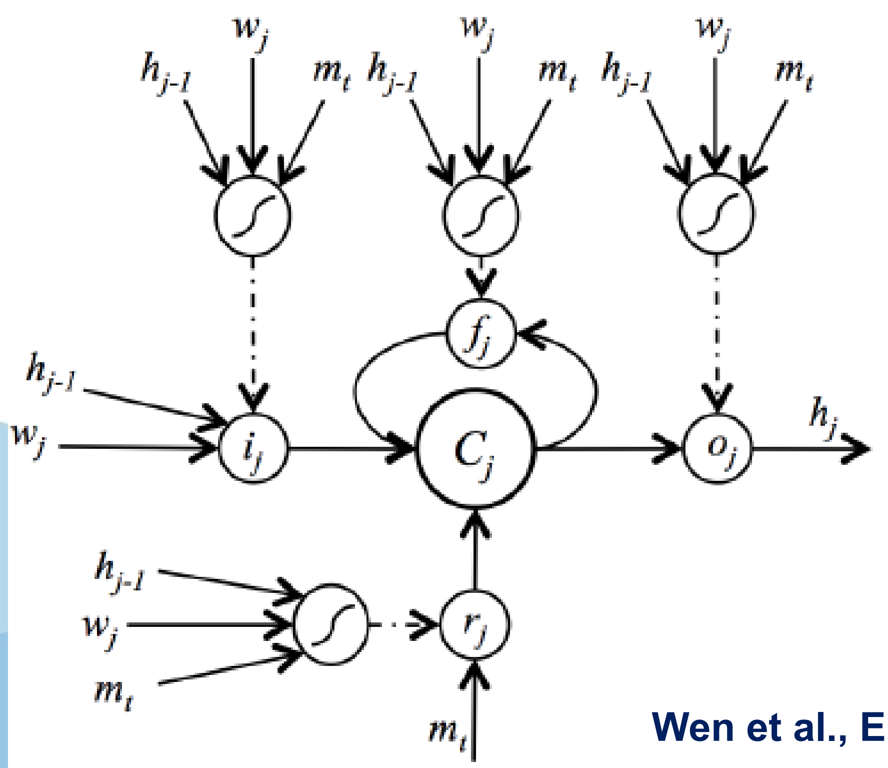
Ghosh et al., KDD'16 Workshop

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i + \mathbf{W}_{\mathbf{Ti}}\mathbf{T}) \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f + \mathbf{W}_{\mathbf{Ti}}\mathbf{T}) \\
c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c + \mathbf{W}_{\mathbf{Ti}}\mathbf{T}) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o + \mathbf{W}_{\mathbf{Ti}}\mathbf{T}) \\
h_t &= o_t \tanh(c_t)
\end{aligned}
$$



Conditional Generation Network

- **Memory type LSTM**



$$\begin{pmatrix} \mathbf{i}_j \\ \mathbf{f}_j \\ \mathbf{o}_j \\ \mathbf{r}_j \end{pmatrix} = \begin{pmatrix} \text{sigmoid} \\ \text{sigmoid} \\ \text{sigmoid} \\ \text{sigmoid} \end{pmatrix} \mathbf{W}_{4n,3n} \begin{pmatrix} \mathbf{m}_t \\ \mathbf{w}_j \\ \mathbf{h}_{j-1} \end{pmatrix}$$

$$\hat{\mathbf{c}}_j = \tanh\left(\mathbf{W}_c(\mathbf{w}_j \oplus \mathbf{h}_{j-1})\right)$$

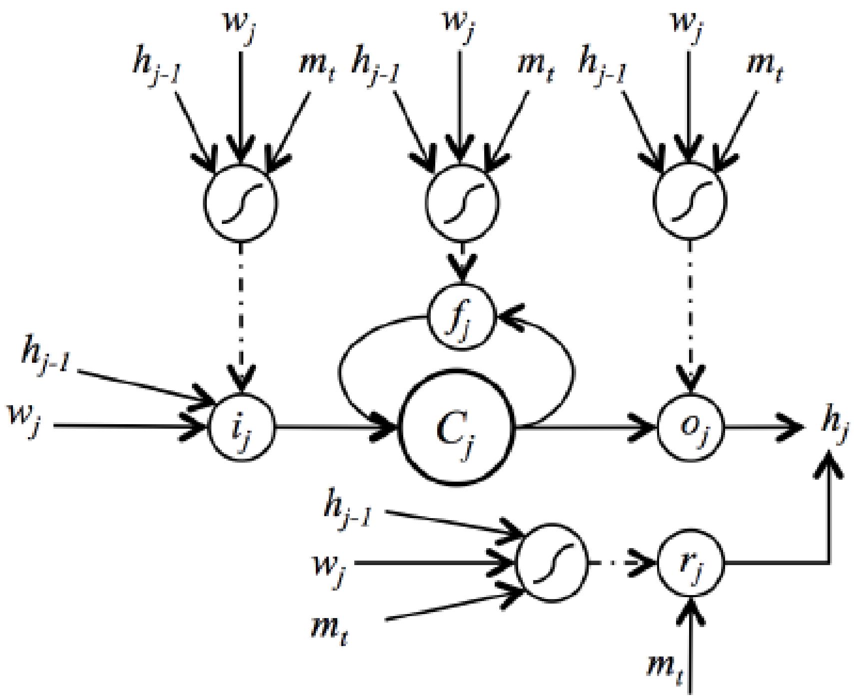
$$\mathbf{c}_j = \mathbf{f}_j \odot \mathbf{c}_{j-1} + \mathbf{i}_j \odot \hat{\mathbf{c}}_j + \mathbf{r}_j \odot \mathbf{m}_t$$

$$\mathbf{h}_j = \mathbf{o}_j \odot \tanh(\mathbf{c}_j)$$

Wen et al., EMNLP'16

Conditional Generation Network

- **Hybrid type LSTM**



$$\begin{pmatrix} \mathbf{i}_j \\ \mathbf{f}_j \\ \mathbf{o}_j \\ \mathbf{r}_j \end{pmatrix} = \begin{pmatrix} \text{sigmoid} \\ \text{sigmoid} \\ \text{sigmoid} \\ \text{sigmoid} \end{pmatrix} \mathbf{W}_{4n,3n} \begin{pmatrix} \mathbf{m}_t \\ \mathbf{w}_j \\ \mathbf{h}_{j-1} \end{pmatrix}$$

$$\hat{\mathbf{c}}_j = \tanh\left(\mathbf{W}_c(\mathbf{w}_j \oplus \mathbf{h}_{j-1})\right)$$

$$\mathbf{c}_j = \mathbf{f}_j \odot \mathbf{c}_{j-1} + \mathbf{i}_j \odot \hat{\mathbf{c}}_j$$

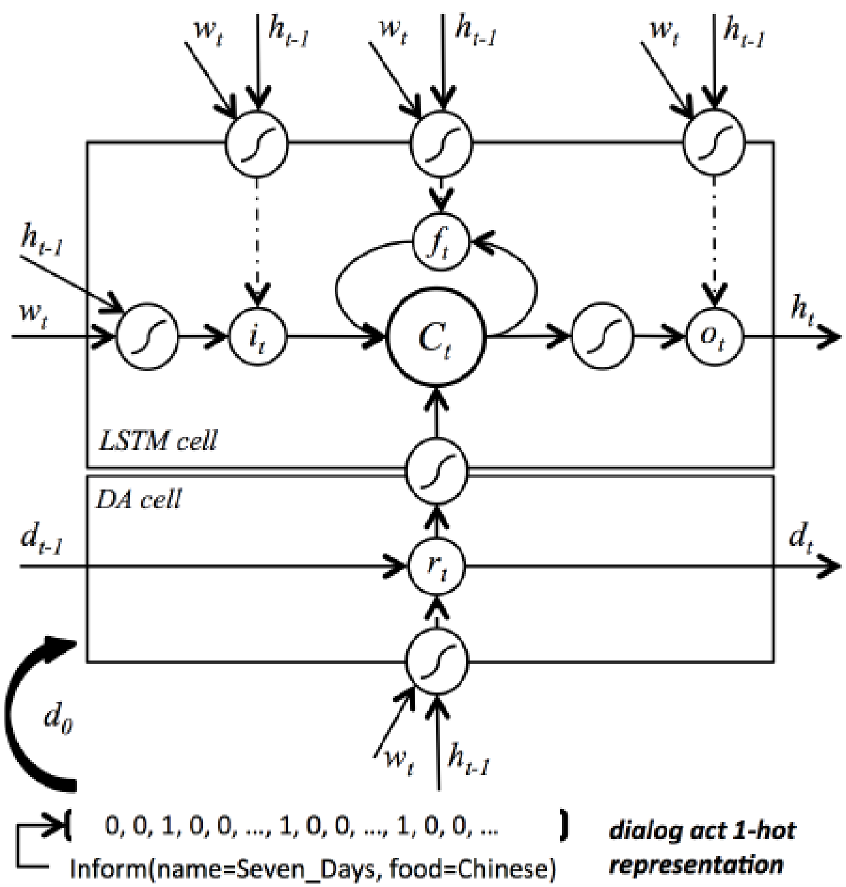
$$\mathbf{h}_j = \mathbf{o}_j \odot \tanh(\mathbf{c}_j) + \mathbf{r}_j \odot \mathbf{m}_t$$

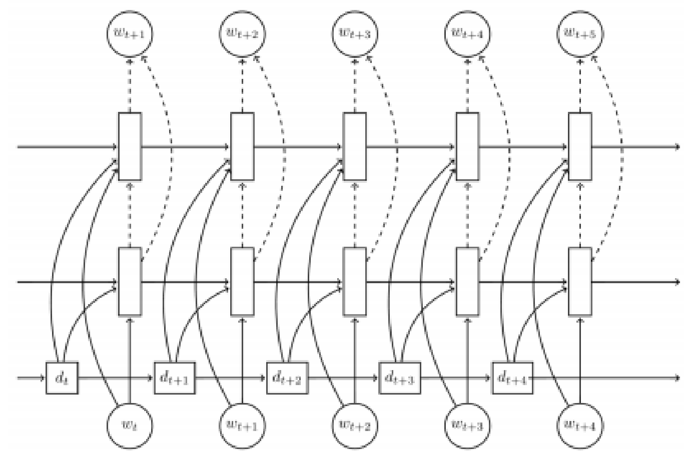
Wen et al., EMNLP'16

Semantically Conditioned LSTM

- **Semantic Controlled LSTM**

Wen et al., EMNLP'15





$$\mathbf{i}_t = \sigma(\mathbf{W}_{wi}\mathbf{w}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1})$$
$$\mathbf{f}_t = \sigma(\mathbf{W}_{wf}\mathbf{w}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1})$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_{wo}\mathbf{w}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1})$$
$$\hat{\mathbf{c}}_t = tanh(\mathbf{W}_{wc}\mathbf{w}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1})$$
$$\mathbf{r}_t = \sigma(\mathbf{W}_{wr}\mathbf{w}_t + \alpha\mathbf{W}_{hr}\mathbf{h}_{t-1})$$
$$\mathbf{d}_t = \mathbf{r}_t \odot \mathbf{d}_{t-1}$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \hat{\mathbf{c}}_t + tanh(\mathbf{W}_{dc}\mathbf{d}_t)$$
$$\mathbf{h}_t = \mathbf{o}_t \odot tanh(\mathbf{c}_t)$$

