

Puzzle--How Long Was He Walking?

Every day, Jack arrives at the train station from work at 5 pm.

His wife leaves home in her car to meet him there at exactly 5 pm, and drives him home. One day, Jack gets to the station an hour early, and starts walking home, until his wife meets him on the road. They get home 30 minutes earlier than usual. How long was he walking?

Distances are unspecified. Speeds are unspecified, but constant.

Give a number which represents the answer in minutes.

Puzzle--How Long Was He Walking?

Every day, Jack arrives at the train station from work at 5 pm.

His wife leaves home in her car to **meet him there at exactly 5 pm**, and drives him home. One day, Jack gets to the station **an hour early**, and starts walking home, until his wife meets him on the road. They get home **30 minutes earlier** than usual. How long was he walking?

Distances are unspecified. Speeds are unspecified, but constant.

Give a number which represents the answer in minutes.

Solution

The best way to think about this problem is to consider it from the perspective of the wife. Her round trip was decreased by 30 minutes, which means each leg of her trip was decreased by 15 minutes. Therefore, she met Jack at **4:45pm**.

Since Jack started walking at 4:00pm, he must have been walking for 45 minutes.

AAAI 2019 conference

CHEN Wang

AAAI-19 Outstanding Paper Award

- [How to Combine Tree-Search Methods in Reinforcement Learning](#)

Yonathan Efroni *
Technion, Israel

Gal Dalal
Technion, Israel

Bruno Scherrer
INRIA, Villers les Nancy, France

Shie Mannor
Technion, Israel

- Honorable Mention: [Solving Imperfect-Information Games via Discounted Regret Minimization](#)

Noam Brown
Computer Science Department
Carnegie Mellon University
noamb@cs.cmu.edu

Tuomas Sandholm
Computer Science Department
Carnegie Mellon University
sandholm@cs.cmu.edu

AAAI-19 Outstanding Student Paper Award

- [Zero Shot Learning for Code Education: Rubric Sampling with Deep Learning Inference](#)

Mike Wu¹, Milan Mosse¹, Noah Goodman^{1,2}, Chris Piech¹

¹ Department of Computer Science, Stanford University, Stanford, CA 94305

² Department of Psychology, Stanford University, Stanford, CA 94305
{wumike, mmosse19, ngoodman, piech}@stanford.edu

- Honorable Mention: [Learning to Teach in Cooperative Multiagent Reinforcement Learning](#)

Shayegan Omidshafiei^{1,2}
shayegan@mit.edu

Dong-Ki Kim^{1,2}
dkkim93@mit.edu

Miao Liu^{2,3}
miao.liu1@ibm.com

Gerald Tesauro^{2,3}
gtesauro@us.ibm.com

Matthew Riemer^{2,3}
mdriemer@us.ibm.com

Christopher Amato⁴
camato@ccs.neu.edu

Murray Campbell^{2,3}
mcam@us.ibm.com

Jonathan P. How^{1,2}
jhow@mit.edu

¹LIDS, MIT

²MIT-IBM Watson AI Lab

³IBM Research

⁴CCIS, Northeastern University

AAAI-19 Tutorials

- [Tutorial Page](#)

