GTC China 2020

Recent Advances in Machine Writing and Translation – Algorithms and Challenges

Lei Li
ByteDance AI Lab

Volctrans

12/19/2020

Revolution in Information Creation and Sharing

- New media platforms

- Tremendous improvement in the efficiency and quality of content creation
- Massive distribution of personalized information

Why is NLG important?

Machine Translation

Machine Writing

ChatBOT

Question Answering

Machine Translation has quietly increased international trade by over 10%

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 65, No. 12, December 2019, pp. 5449–5460
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform

Erik Brynjolfsson,[a] Xiang Hui,[b] Meng Liu[b]

[a] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [b] Marketing, Olin School of Business, Washington University in St. Louis, St. Louis, Missouri 63130

Contact: erikb@mit.edu, http://orcid.org/0000-0002-8031-6990 (EB); hui@wustl.edu, http://orcid.org/0000-0001-7595-3461 (XH); mengl@wustl.edu, http://orcid.org/0000-0002-5512-7952 (ML)

Received: April 18, 2019
Revised: April 18, 2019
Accepted: April 18, 2019
Published Online in Articles in Advance: September 3, 2019

https://doi.org/10.1287/mnsc.2019.3388

Abstract. Artificial intelligence (AI) is surpassing human performance in a growing number of domains. However, there is limited evidence of its economic effects. Using data from a digital platform, we study a key application of AI: machine translation. We find that the introduction of a new machine translation system has significantly increased international trade on this platform, increasing exports by 10.9%. Furthermore, heterogeneous treatment effects are consistent with a substantial reduction in translation costs. Our results provide causal evidence that language barriers significantly hinder trade and that AI has already begun to improve economic efficiency in at least one domain.

History: Accepted by Joshua Gans, business strategy.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2019.3388.

Keywords: artificial intelligence • international trade • machine translation • machine learning • digital platforms

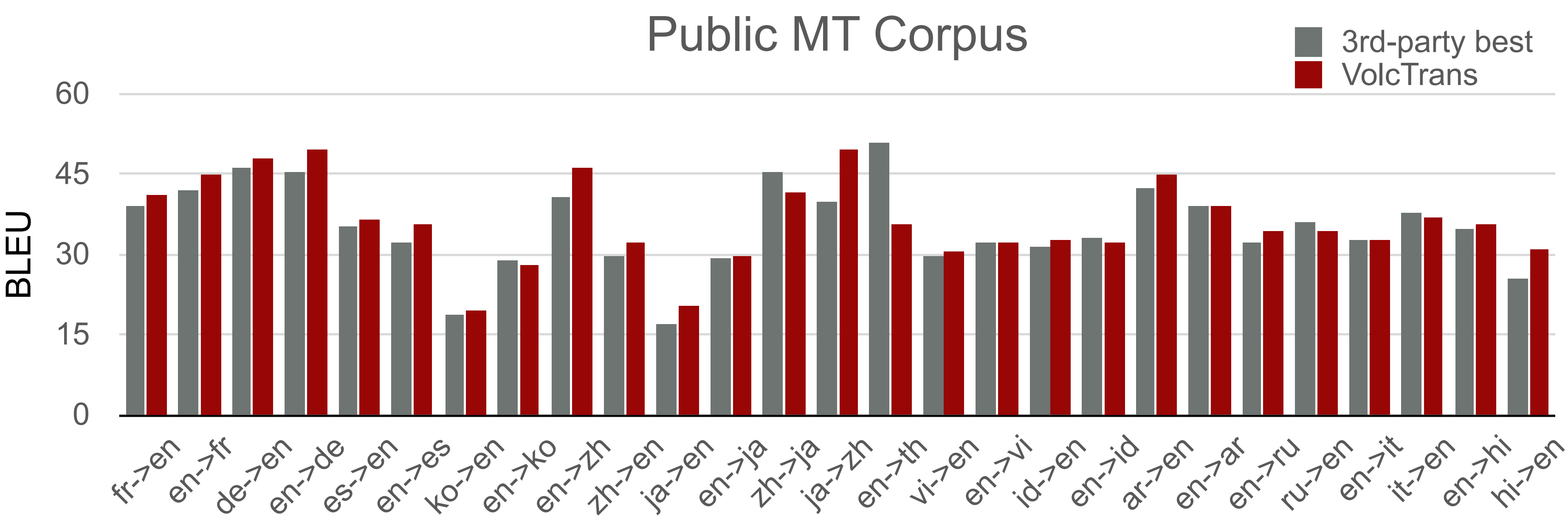
Machine Translation at ByteDance

50+
Clients

50+
languages

Five champions
in WMT 20
including
Chinese-to-English
German-to-English
German-to-French