Generation Overview

- **Case studies**

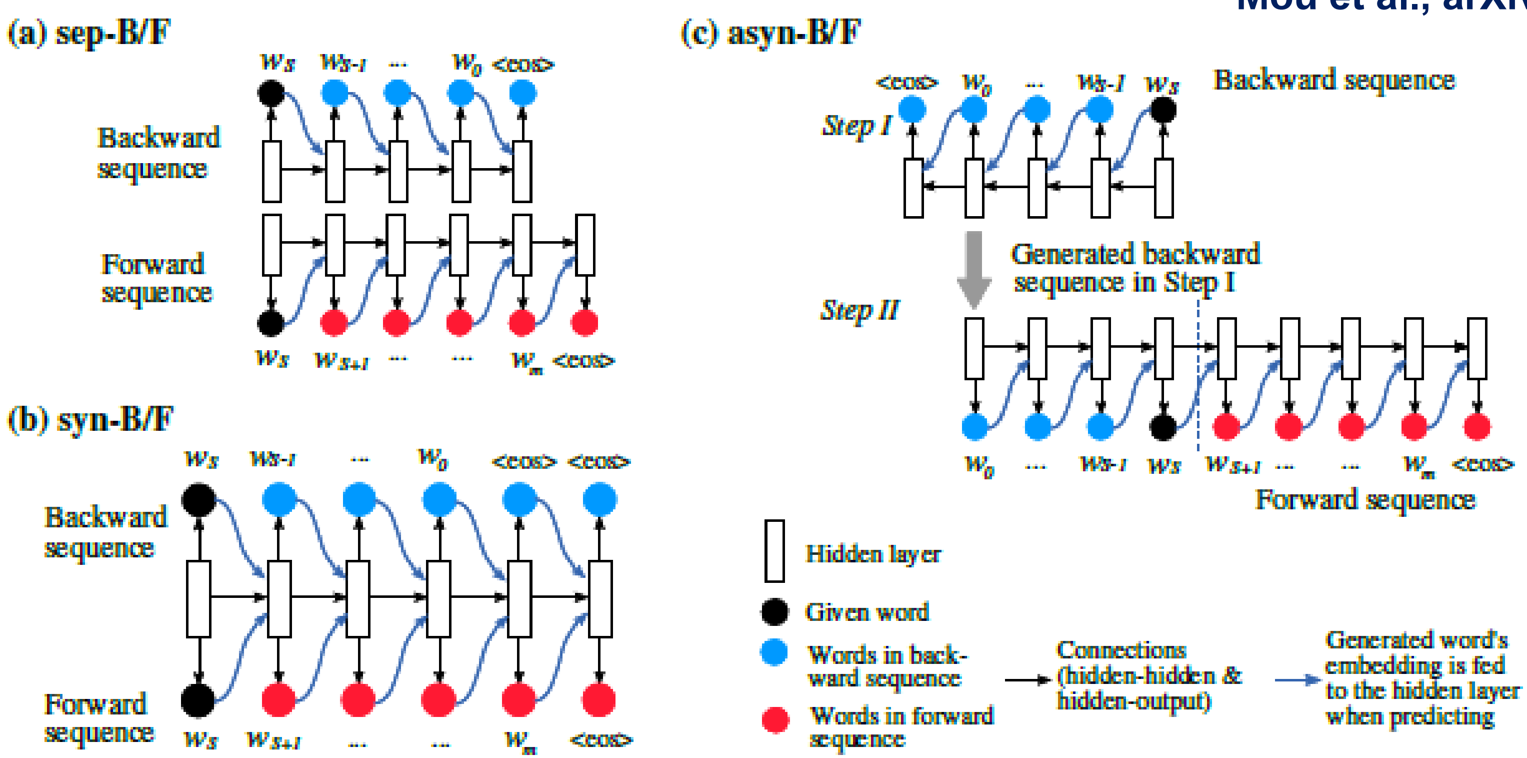
Wen et al., EMNLP'15

#	Example Dialogue Acts and Realizations from SF Restaurant Domain
1	inform(name="red door cafe", goodformeal="breakfast", area="cathedral hill", kidsallowed="no") *red door cafe is a good restaurant for breakfast in the area of cathedral hill and does not allow children .* *red door cafe is a good restaurant for breakfast in the cathedral hill area and does not allow children .* *red door cafe is a good restaurant for breakfast in the cathedral hill area and does not allow kids .* *red door cafe is good for breakfast and is in the area of cathedral hill and does not allow children .* *red door cafe does not allow kids and is in the cathedral hill area and is good for breakfast .*
2	informonly(name="dosa on fillmore and kiss seafood", pricerange="expensive", near="lower pacific heights") *there is no place other than dosa on fillmore and kiss seafood that are expensive near to lower pacific heights .* *dosa on fillmore and kiss seafood is the only expensive restaurant near lower pacific heights .* *the only listed restaurant near lower pacific heights in the expensive price range is dosa on fillmore and kiss seafood .* *i apologize , dosa on fillmore and kiss seafood is the only expensive restaurant near lower pacific heights .* *i apologize , dosa on fillmore and kiss seafood are the only expensive restaurants near lower pacific heights .*
#	**Example Dialogue Acts and Realizations from SF Hotel Domain**
3	inform(type="hotel",count="182",dogsallowed="dontcare") *there are 182 hotels if you do not care whether dogs are allowed .* *there are 182 hotels if you do not care whether they allow dogs .* *182 hotels are available if dogs allowed or not is not an issue .* *there are 182 hotels if allowing dogs or not is not an issue .* *there are 182 hotels if whether dogs are allowed does not matter .*
4	informonly(name="red victorian bed breakfast",acceptscreditcards="yes",near="haight",hasinternet="yes") *red victorian bed breakfast is the only hotel near haight and accepts credit cards and has internet .* *red victorian bed breakfast is the only hotel near haight and has internet and accepts credit cards .* *red victorian bed breakfast is the only hotel near haight that accept credit cards and offers internet .* *the red victorian bed breakfast has internet and near haight , it does accept credit cards .* *the red victorian bed breakfast is the only hotel near haight that accepts credit cards , and offers internet .*

Language Generation

- **Constrained language generation**
- **Models: Backward/Forward Language Modeling**
  - sep-B/F v.s. syn-B/F v.s. asyn-B/F

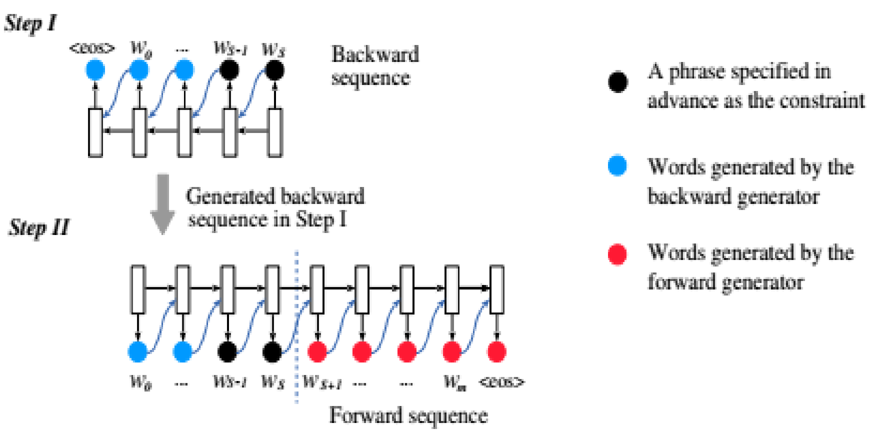
Mou et al., arXiv'15

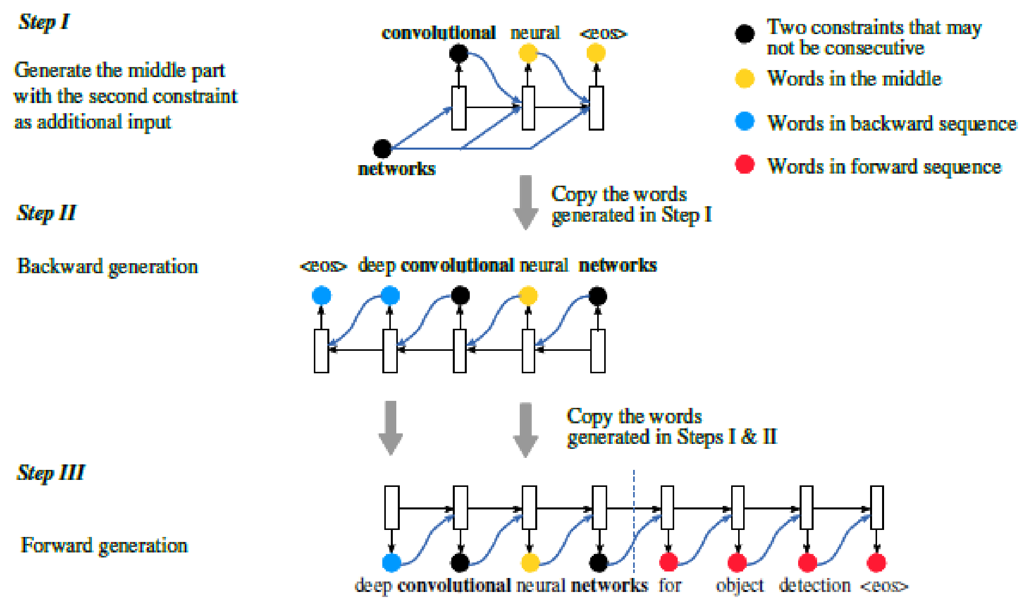


Extensions & Applications

- **Extensions**
  - Constraints of phrases
  - Constraints of multi-terms

Mou et al., arXiv'15





- Applications
  - Two-step conversation generation
    - Step 1: keyword generation
    - Step 2: Backward/Forward language generation

2-Step Conversation

- **Overview**
  - Step I: predict a keyword using PMI
  - Step II: sequence generation with the predicted keyword
- Keyword prediction
  - For a query word and a reply word:

$$\mathrm{PMI}(w_q, w_r) = \log \frac{p(w_q, w_r)}{p(w_q)p(w_r)} = \log \frac{p(w_q|w_r)}{p(w_q)}$$

  - For all the words in the query

$$\mathrm{PMI}(w_{q_1} \cdots w_{q_n}, w_r) = \log \frac{p(w_{q_1} \cdots w_{q_n}|w_r)}{p(w_{q_1} \cdots w_{q_n})}$$

$$\approx \log \frac{\prod_{i=1}^{n} p(w_{q_i}|w_r)}{\prod_{i=1}^{n} p(w_{q_i})} = \sum_{i=1}^{n} \log \frac{p(w_{q_i}|w_r)}{p(w_{q_i})} = \sum_{i=1}^{n} \mathrm{PMI}(w_{q_i}, w_r)$$

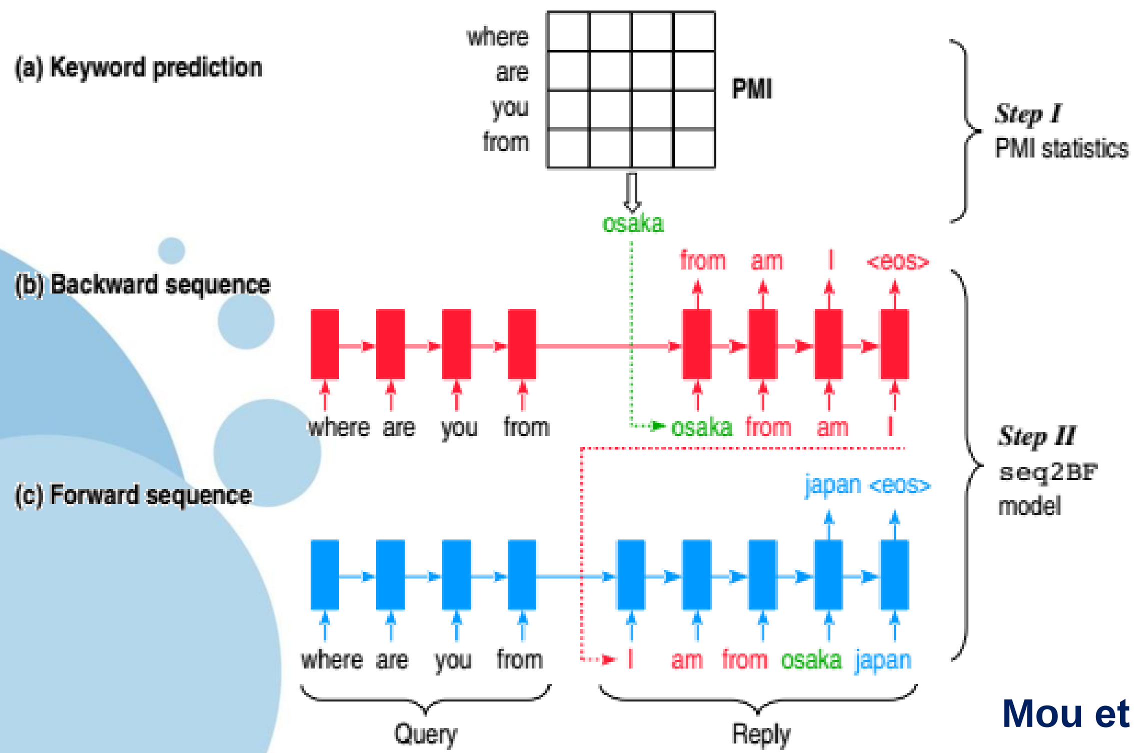
- PMI prefers a word that is most “mutually informative” with the query