AAAI-19 Workshops

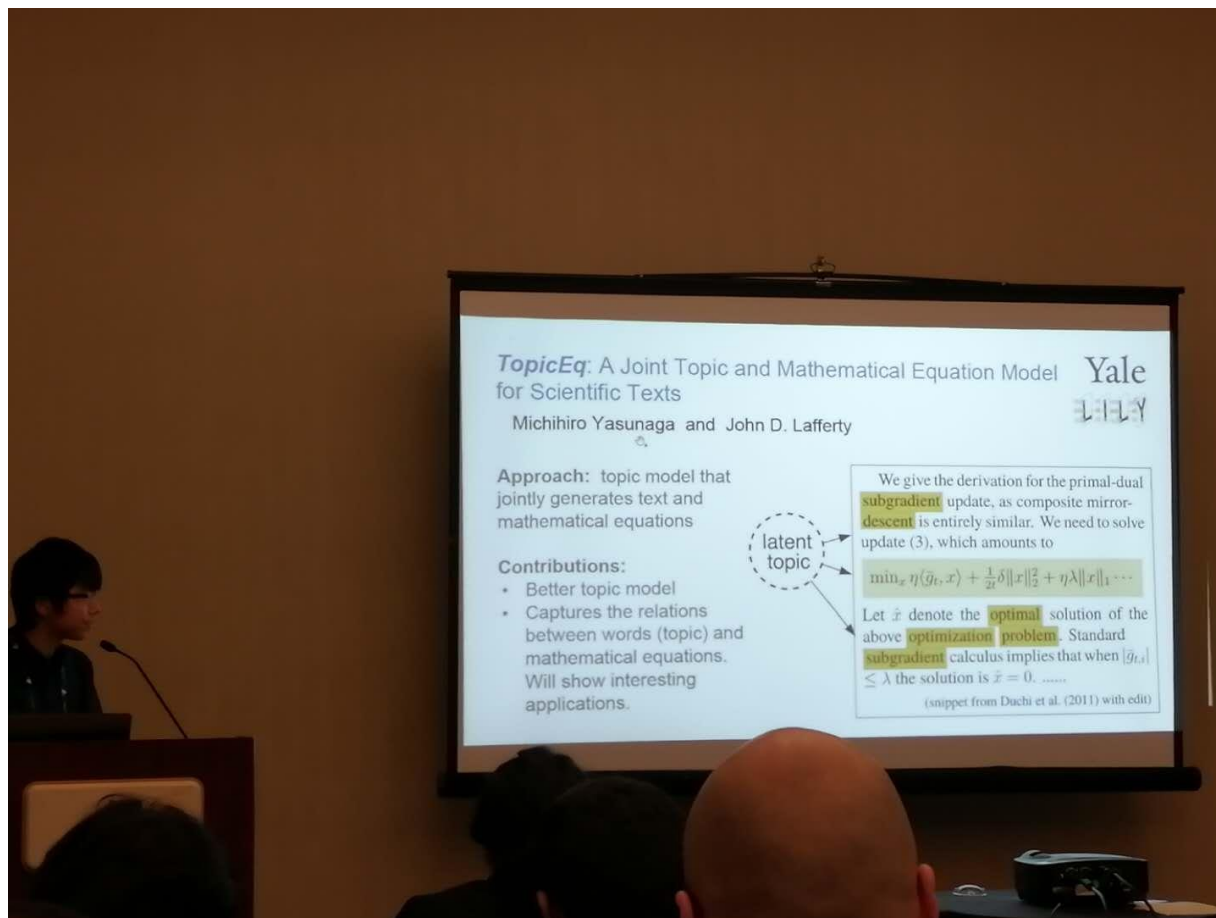
- [Workshop Page](#)

AAAI-19 Invited Talks

- [Invited Talk Page](#)

AAAI-19 Text Generation Papers

TopicEq: A Joint Topic and Mathematical Equation Model for Scientific Texts



Black holes in Einstein gravity. As a warm-up exercise, in this section, we will briefly review the observation made by Padmanabhan [14] by generalizing his discussion to a more general spherically symmetric case. In **Einstein's** general **relativity**, the **gravitational** field equations are

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu}$$

where $G_{\mu\nu}$ is Einstein **tensor** and $T_{\mu\nu}$ is the energy-momentum tensor of matter field. On the other hand, for a general static, spherically symmetric **spacetime**, its metric can be written down as

(snippet from Cai and Ohta (2010))

We give the derivation for the primal-dual **subgradient** update, as composite mirror-**descent** is entirely similar. We need to solve update (3), which amounts to

$$\min_x \eta \langle \bar{g}_t, x \rangle + \frac{1}{2t} \delta \|x\|_2^2 + \frac{1}{2t} \langle x, \text{diag}(s_t)x \rangle + \eta \lambda \|x\|_1$$

Let \hat{x} denote the **optimal** solution of the above **optimization problem**. Standard **subgradient** calculus implies that when $|\bar{g}_{t,i}| \leq \lambda$ the solution is $\hat{x} = 0$. Similarly, when $\bar{g}_{t,i} \leq -\lambda$, then $\hat{x} > 0$, the objective is differentiable, and the **solution** is obtained by setting the **gradient** to zero.

(snippet from Duchi et al. (2011))

Figure 1: The words in a given technical context often characterize the distinctive types of equations used, and vice versa. **Top** topic: Relativity; **bottom** topic: Optimization.

AAAI-19 Text Generation Papers

Knowledge-driven Encode, Retrieve, Paraphrase for Medical Image Report Generation
Christy Y. Li, Xiaodan Liang, Zhiting Hu, Eric Xing

Medical Image Report
Indication: 65-year-old male.
Finding: The cardiomeastinal silhouette is normal in size and contour. Minimal opacification of right apex. No pneumothorax or large pleural effusion. The osseous structures are grossly normal.
Impression: Worsening minimal opacification of right apex suggesting worsening malignancy or malignancy with postobstructive pneumonia.

Abnormality graph

- Hyperexpansion of lungs (0.79)
- Focal airspace consolidation (0.01)
- Elevated heart size (0.94)
- Low lung volumes (0.00)
- Degenerative changes of spine (0.00)

Disease graph

- Consolidation
- Pleural effusion
- Degenerative disease
- Emphysema

Report

Degenerative changes of the spine. No pleural effusion. There is hyperexpansion of the lungs suggesting underlying emphysema. No focal airspace consolidation. Heart size is normal.

AAAI-19 Text Generation Papers

CGMH: Constrained Sentence Generation by Metropolis-Hastings Sampling



Step 0: Key words

BMW sports

Step 1: Insertion

Accept

BMW sports car

Step 2: Insertion

Accept

BMW the sports car

...

Step 6: Insertion

Accept

BMW , the sports car of daily life

Step 7: Replacement

Accept

BMW , the sports car of Future life

Step 8: Insertion

Accept

BMW , the sports car of the Future life

Step 9: Deletion

Reject

BMW , ~~the~~ sports car of the Future life

Step 10: Deletion

Accept

BMW , the sports car of the Future ~~life~~



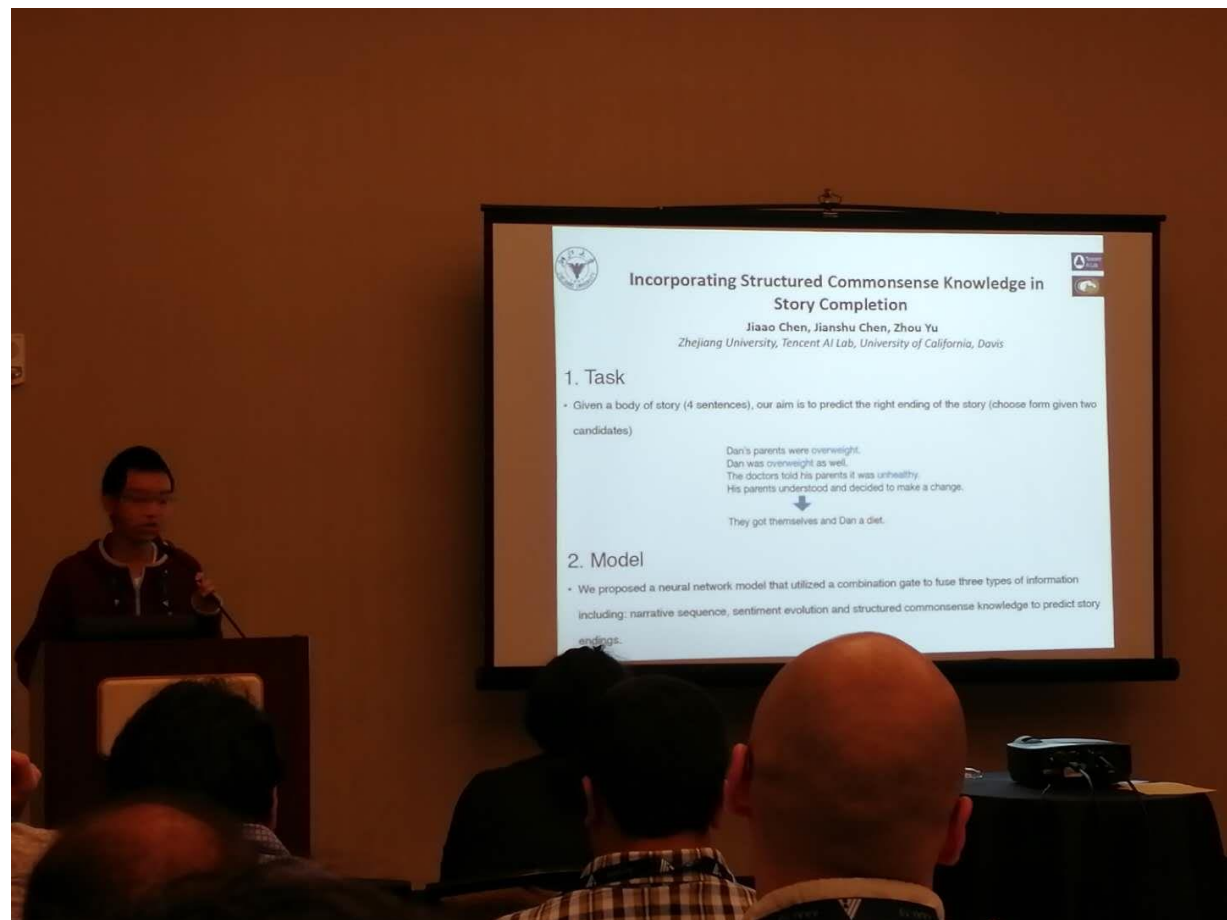
Output:

BMW , the sports car of the Future

Figure 1: CGMH generates a sentence with the constraint of keyword inclusion. At each step, CGMH proposes a candidate modification of the sentence, which is accepted or rejected according to a certain acceptance rate.