Mou et al., COLING’16

Content Introducing

- **Generation Process**
  - Backward sequence
  - Forward sequence

$$p\left(\begin{array}{c} r_{k-1}\cdots r_1 \\ \hdashline r_{k+1}\cdots r_m \end{array}\middle| r_k, q\right) = \prod_{i=1}^{k-1} p^{(\mathrm{bw})}(r_{k-i}|r_k, q, \cdot) \prod_{i=1}^{m-k} p^{(\mathrm{fw})}(r_{k+i}|r_k, q, \cdot)$$



Mou et al., COLING'16

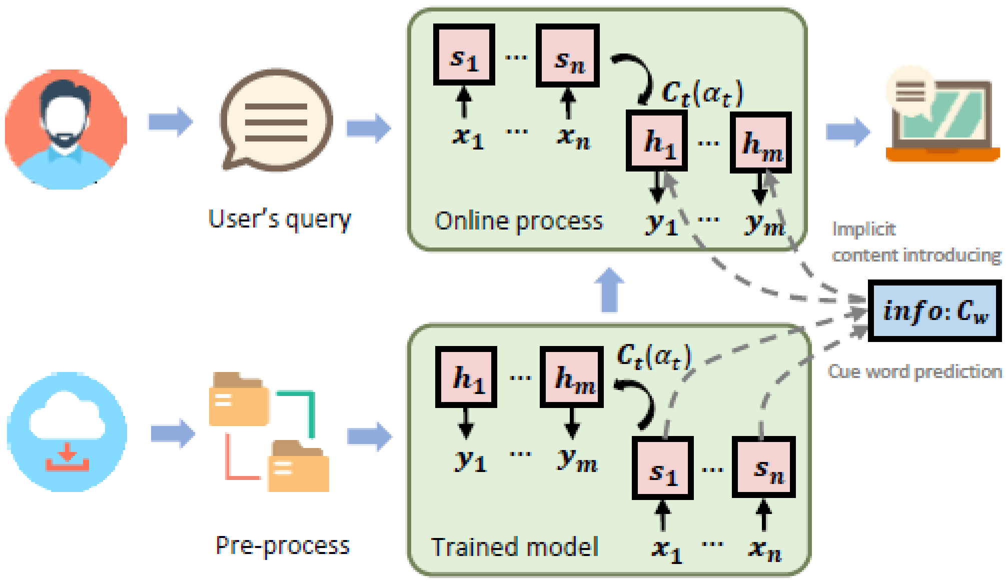
Content Introducing Case

- **Case studies**

	Chinese	English (translated)
Query	李有男友公开过了	It's known that *Li*† has a boyfriend.
Grountruth	都已经分了 之前李的贴吧都在讨论了	Broken up. There's discussion in her *Tieba*.‡
seq2seq	我是男的	I am a male
$\mathrm{seq2BF}_{-}$	我的男友	My boyfriend
$\mathrm{seq2BF}_{+}$	有**绯闻**男友	Has a **rumor**ed boyfriend
Query	人大复试飘过	Passed second-round exam of Renming Univ.
Grountruth	这么牛，什么专业	Cool, what's your major
seq2seq	我也是！	Me too!
$\mathrm{seq2BF}_{-}$	我也是！	Me too!
$\mathrm{seq2BF}_{+}$	**分数**是什么	What is your **score**
Query	挺漂亮的 祝福祝福	So beautiful, congratulations
Grountruth	下手慢了哈。现在	You're late till now
seq2seq	谢谢	Thank you
$\mathrm{seq2BF}_{-}$	是的，谢谢	Yes, thank you
$\mathrm{seq2BF}_{+}$	谢谢**夸奖**	Thank you for **praising**
Query	我要换头像了！	Want to change a photo
Grountruth	打算换成什么啊～	What do you like...
seq2seq	我是你的头像	I'm in your photo
$\mathrm{seq2BF}_{-}$	我是你的头像	I'm in your photo
$\mathrm{seq2BF}_{+}$	第一张**图像**是谁	Who is in your first **photo**

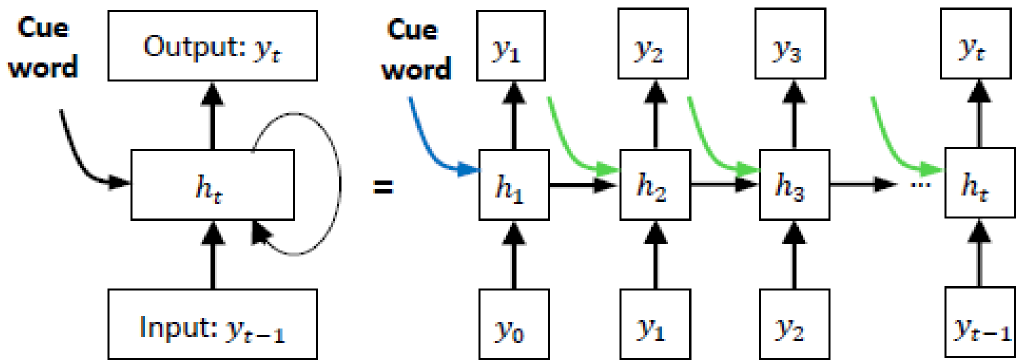
Implicit Content Introducing

- Hard constraint vs soft constraint
- Does the constraint word really need to exist, explicitly?

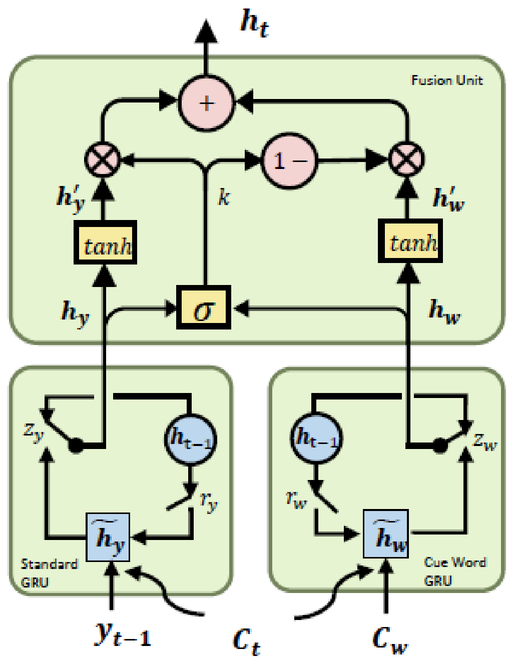


Soft Constraint

- **Two ways to inject information**



- **Model**
  - Standard GRU
  - Cue word GRU
  - Fusion units



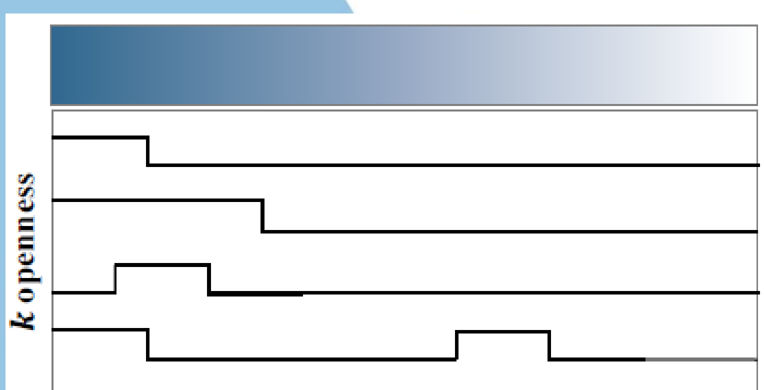
Yao et al., EMNLP’17

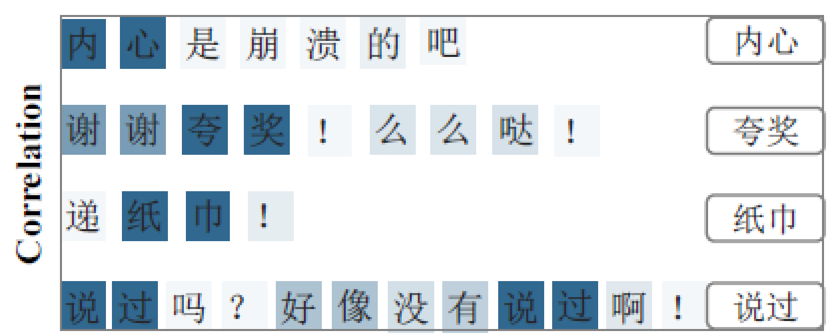
Illustrations

- **An example query**

Query (**Cue word**)	班主任还拍了我超级丑的照片已被笑死.(**上镜**) The teacher took a photo of me; it was really ugly and people laughed at me. (**Photogenic**)	Related Criterion	Labels
Reply1	谁的照片？ Whose photo?	Logic Consistency	Unsuitable
Reply2	什么时候拍的？ When did he took the photo?	Implicit Relevance	Neutral
Reply3	抱抱。Give you a hug.	Implicit Relevance	Neutral
Reply4	我拍照也都是巨丑的！ My photos are also ugly!	——	Suitable

- **Visualization**





Diversity in Conversation

- **A well-known problem for conversation generation**
  - Diversity-promoting
- **Maximum mutual information criterion**
  - Standard objective

$$\hat{T} = \arg\max_{T} \left\{ \log p(T|S) \right\}$$

  - MMI objective

$$\log \frac{p(S,T)}{p(S)p(T)}$$

Li et al., NAACL'16

Input: What are you doing?			
-0.86	I don't know.	-1.09	Get out of here.
-1.03	I don't know!	-1.09	I'm going home.
-1.06	Nothing.	-1.09	Oh my god!
-1.09	Get out of the way.	-1.10	I'm talking to you.
Input: what is your name?			
-0.91	I don't know.	...	
-0.92	I don't know!	-1.55	My name is Robert.
-0.92	I don't know, sir.	-1.58	My name is John.
-0.97	Oh, my god!	-1.59	My name's John.
Input: How old are you?			
-0.79	I don't know.	...	
-1.06	I'm fine.	-1.64	Twenty-five.
-1.17	I'm all right.	-1.66	Five.
-1.17	I'm not sure.	-1.71	Eight.

Diversity in Conversation

- **MMI objective**

$$\hat{T} = \arg\max_{T} \left\{ \log p(T|S) - \log p(T) \right\}$$

- **Penalty parameter**

$$\hat{T} = \arg\max_{T} \left\{ \log p(T|S) - \lambda \log p(T) \right\}$$

- **Bayes theorems**

$$\log p(T) = \log p(T|S) + \log p(S) - \log p(S|T)$$

- Final representation

$$\hat{T} = \arg\max_{T} \left\{ (1-\lambda) \log p(T|S) + \lambda \log p(S|T) - \lambda \log p(S) \right\}$$
$$= \arg\max_{T} \left\{ (1-\lambda) \log p(T|S) + \lambda \log p(S|T) \right\}$$

Input: What are you doing?
1. I've been looking for you.
2. I want to talk to you.
3. Just making sure you're OK.
Input: What is your name?
1. Blue!
2. Peter.
3. Tyler.
Input: How old are you?
1. Twenty-eight.
2. Twenty-four.
3. Long.