AAAI-19 Text Generation Papers

Incorporating Structured Commonsense Knowledge in Story Completion



Dan's parents were **overweight**.
Dan was **overweight** as well.
The doctors told his parents it was **unhealthy**.
His parents understood and decided to make a change.



They got themselves and Dan a **diet**.

(a) An example story



(b) Clues in ConceptNet

Figure 1: (a) shows an example story from ROCStories dataset, words in colors are key-words. (b) shows the key-words and their relations in ConceptNet Knowledge Graph

AAAI-19 Summarization Papers

Abstractive Text Summarization by Incorporating Reader Comments

Abstractive Text Summarization by Incorporating Reader Comments
 Shen Gao, Xuying Chen, Piji Li, Zhaochun Ren, Lidong Bing, Dongyan Zhao, Rui Yan
 shengao@pku.edu.cn

北京大學 PEKING UNIVERSITY

Motivation

Model

Ablation study

Assignment	cF0ss		
RASG w/o DM	RASG w/o Denoising Module		
RASG w/o G	RASG w/o Gap content		
RASG w/o GT	RASG w/o Goal Tracker		
RASG w/o GTD	RASG w/o Goal Tracker Discriminator		
RASG w/o DM	27.29	11.01	24.64
RASG w/o G	28.03	11.24	25.28
RASG w/o GT	22.75	9.17	20.8
RASG w/o GTD	19.3	6.95	17.7
RASG	30.33	12.39	27.16

Human Evaluation

	Fluency	Consistency
S2S	2.17	1.98
CGU	2.2	2.08
RASG	2.65	2.48

Overall performance

	ROUGE-1	ROUGE-2	ROUGE-L
S2S	23.86	9.86	23.83
S2SR	24.7	10	24.5
CGU	27.32	11.36	25.49
RASG	30.33	12.39	27.16
LEAD1	5.51	1.71	4.94
TextRank	13.5	4.55	11.46

Overall performance

Cosine distance between decoding attention and reader attention.

Recall score of denoising module.

Table 1: Examples of the text summarization. The text in red denotes the focused aspect by the good summary, while the text in blue is described by the bad summary. The text with underline is the focused aspect by reader comments.

document	On August 28, according to a person familiar with the matter, <u>Toyota Motor Corporation will invest 500 million U.S. dollars into the Uber, a taxi service company, with a valuation of up to 72 billion U.S. dollars.</u> The investment will focus on driverless car technology. However, its development path is not smooth. In March of this year, a <u>Uber driverless car hit a woman and caused her death.</u> In last year, Softbank also invested into Uber with a valuation of \$48 billion.
comments	Toyota's investment in Uber is a wise choice. \$500 million investment is really a lot of money!
good summary	Toyota invests \$500 million into Uber with a valuation of \$72 billion
bad summary	An Uber driverless car hits a passerby to death

AAAI-19 Summarization Papers

Towards Personalized Review Summarization via User-aware Sequence Network

Junjie Li, Haoran Li, Chengqing Zong

National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing, China

Review: The hotel is right next to the airport (my room had a view of the runways) but the noise is pretty well dampened so that is not an issue at all. Very convenient to the airport obviously, but also the main highways. Room was clean and comfortable, no complaints there. The price is a little high, but it is ok for me.

Bob Summary: very quite room in a great location.

John Summary: expensive hotel near by airport.

Alice Summary: clean and comfortable rooms, i love !!!

Personalized review summarization is motivated by that different users are likely to generate different summaries for the same review, according to their own experiences, thoughts, or writing styles. It can:

- help users who read there reviews to choose products.
- help review owners to summary review.

#reviews	536,255
#users	19,400
#summaries	536,255
#words/review	154.79
#reviews/user	27.64
#words/summary	7.60

We take the title as the reference summary of the review. To filter meaningless titles, we propose three filters:

- Aspect-based filter
- Title length filter
- Compression ratio filter

AAAI-19 Summarization Papers

DeepChannel: Saliency Estimation by Contrastive Learning for Extractive Document Summarization

Motivation

- Existing summarization methods have following drawbacks:
 - End-to-end models: demands large training corpus, ...
 - Noisy Channel^[1, 2]: fails to capture semantics, ...
- DeepChannel combines neural network and the idea of channel model, carrying forward their advantages and overcoming their disadvantages.

DeepChannel

- DeepChannel aims to learn a saliency score $P(D|S)$, which models the conditional probability of expanding S to D , for any document-summary pair (D, S) via neural network (Figure 1).