Li et al., NAACL'16

Persona in Conversation

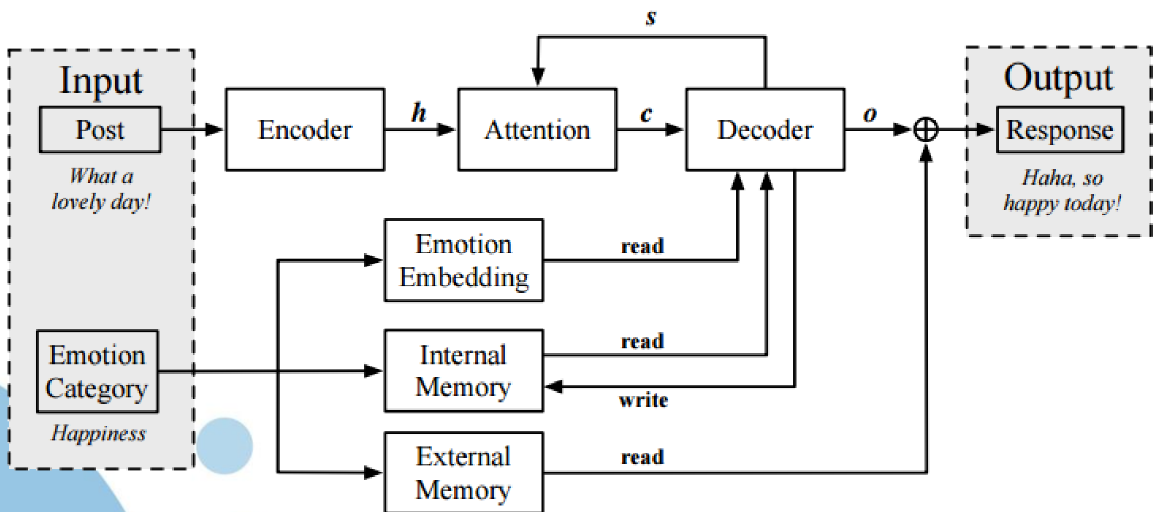
- **What is persona and why?**
  - Motivation examples
  - Incoherent personality makes it impossible for current systems to pass the Turing test
  - Embedding based persona

message	Where do you live now?
response	I live in Los Angeles.
message	In which city do you live now?
response	I live in Madrid.
message	In which country do you live now?
response	England, you?
message	Where were you born?
response	I was born in Canada.
message	Where are you from?
response	England, you?
message	Where did you grow up?
response	I grew up in Texas.
message	How old are you?
response	16 and you?
message	What's your age?
response	18.
message	What is your major?
response	I'm majoring in psychology
message	What did you study in college?
response	English lit.

Li et al., NAACL'16

Emotion in Conversation

- **Emotion is important**
  - Emotion classification
  - Emotion fusion



Post	Model		Response(Original)	Response(Translated)
比情人节更庸俗的事就是，对情人节的接待规格待遇提出要求。 What is more vulgar than Valentine's day is to make specific demands on the reception of the Valentine's day.	Seq2Seq		这才是真正的情人节。	This is a real Valentine's day.
	ECM	*Like*	情人节快乐！	Happy Valentine's Day!
		Happiness	哈哈，这也太浪漫了吧！	Aha, this is too romantic!
		Sadness	我也想要这样的情人节。	I also want this kind of Valentine's day, sigh.
		Disgust	这就是所谓的情人节！	This is the so-called Valentine's day!
		Anger	这是赤裸裸的炫耀啊！	This is shameless show-off!

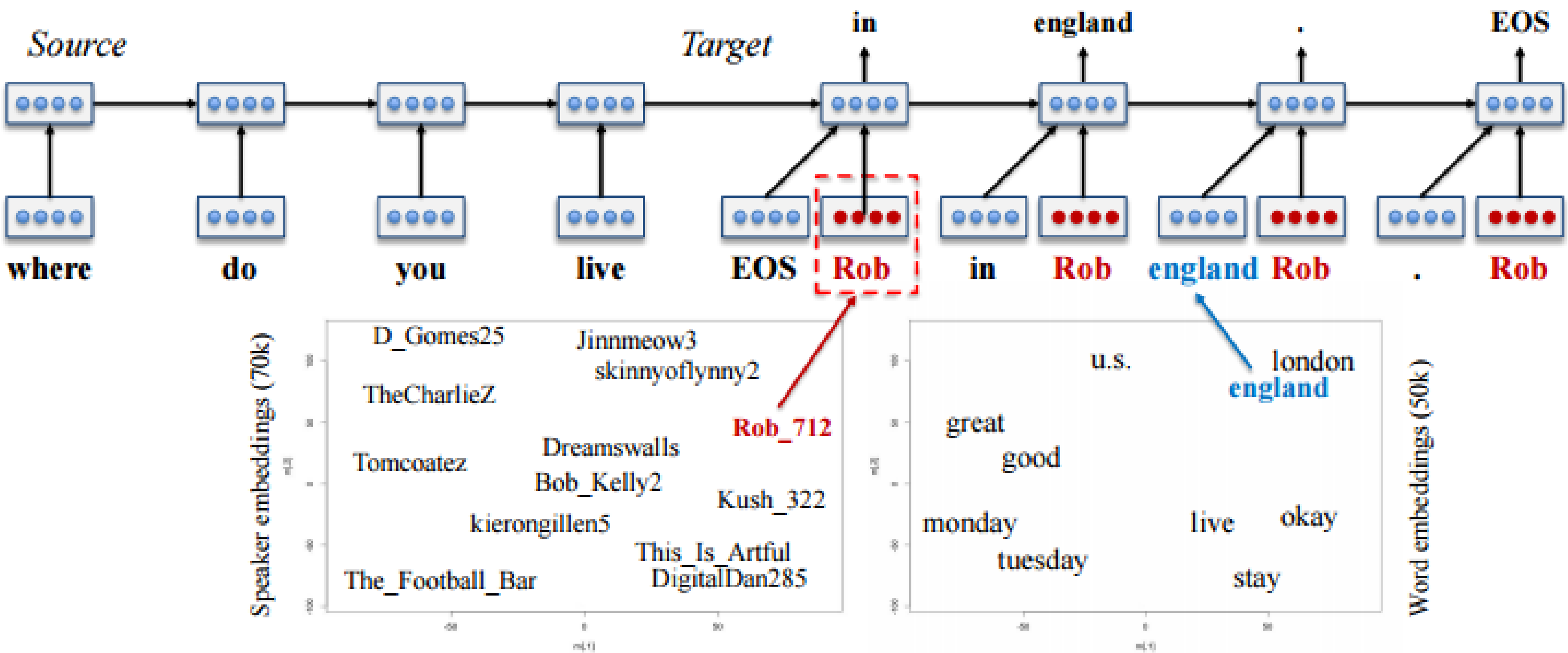
Zhou et al., arXiv'17

Speaker Model

- **What is persona and why?**
  - Persona as additional input
  - Possible benefits: inference in persona

$$\begin{bmatrix} i_t \\ f_t \\ o_t \\ l_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} W \cdot \begin{bmatrix} h_{t-1} \\ e_t^s \\ v_i \end{bmatrix}$$

Li et al., ACL'16



Model Extension

- **Speaker-addressee model**
  - Speaker vector
  - Addressee vector

$$V_{i,j} = \tanh(W_1 \cdot v_i + W_2 \cdot v_2)$$

Li et al., ACL'16

- **Reranking**
  - Persona fit?
  - Message fit?
  - Length penalty

$$\log p(R|M,v) + \lambda \log p(M|R) + \gamma |R|$$

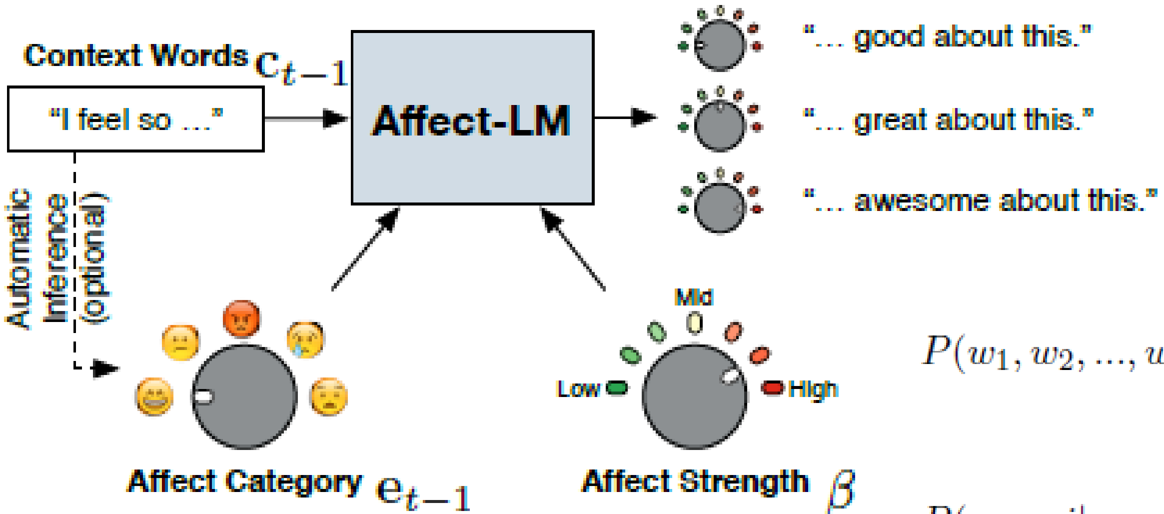
message	Where is your hometown?
baseline	I was born in Canada.
user1	I'm from England.
user2	I'm from Manchester.
user3	I'm from Liverpool.
user4	England. You?
user5	My hometown.
user6	I'm from Texas.
user7	I'm from LA.
user8	England.
user9	I was born here.
user10	I was born in the us.

User1	
message	Where is your hometown?
response	I'm from England.
message	Where are you from?
response	I'm from England.
message	In which city do you live now?
response	I live in London.
message	In which country do you live now?
response	I live in England.
message	What is your major?
response	Business. You?
message	What did you study in college?
response	I did business studies.
message	How old are you?
response	I'm 18.
message	What is your age?
response	I'm 18.

Semantically Conditioned LSTM

- **Affect-LM: customizable affective text generation**

Ghosh et al., ACL 2017



$$P(w_1, w_2, ..., w_M) = \prod_{t=1}^{t=M} P(w_t | w_1, w_2, ..., w_{t-1})$$

$$P(w_t = i | \mathbf{c}_{t-1}) = \frac{\exp(\mathbf{U_i}^T \mathbf{f}(\mathbf{c}_{t-1}) + b_i)}{\sum_{j=1}^{V} \exp(\mathbf{U_j}^T \mathbf{f}(\mathbf{c}_{t-1}) + b_j)}$$

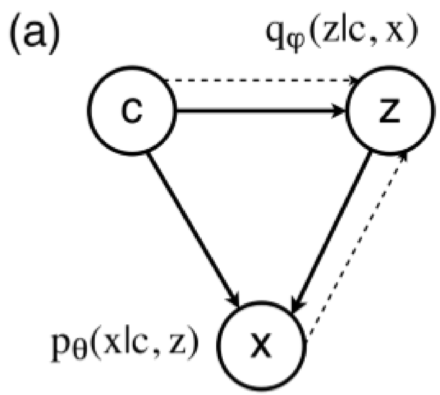
$$P(w_t = i | \mathbf{c}_{t-1}, \mathbf{e}_{t-1}) = \frac{\exp(\mathbf{U_i}^T \mathbf{f}(\mathbf{c}_{t-1}) + \beta \mathbf{V_i}^T \mathbf{g}(\mathbf{e}_{t-1}) + b_i)}{\sum_{j=1}^{V} \exp(\mathbf{U_j}^T \mathbf{f}(\mathbf{c}_{t-1}) + \beta \mathbf{V_j}^T \mathbf{g}(\mathbf{e}_{t-1}) + b_j)}$$