Training

$$L = L_{\text{contrastive}} + \alpha L_{\text{penal}}$$
$$L_{\text{penal}} = -(\log P(D|S_{\text{good}}) - \log P(D|S_{\text{bad}}))$$

- Penalization loss encourages sharp attention distribution.

• Then we recursively extract the most salient sentences from original document to form an informative summary.

Jiaxin Shi^{1*}, Chen Liang^{1*}, Lei Hou¹, Juanzi Li¹, Zhiyuan Liu¹, Hanwang Zhang²
¹Tsinghua University, ²Nanyang Technological University

[DeepChannel: Saliency Estimation by Contrastive Learning for Extractive Document Summarization](#)

D : Rutgers University has banned fraternity and sorority house parties at its main campus in New Brunswick, New Jersey, for the rest of the spring semester after several alcohol-related problems this school year, including the death of a student.

S_1 : Rutgers University has banned fraternity and sorority house parties because of an alcohol-related accident that led to the death of a student.

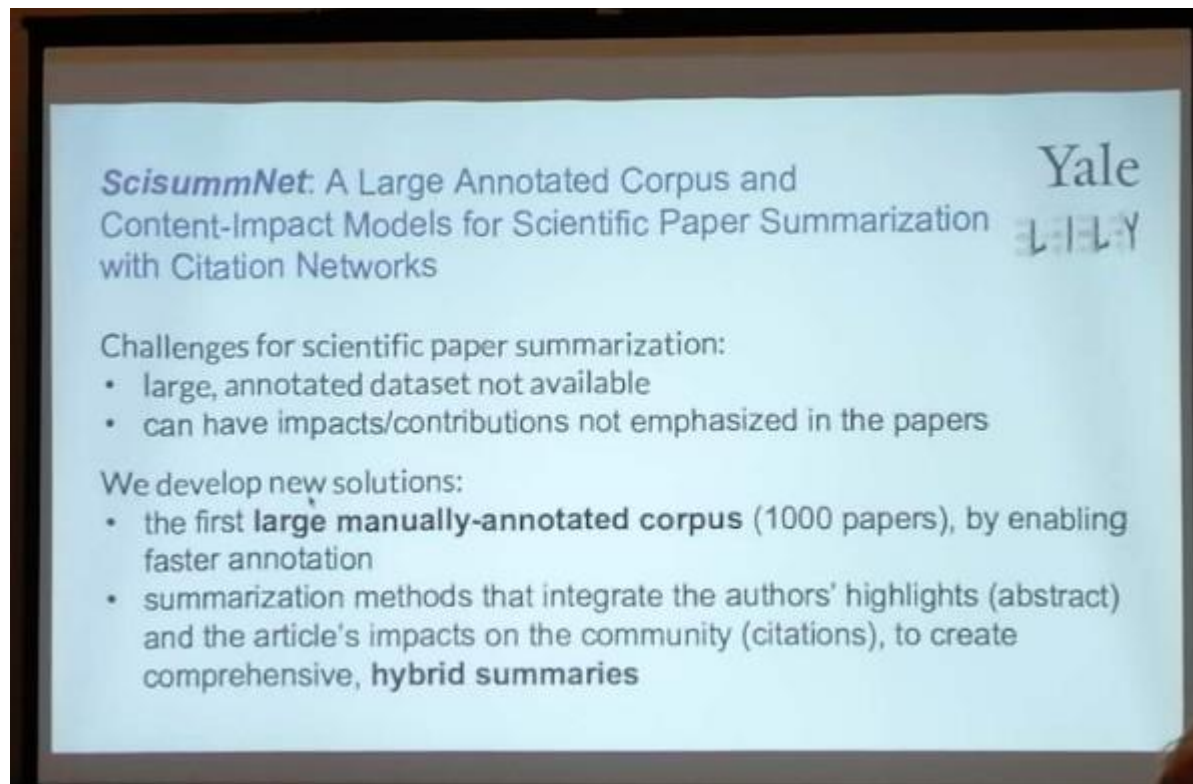
S_2 : The main campus of Rutgers University is located in New Brunswick, New Jersey.

Table 1: Examples of different degrees of saliency. We consider $P(D|S_1) > P(D|S_2)$ because S_1 contains more important information compared with S_2 and thus is more salient for yielding D .

$$P(S|D) \rightarrow P(D|S)$$

Better Summarization Better Reconstruction

AAAI-19 Summarization Papers



ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks

Yale

Challenges for scientific paper summarization:

- large, annotated dataset not available
- can have impacts/contributions not emphasized in the papers

We develop new solutions:

- the first **large manually-annotated corpus** (1000 papers), by enabling faster annotation
- summarization methods that integrate the authors' highlights (abstract) and the article's impacts on the community (citations), to create comprehensive, **hybrid summaries**

[ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks](#)

Paper ID: P06-1005

Paper Title: Bootstrapping Path-Based Pronoun Resolution

Abstract:

We present an approach to pronoun resolution based on syntactic paths. Through a simple bootstrapping procedure, we learn the likelihood of coreference between a pronoun and a candidate noun based on the path in the parse tree between the two entities. This path information enables us to handle previously challenging resolution instances, and also robustly addresses traditional syntactic coreference constraints. Highly coreferent paths also allow mining of precise probabilistic gender/number information. We combine statistical knowledge with well known features in a Support Vector Machine pronoun resolution classifier. Significant gains in performance are observed on several datasets. **(mostly about technique)**

Citation Sentences:

Bergsma and Lin (2006) determine the like-lihood of coreference along the syntactic path connecting a pronoun to a possible antecedent, by looking at the distribution of the path in text. **(about technique)**

We use the approach of Bergsma and Lin (2006), both because it achieves state-of-the-art gender classification performance, and because a database of the obtained noun genders is available online. **(about both technique and dataset)**

For the gender task that we study in our experiments, we acquire class instances by filtering the dataset of nouns and their genders created by Bergsma and Lin (2006). **(about dataset)**

Figure 1: Abstract and citations of (Bergsma and Lin 2006). The abstract emphasizes their pronoun resolution techniques and improved performance; the citation sentences reveal that their noun gender dataset is also a major contribution to the research community, but it is not covered in the abstract.

AAAI-19 Transformer Papers

Gaussian Transformer: a Lightweight Approach for Natural Language Inference

IR
Gaussian Transformer

Embedding Block: tokens \rightarrow vectors

Encoding Block: capture the local & global structure of sequences

Interaction Block: align & infer the two sentences multiple times

Comparison Block: compare them & make the final prediction

The diagram illustrates the architecture of the Gaussian Transformer. It starts with an **Embedding Block** that takes a **Passage** and outputs **Word & Char Embeddings**. These are processed by **M** **Encoding Blocks**, each containing **Add & Norm**, **Feed Forward**, **Self-attention**, and **Gaussian Prior** layers. The output is then processed by **N** **Interaction Blocks**, which include **Add & Norm**, **Inter-attention**, **Add & Norm**, **Self-attention**, and **Gaussian Prior** layers. The final output goes through a **Comparison Block** with **Scaled Sum**, **Feed Forward**, and **Concatenate** layers, leading to **Output Probabilities**.

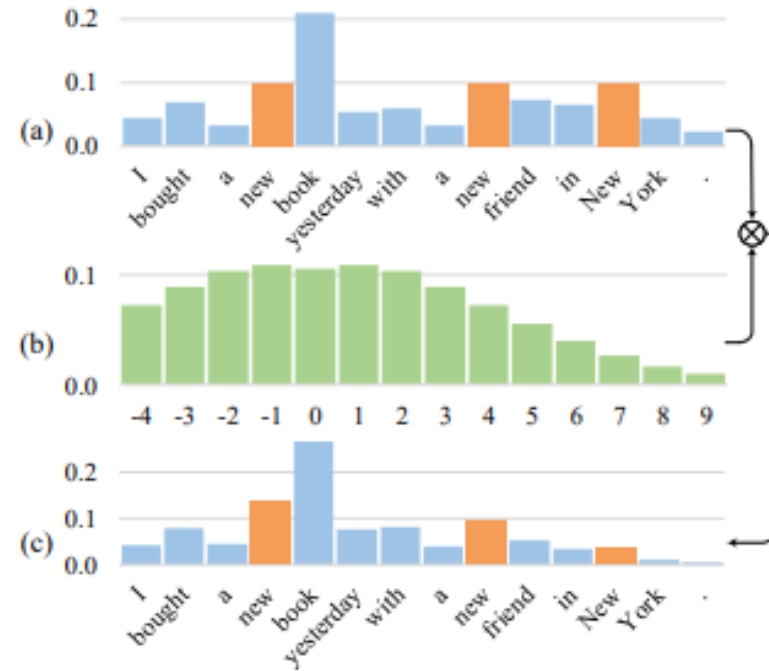


Figure 1: Probabilities of each token attending to current central word 'book': (a) illustrates the vanilla self-attention, where the word 'new' appeared in different positions obtain the same importance, which is inconsistent with our experience that adjacent words matters; (b) depicts a Gaussian distribution over distance (x-axis) that encourages focusing on neighboring tokens; (c) draws the attention corrected by the Gaussian prior, where the first 'new' is more important.