Five specific affect categories e_{t-1}
B is affect strength

Conditional VAE

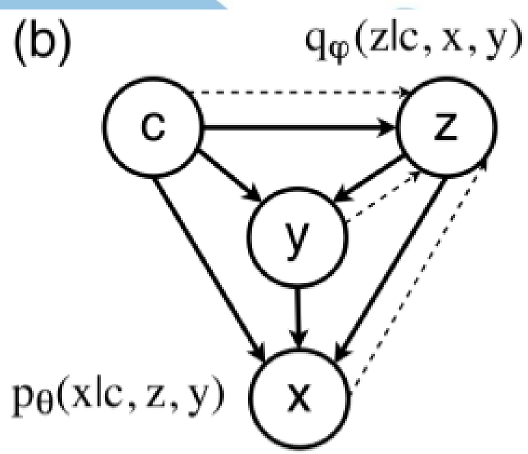
- **Conditional Variational Auto Encoder (CVAE)**

zhao et al., ACL'17



$$\begin{aligned}\mathcal{L}(\theta,\phi;x,c) &= -KL(q_\phi(z|x,c)\|p_\theta(z|c)) \\ &+ \mathbf{E}_{q_\phi(z|c,x)}[\log p_\theta(x|z,c)] \\ &\le \log p(x|c)\end{aligned}$$

- **Knowledge-Guided CVAE (kgCVAE)**

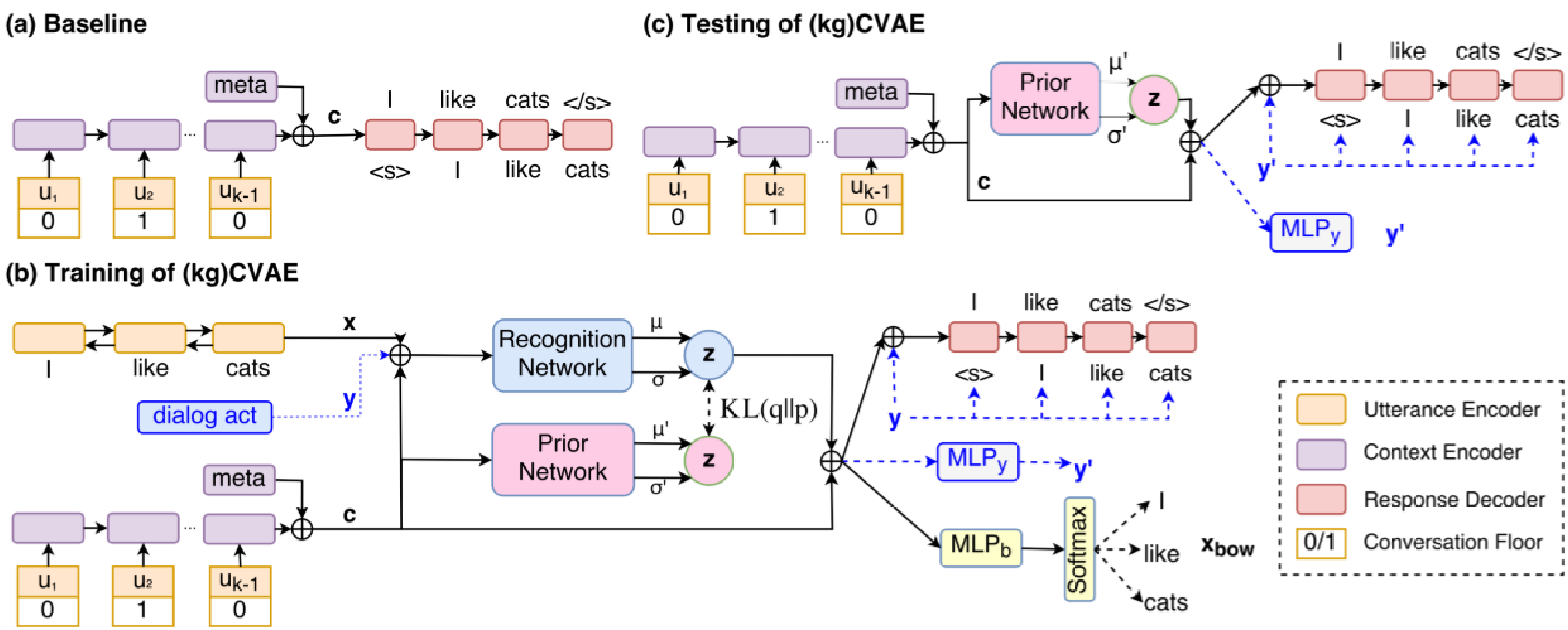


$$\begin{aligned}\mathcal{L}(\theta,\phi;x,c,y) &= -KL(q_\phi(z|x,c,y)\|P_\theta(z|c)) \\ &+ \mathbf{E}_{q_\phi(z|c,x,y)}[\log p(x|z,c,y)] \\ &+ \mathbf{E}_{q_\phi(z|c,x,y)}[\log p(y|z,c)]\end{aligned}$$

Conditional VAE

- **Conditional Variational Autoencoder**

zhao et al., ACL'17

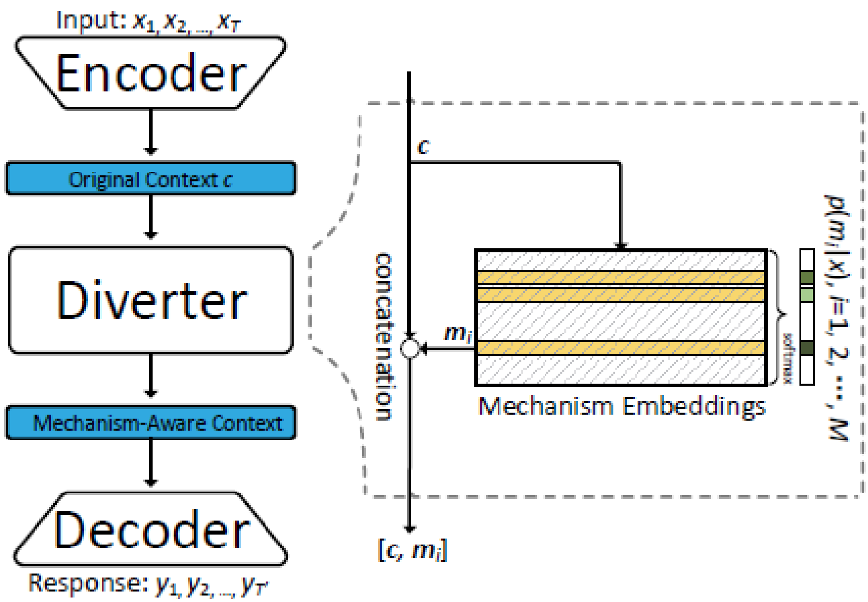


$$\log p(x_{bow}|z,c) = \log \prod_{t=1}^{|x|} \frac{e^{f_{x_t}}}{\sum_j^V e^{f_j}}$$

Mechanism Aware Generation

Mechanism-aware Neural machine for Dialogue Generation

Zhou et al., AAAI 2017



$$p(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{M} p(\boldsymbol{y}, m_i|\boldsymbol{x}) = \sum_{i=1}^{M} p(m_i|\boldsymbol{x}) p(\boldsymbol{y}|m_i, \boldsymbol{x})$$

$$p(m_i|\boldsymbol{x}) = \frac{\exp g(\boldsymbol{m}_i, \boldsymbol{c})}{\sum_{k=1}^{M} \exp g(\boldsymbol{m}_k, \boldsymbol{c})}$$

Encoder: returns context c
Diverter: calculate distribution over mechanism embeddings and return $[c, m_i]$
Decoder: generate response according to $[c, m_i]$

$$g(\boldsymbol{m}_i, \boldsymbol{c}) = \boldsymbol{m}_i^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{t}$$

$$\boldsymbol{t} = [\max\{\widetilde{t}_{2j-1}, \widetilde{t}_{2j}\}]^{\mathrm{T}}_{j=1,2,\cdots,l_c}$$

$$\widetilde{\boldsymbol{t}} = \boldsymbol{W}_c \boldsymbol{c}$$

Mechanism Aware Generation

$$\sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{D}^c} \log p(\boldsymbol{y}|\boldsymbol{x}) = \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{D}^c} \log \sum_{i=1}^{M} p(m_i|\boldsymbol{x})p(\boldsymbol{y}|m_i,\boldsymbol{x})$$

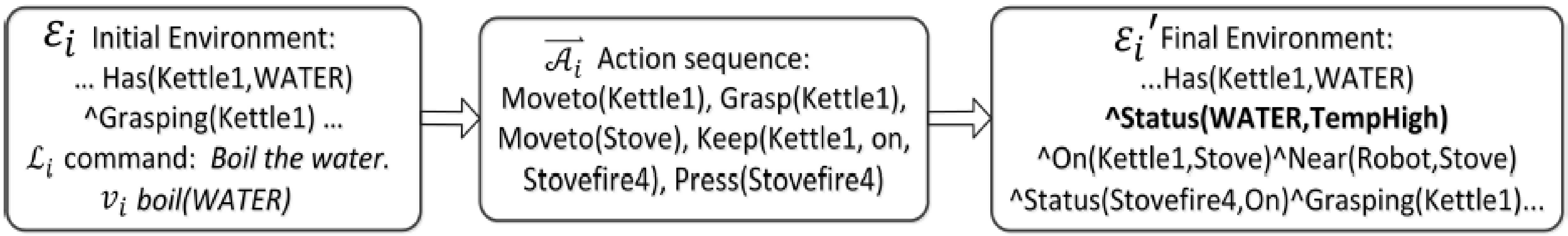
$$p(\boldsymbol{y}, m_l|\boldsymbol{x}) = p(m_l|\boldsymbol{x})p(\boldsymbol{y}|m_l,\boldsymbol{x})$$

Questions	Mechanism 1	Mechanism 2	Mechanism 3	Mechanism 4
我要坐地铁回家 I take subway back home	好的，路上注意安 Ok, be safe on the subway	真是个不错的选择 That's a good choice	坐车去哪里 Where are you going on the subway	坐车一定要注意安 Be careful on the subway
生活中怎样感到开心快乐 How do I feel happy in my life	开心就来和我聊天 Have a talk with me if you are happy	看来你今天心情很不错 You seem to be happy today	有什么开心的事情 What are the happy things you have experieced	和我聊天就很开心 Talk with me and you will be happy
明天考试、求安慰 I have a test tomorrw, and I need some solace	好好复习就好了 Just work hard on the reviews	我会为你加油的 I will stand by you	你要去哪里试试 Where are you taking the test	我会一直安慰你的 I will always be on your side
上班好累啊 The work is really hard	那就好好休息 Take a good rest	要懂得劳逸结合 You need to balance work and life	你是不是累了 Are you tired	上班就是这样 That is work

Reinforcement Learning

Interactive learning of grounded verb semantics

She and Chai ACL 2017



The acquired verb representation (i.e., a goal state hypothesis): boil(x): Status(x,TempHigh)

An example of acquiring state-based representation for verb semantics

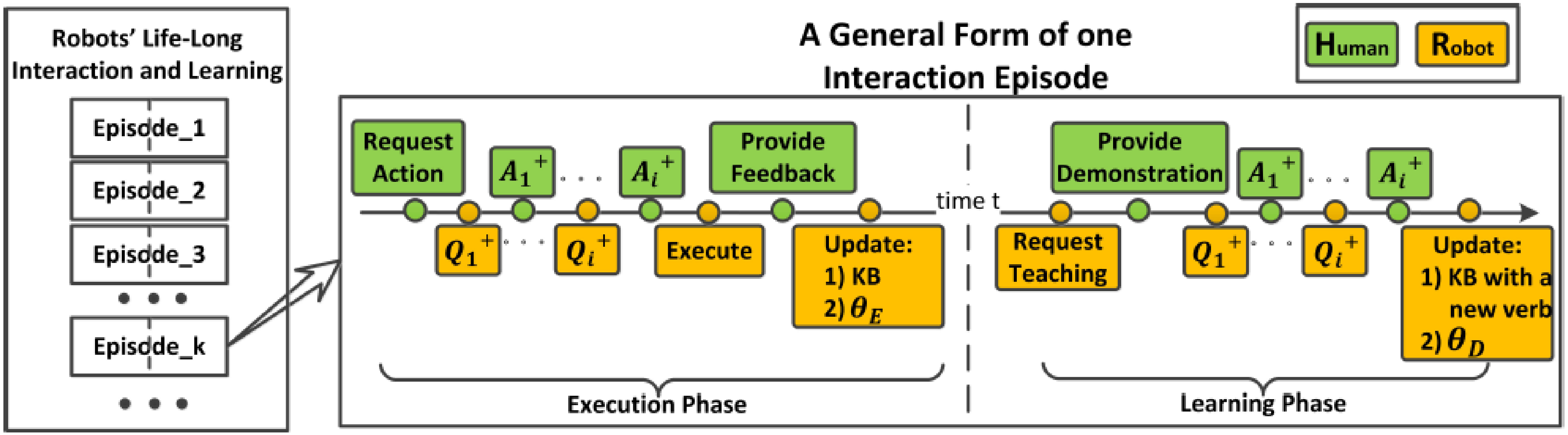
environment $\mathcal{E}_i$

a language command $\mathcal{L}_i$

primitive action sequence $\overrightarrow{\mathcal{A}_i}$

final environment $\mathcal{E}_i'$ that results from the execution of $\overrightarrow{\mathcal{A}_i}$ in $\mathcal{E}_i$