AAAI-19 Transformer Papers

Hierarchical Attention Networks for Sentence Ordering

Tianming Wang and Xiaojun Wan
 Institute of Computer Science and Technology, Peking University
 The MOE Key Laboratory of Computational Linguistics, Peking University
 (wangtm, wanxiaojun}@pku.edu.cn

Task Description

Sentence ordering is a task of organizing a set of sentences into a coherent text.

Formally, given an out-of-order set of N sentences $\{s_1, s_2, \dots, s_N\}$, our goal is to find the gold order $O = (o_1, o_2, \dots, o_N)$ for these sentences, which maximize the probability:

$$\sum_{i=1}^N \log p(s_{o_i} | s_{o_1}, s_{o_2}, \dots, s_{o_{i-1}})$$

Dataset

arXiv: abstract of papers on arXiv website.
VIST: visual storytelling dataset. Only story text is used.
ROCStory: commonsense story dataset.

Hierarchical Attention

we propose a novel hierarchical attention network which captures word clues and dependencies between sentences to address the task.

Word Attention
 Word clues play an important role in sentence ordering. We employ a multi-head attention layer over the word-level LSTM encoder to let the model pay more attention to word clues.

Sentence Attention
 We use self-attention mechanism to capture the dependencies between sentences and adjust their representations.

Experiments

Evaluation Metric: Kendall's τ (a metric of rank correlation) and Perfect Match Ratio (the ratio of cases of exact match). We also test the robustness of our model by adding noisy sentence into the given set. Our model can achieve very high accuracy for discriminating noise.

Methods	arXiv		VIST		ROCStory	
	τ	PMR	τ	PMR	τ	PMR
UITS (Mansour, Sun, Qiu, and Huang 2016)	0.6994	0.5241	-	-	-	-
Multi-Head Attention (Nguyen et al. 2016)	0.7578	0.6004	0.6842	0.7214	-	-
Seq2Seq (Sutskever, Vinyals, and Le 2014)	0.6802	0.5100	0.6862	0.7200	0.5400	0.6700
LSTM (Klein, Shih, and Manning 2016)	0.7201	0.5817	0.6889	0.7200	0.7111	0.7064
Word2Vec (Mikolov et al. 2013)	0.7100	0.6210	0.6820	0.7100	0.7020	0.7020
Proposed	0.7506	0.6485	0.6902	0.7300	0.7320	0.7362

Strategy	Methods	arXiv	VIST	ROCStory
		acc	acc	acc
I noise	random	0.1819	0.1667	0.1667
	Our	0.9664	0.8462	0.9082
O/I noise	random	0.2855	0.5833	0.5833
	Our	0.9370	0.9151	0.9008

Visualization

Using transformer to encode a **set** of sentences ---- removing the position encoding

AAAI-19 Transformer Papers

- [Neural Speech Synthesis with Transformer Network](#)

Naihan Li^{1,4}, Shujie Liu², Yanqing Liu³, Sheng Zhao³, Ming Liu^{1,4}, Ming Zhou²

¹University of Electronic Science and Technology of China

²Microsoft Research Asia

³Microsoft STC Asia

⁴CETC Big Data Research Institute Co.,Ltd, Guizhou, China

- [Tied Transformers: Neural Machine Translation with Shared Encoder and Decoder](#)