Reinforcement Learning



KB stands for knowledge base

θ_E stands for Interaction Strategy for Execution

θ_D stands for Interaction Strategy for Learning

Reinforcement Learning

Interactive learning of grounded verb semantics

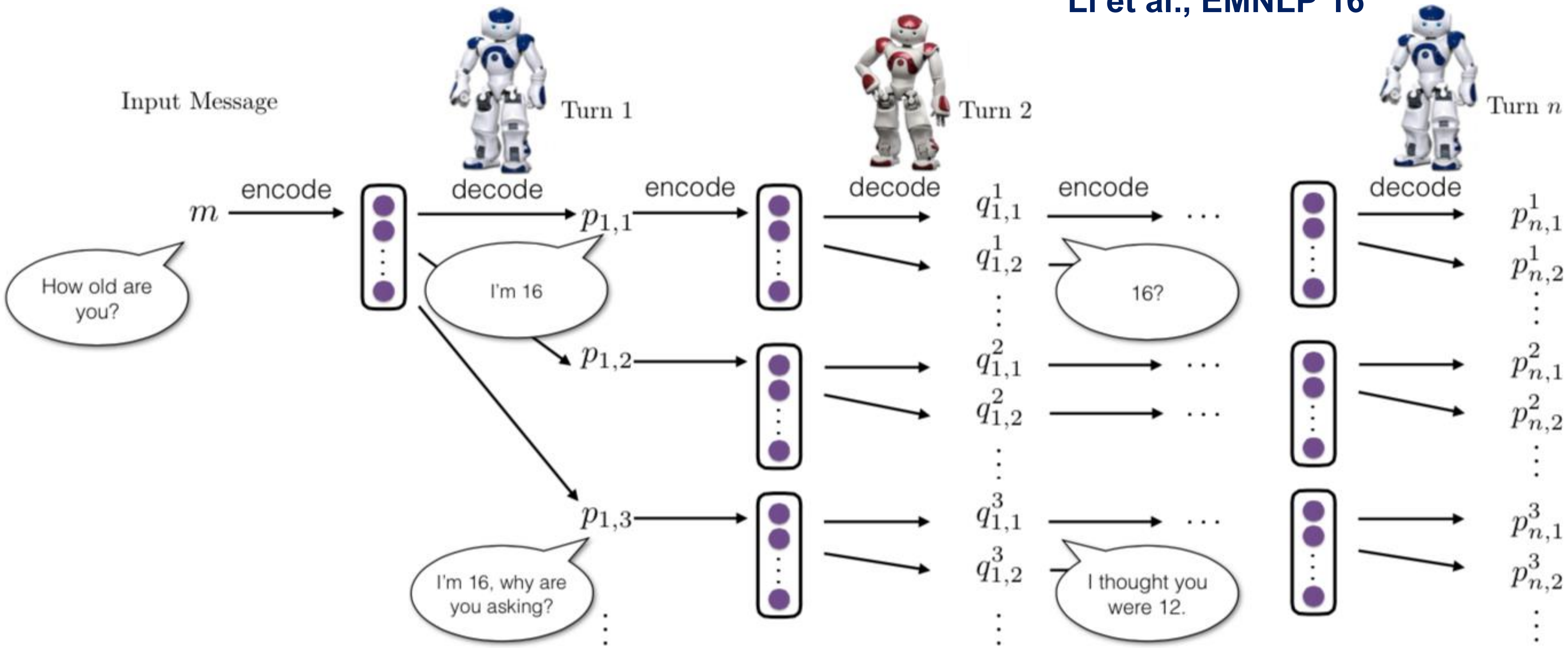
She and Chai ACL 2017

Only demonstration(She and Chai, 2016)	Interactive learning (this work)
Execution Example H1: Boil the water. R1: (*select a hypothesis and execute the action accordingly*) H2: (*provide true/false feedback on whether the execution is correct*)	**Execution Example** H1: Boil the water. R1: I see a silver object on the table, is that a pot? H2: Yes. R2: Should I leave the pot on the stove after boil? H3: Yes. R3: Ok. (*select a hypothesis to execute*) H4: Thank you.
Learning Example H3: Boil the water. R3: I don't know how to boil, could you show me an example? H4: (*human shows the robot how to achieve the corresponding action through a sequence of primitive actions*) R4: (*indicate the acceptance and induce a hypothesis to represent verb semantics*)	**Learning Example** H5: Boil the water. R5: I don't know how to boil, could you show me an example? H6: Ok. (*demonstrate the task, at the end put the pot on a heat pad*) R6: So after boiling something, should I put it on a heat pad? H7: No, you don't have to. R7: I see. (*induce goal state hypothesis*)

Reinforcement Learning

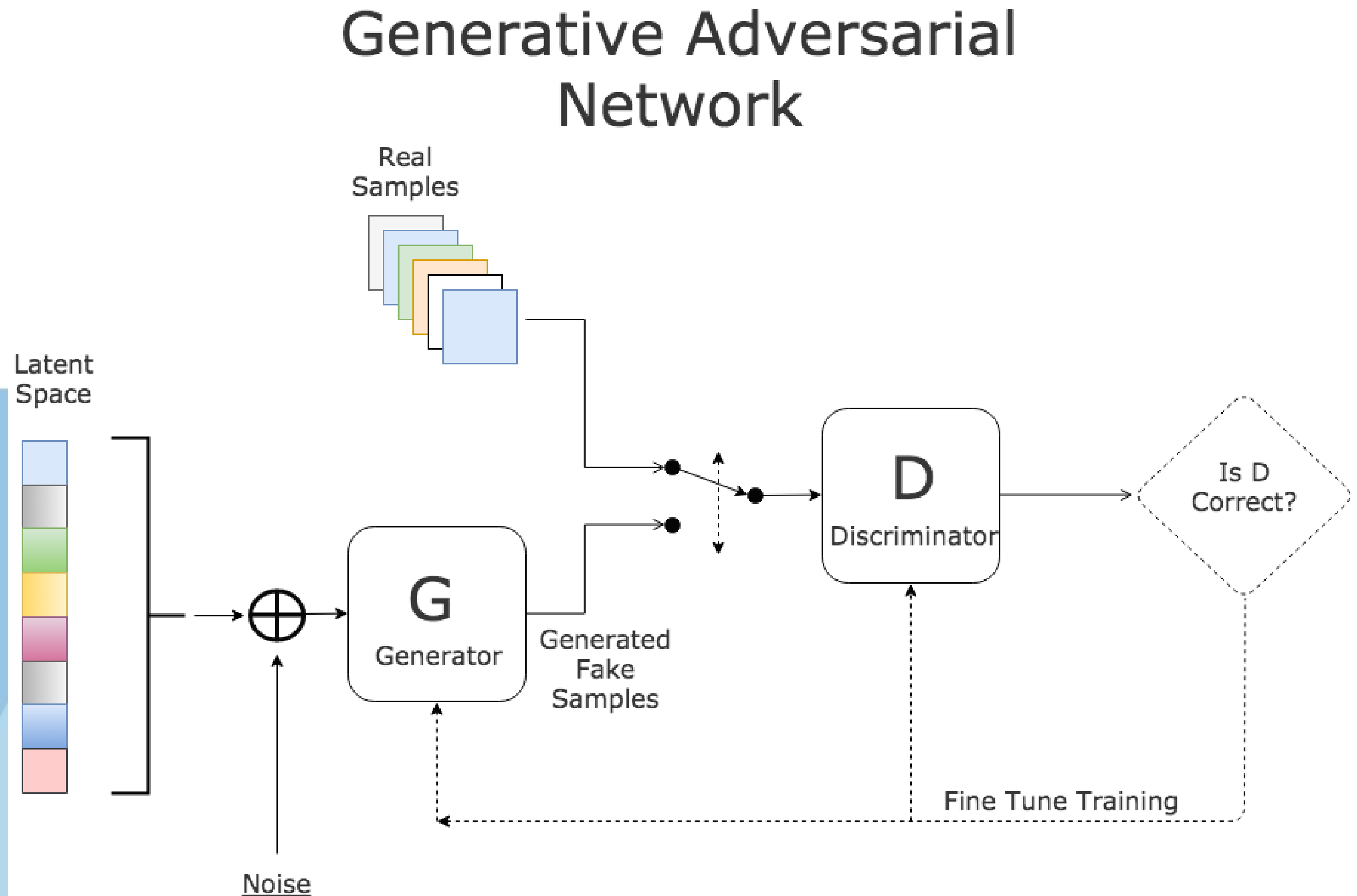
- **Modeling the future direction by RL**
  - Conversation between two virtual agents
  - Explore the space of possible actions while learning to maximize expected reward

Li et al., EMNLP'16



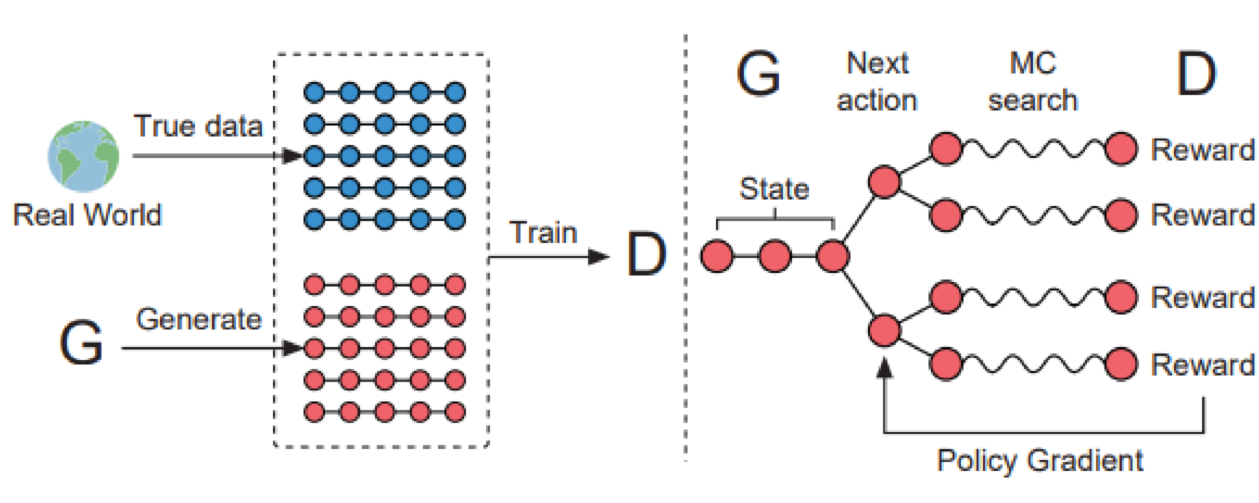
GAN

- **Idea of adversarial training: GAN**



GAN in NLP

- **Languages are different from images**
  - Consecutive distributions vs. discrete distributions
- **SeqGAN**
  - NO intermediate reward
  - Policy: to maximize the expected reward of the whole sequence after generation
  - Generator and discriminator



Yu et al., AAAI'17

GAN in Conversation

- **Seq2SeqGAN**
  - **With the input sequence, using the idea of SeqGAN**
- **Generative model with discrimination via policy**
- **Reward with every generation step**
- **Adversarial evaluation**

Input	tell me ... how long have you had this falling sickness ?
Vanilla-MLE	i 'm not a doctor .
Vanilla-Sample	well everything you did was totally untrue .
REINFORCE	i don 't know how long it 's been .
REGS Monte Carlo	A few months, I guess .
Input	so i had the doctors test sammy 's response to conditioning .
Vanilla-MLE	sammy wrote the test sammy wrote the test .
Vanilla-Sample	objects pick objects up objects objects objects objects objects
REINFORCE	i 'm not sure that 's relevant .
REGS Monte Carlo	so he took the pills .
Input	what are you going to do ?
Vanilla-MLE	i 'm going to the movies .
Vanilla-Sample	get him outta here first !
REINFORCE	i 'm going to get you .
REGS Monte Carlo	i 'm going to get you a new car .
Input	they fear your power your intellect .
Vanilla-MLE	you 're the only one who knows what 's going on .
Vanilla-Sample	when they are conquered and you surrender they will control all of us .
REINFORCE	i 'm afraid i 'm not ready yet .
REGS Monte Carlo	i 'm not afraid of your power .

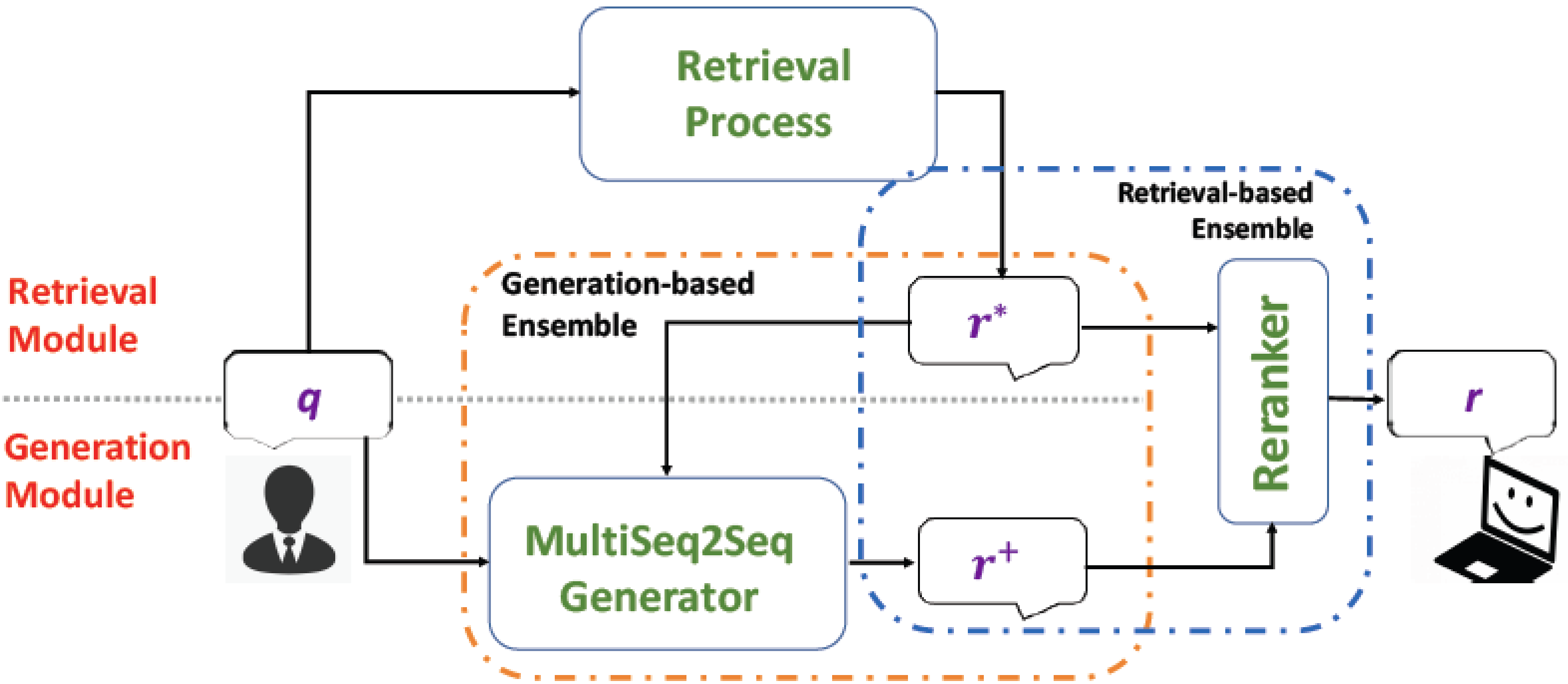
Li et al., EMNLP'17

GENERATION + RETRIEVAL CONVERSATION SYSTEM

Motivation

- **Why**
  - Retrieval is not enough?
  - Generation is not enough as well?
- **Retrieval + generation framework**

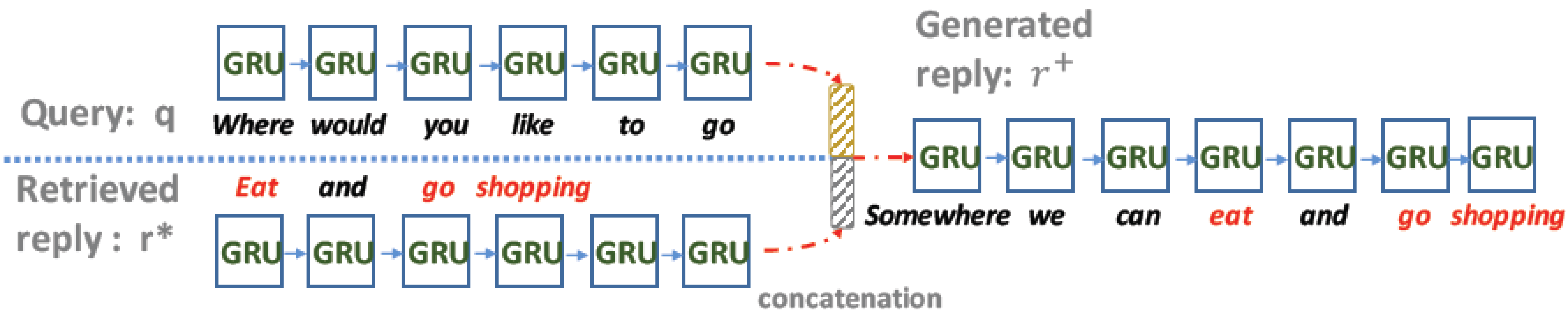
Song et al., arXiv'16



Methodology

- **Generation process**
  - Component ensemble

Song et al., arXiv’16



	Utterance (Translated)	Reranked
Query	这款手机的拍照效果真不错 (This mobile phone’s photo effect is pretty good.)	
Retrieved	所以好心动 (I really have a crush on it.)	
`biseq2seq`	心动不如行动 (Rushing for it rather than having a crush on it.)	√
`seq2seq`	哈哈 (Haha)	
Query	有给孕妇推荐的保健品啥的么？ (Could you please recommend some health products for pregnant woman?)	
Retrieved	打算上孕妇装？ (Do you want metarnity clothes?)	
`biseq2seq`	打算买护肤品么？ (Do you want some skin care products?)	√
`seq2seq`	不好意思 (I’m sorry.)	

EVALUATION

Automatic Evaluation

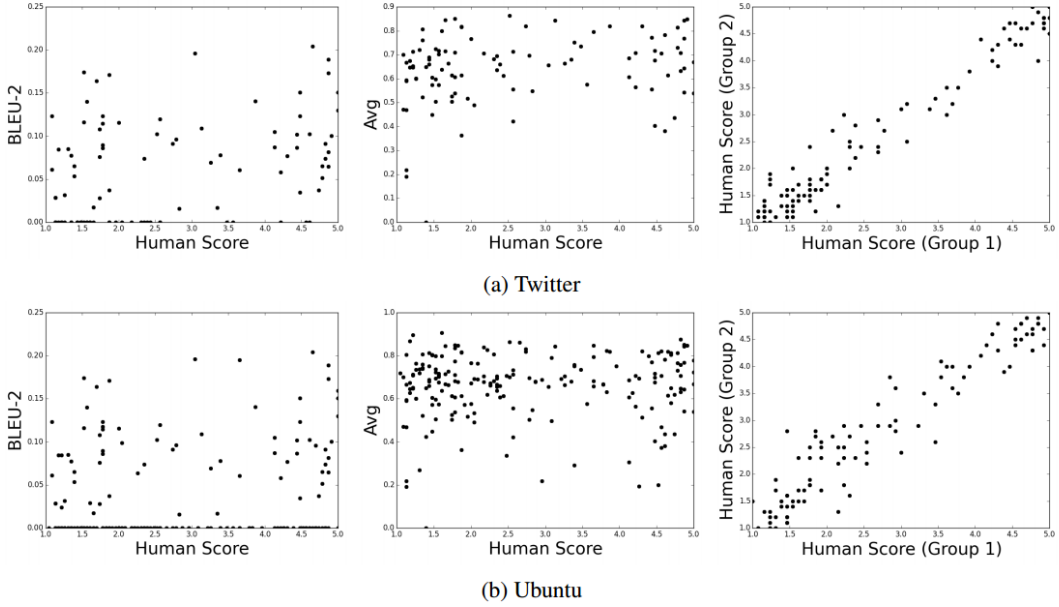
- **Automatic evaluation is crucial to the research of language generation tasks such as:**
  - Machine translation
    - BLEU, NIST, METEOR, …
    - Shared tasks for evaluation metrics in WMT
  - Text summarization
    - ROUGE, Pyramid, …
  - Dialog systems
    - ????

Typically Used Metrics in Dialog Systems

- **Human evaluation (widely used)**
  - Pair-wise
  - Point-wise
- **Automatic evaluation metrics**
  - **BLEU** [Ritter et al., 2011; Li et al., 2015; Sordoni et al., 2015; Song et al., 2016]
  - **Information**: entropy, perplexity [Serban et al. 2016 and Mou et al. 2016]
  - **Diversity**: distinc-1, distinc-2 [Li et al., 2015]
  - Average response length [Serban et al. 2016, Mou et al. 2016]

How NOT to Evaluate Dialogue Systems

- **Liu et al. (2016) show the weak correlation of existing metrics (e.g., BLEU, ROUGE and METEOR) with human judgments for dialog systems**
  - Significant diversity in the space of valid replied to a given input.
  - Utterances are typically short and casual in open-domain dialog systems.