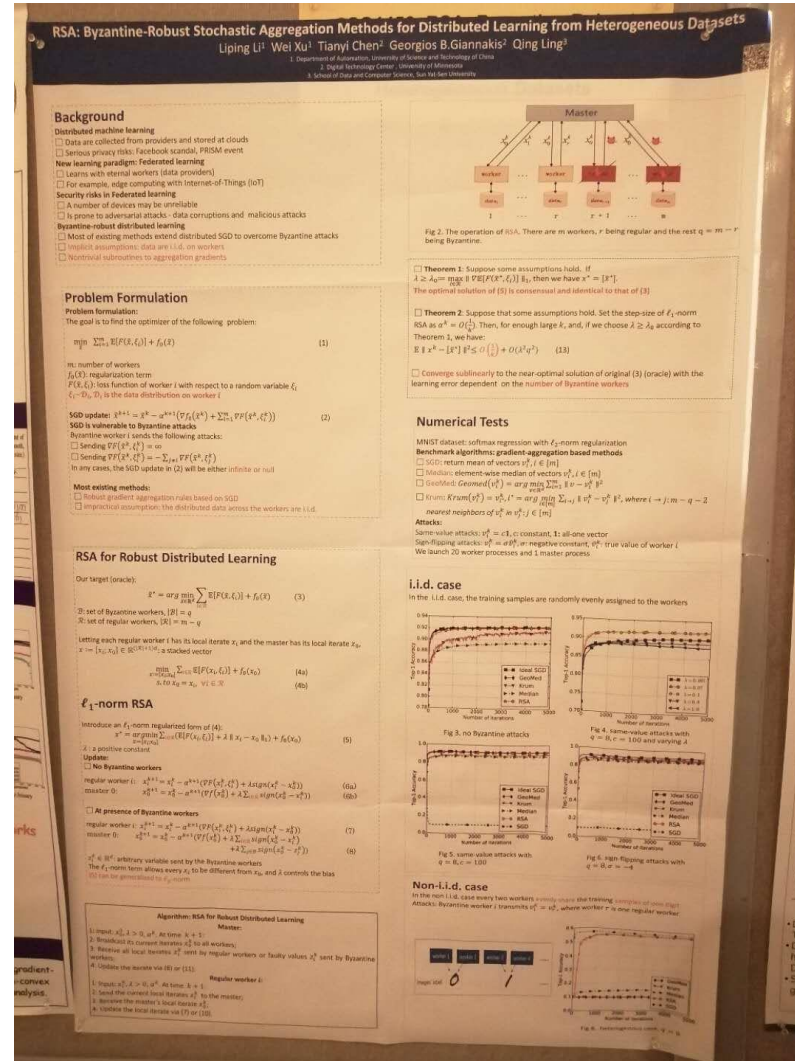
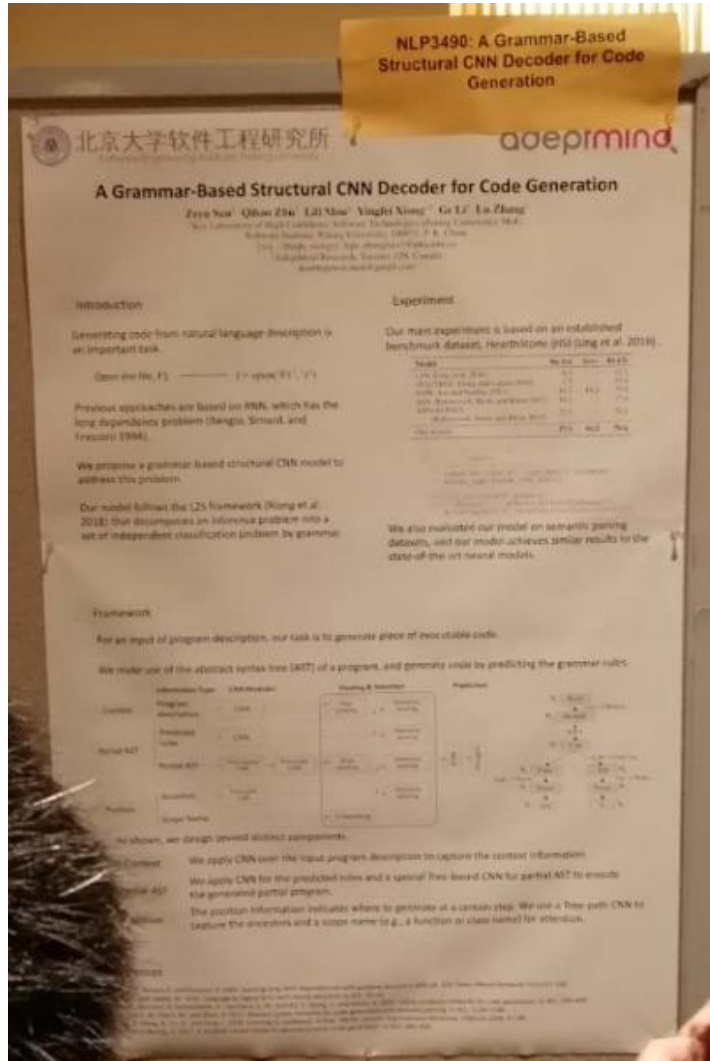
¹Yingce Xia, ²Tianyu He, ¹Xu Tan, ¹Fei Tian, ³Di He and ¹Tao Qin

¹ Microsoft Research, Beijing, China

²University of Science and Technology of China, Anhui, China

³Key Laboratory of Machine Perception, MOE, School of EECS, Peking University
{yinxia,xuta,fetia,taoqin}@microsoft.com; hetianyu@mail.ustc.edu.cn; di_he@pku.edu.cn

AAAI-19 Other Papers



AAAI-19 Internship

- Singapore, Research Institute (similar to MSRA)



Steven Hoi
Contact: Wechat

?

Report to



Richard Socher

- Tokyo, ...
- Amazon, Tong Zhao, ...

AAAI-19 游玩篇



Diamond Head



Pearl Harbor



Hanauma Bay



... Beach