Liu et al., EMNLP'16

ADEM Model

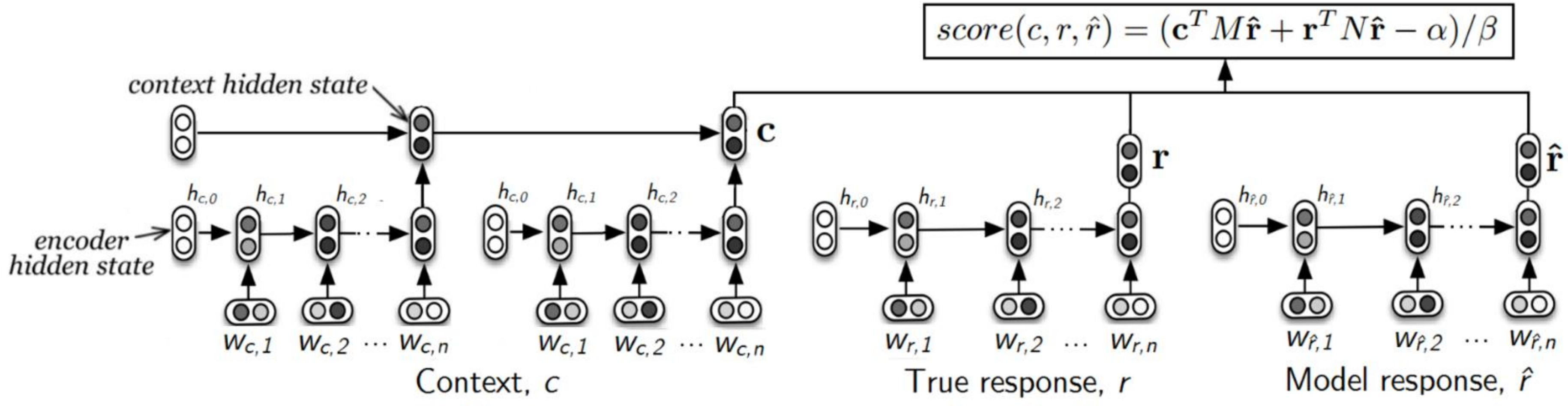


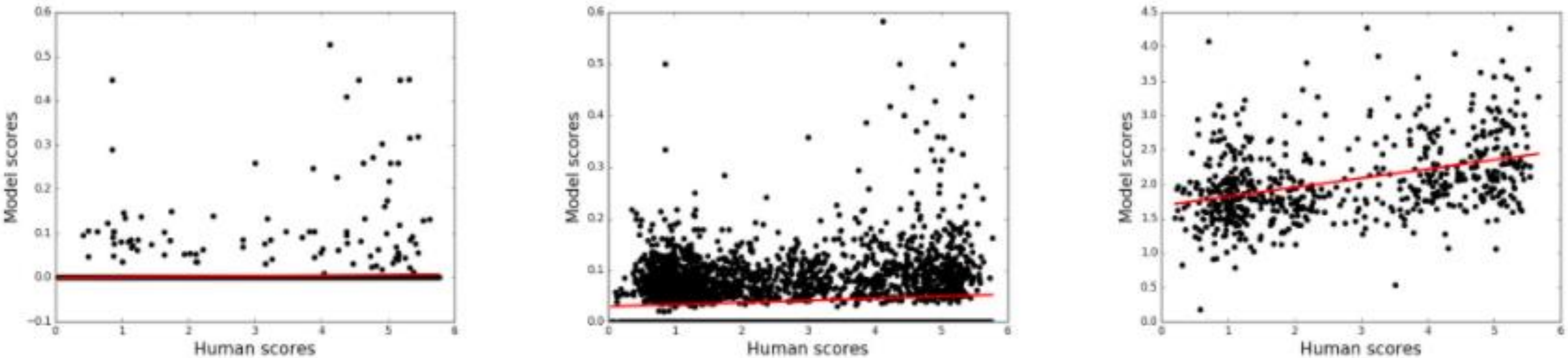
Figure 2: The ADEM model, which uses a hierarchical encoder to produce the context embedding **c**.

- Predict a score of a reply given its query (context) and a ground truth reply.
- It requires massive human-annotated scores to train the network

Lowe et al., ACL'17

ADEM Results

Context	Reference response	Model responses	Human score	ADEM score
photo to see my television debut go to - some. some on <url> - hehe <url> → it really was you? i thought ppl were recognizing someone who looked like you! were the oysters worth the wait?	yeah it was me . haha i 'd kinda forgotten about it it was filmed a while ago	1) i'm not sure. i just don't know what to do with it.	3	1.602
		2) you heard the horsepower productions remix of lee scratch perry's 'exercising' off his 'mighty upsetter' album?	1	1.513
		3) you wont chug a fuzzy peach navel	1	1.744
		4) they were!	5	3.274
just beat call of duty!! → want a cookie? → yes!! → come get it	im in kenmore at the moment	1) i'm gonna get a new phone some moro	1	1.848
		2) no way man.	5	4.265
		3) wow i just got a free pizza coupon! get yours before theres no more! <url>	1	0.921
		4) i'm going to go to the mall.	1	2.634
am i out of twitter jail yet? testing → yeah. i posted bail → thanks. i am a right chatter tweetbox on sundays. same happened last sunday lol	any news on meeting our user ? i go to the us on friday and i don 't want to miss anything arranged	1) i'm not sure if i'm going to be able to get it.	3	1.912
		2) good to see another mac user in the leadership ranks	4	1.417
		3) awww poor baby hope u get to feeling better soon. maybe some many work days at piedmont	2	1.123
		4) did you tweet too much?	5	2.539

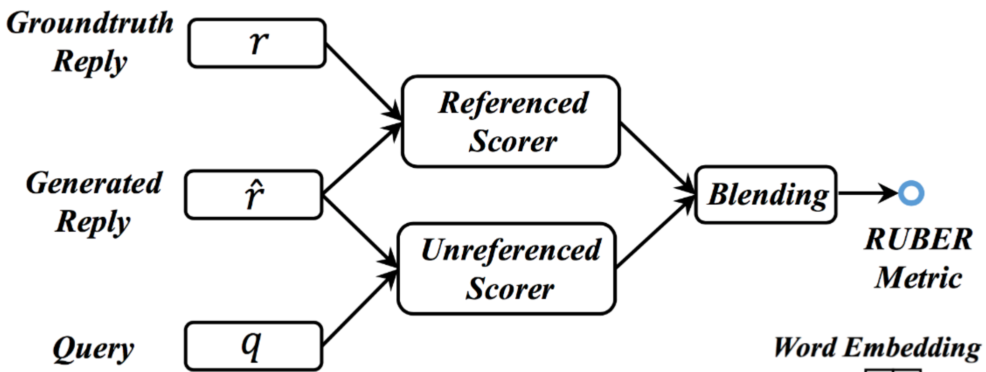


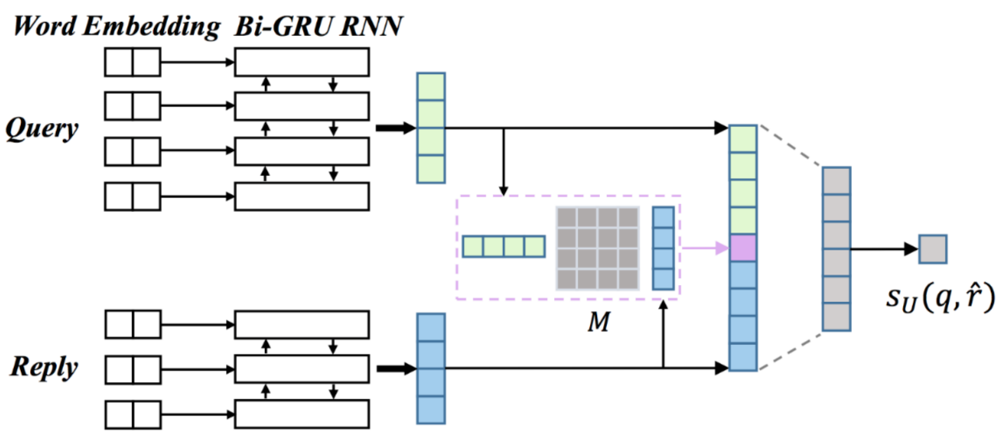
(a) BLEU-2 (b) ROUGE (c) ADEM

- **Higher correlations compared with BLEU and ROUGE**

RUBER

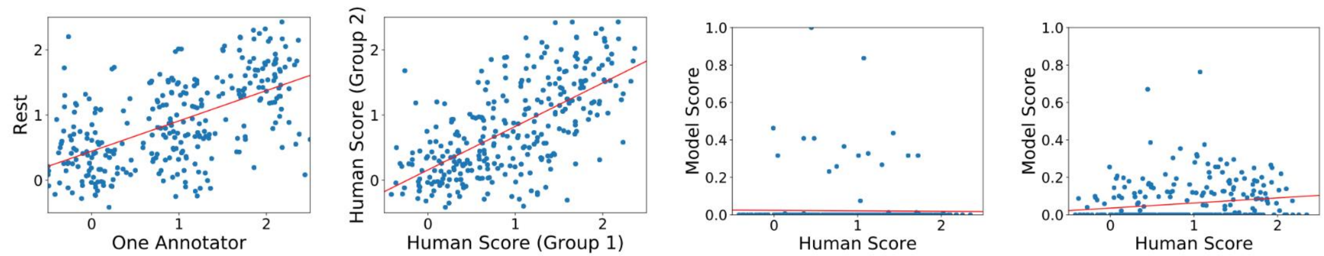
- **RUBER: a Referenced metric and Unreferenced metric Blended Evaluation Routine for open-domain dialog systems**



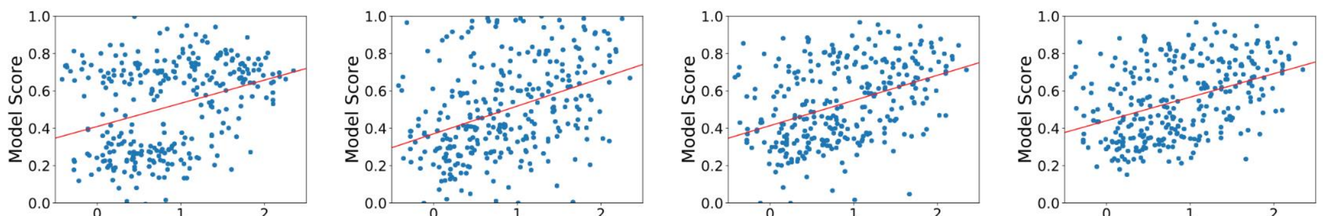


RUBER Results

Metrics		Retrieval (Top-1) Pearson(p-value)	Retrieval (Top-1) Spearman(p-value)	Seq2Seq (w/ attention) Pearson(p-value)	Seq2Seq (w/ attention) Spearman(p-value)
Inter-annotator	Human (Avg)	0.4927(<0.01)	0.4981(<0.01)	0.4692(<0.01)	0.4708(<0.01)
	Human (Max)	0.5931(<0.01)	0.5926(<0.01)	0.6068(<0.01)	0.6028(<0.01)
Referenced	BLEU-1	0.2722(<0.01)	0.2473(<0.01)	0.1521(<0.01)	0.2358(<0.01)
	BLEU-2	0.2243(<0.01)	0.2389(<0.01)	-0.0006(0.9914)	0.0546(0.3464)
	BLEU-3	0.2018(<0.01)	0.2247(<0.01)	-0.0576(0.3205)	-0.0188(0.7454)
	BLEU-4	0.1601(<0.01)	0.1719(<0.01)	-0.0604(0.2971)	-0.0539(0.3522)
	ROUGE	0.2840(<0.01)	0.2696(<0.01)	0.1747(<0.01)	0.2522(<0.01)
	Vector pool (s_R)	0.2844(<0.01)	0.3205(<0.01)	0.3434(<0.01)	0.3219(<0.01)
Unreferenced	Vector pool	0.2253(<0.01)	0.2790(<0.01)	0.3808(<0.01)	0.3584(<0.01)
	NN scorer (s_U)	0.4278(<0.01)	0.4338(<0.01)	0.4137(<0.01)	0.4240(<0.01)
RUBER	Min	0.4428(<0.01)	0.4490(<0.01)	**0.4527**(<0.01)	**0.4523**(<0.01)
	Geometric mean	0.4559(<0.01)	0.4771(<0.01)	0.4523(<0.01)	0.4490(<0.01)
	Arithmetic mean	**0.4594**(<0.01)	**0.4906**(<0.01)	0.4509(<0.01)	0.4458(<0.01)
	Max	0.3263(<0.01)	0.3551(<0.01)	0.3868(<0.01)	0.3623(<0.01)



(a) Human (1 vs. rest) (b) Human (Gropu 1 vs. Group 2) (c) BLEU-2 (d) ROUGE



DISCUSSIONS

Where Are We?

- We are doing well, enough?
- Users have really high expectations

Still A Long Way to Go

- **How is AI in NLP now?**
  - There is still a long way to go

Expectations

- **What can we expect in the future?**
  - Dialogue challenges
    - Multi-turn understanding
    - Semantic understanding
  - Data challenges
    - Vertical domains
  - Intelligence challenges
    - Reasoning
    - Evolution

Thank You

- Q & A
- Email to