Volctrans
translate.volcengine.cn



Simultaneous Speech-to-Text Translation

The New York Times

Soon a Robot Will Be Writing This Headline

Gabriel Alcala

By Alana Semuels

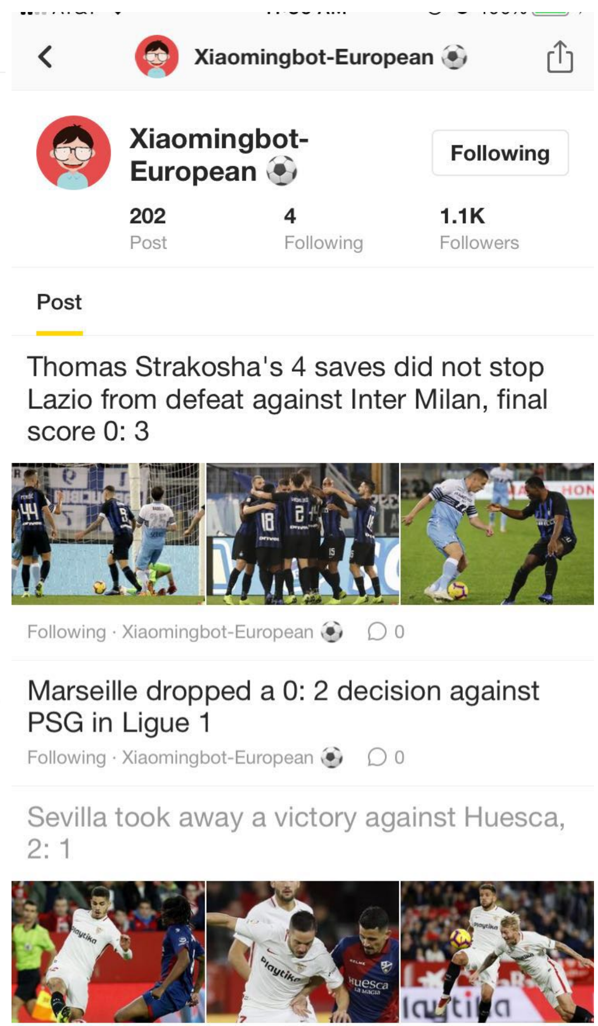
Jan. 14, 2020

Xiaomingbot
Automatic News Writing System

Winning 2017 Wu Wen-tsün Award in AI from CAAI

明くんのW杯
(Japanese)

Beto Bot Copa2018
(Portuguese)



600,000 articles

6 lang

150,000 followers

Xiaomingbot : Multilingual Robot News Reporter

Runxin Xu, Jun Cao, Mingxuan Wang, Jiaze Chen, Hao Zhou, Ying Zeng, Yuping Wang, Li Chen, Xiang Yin, Xijin Zhang, Songcheng Jiang, Yuxuan Wang, Lei Li, ACL 2020.

Snooker Commentary Generation
Combining Visual Understanding with Strategy Prediction

Balls Detection

Balls' Positions at the Beginning

Red0: (180, 542)
Red1: (189, 552)
Red2: (179, 555)
Red3: (184, 561)
Red4: (202, 563)
Red5: (174, 564)
Red6: (189, 569)
Red7:
Red11:(197, 590)
Red12:(241, 595)
Red13:(155, 606)
Red14:(327, 611)
Brown: (183, 163)
Green: (240, 163)
Yellow: (127, 163)
Blue: (183, 366

(positions after mapping)

GraspSnooker [Z. Sun, J. Chen, H. Zhou, D. Zhou, **Lei Li**, M. Jiang, IJCAI19b]

Outline

1. Sequence Generation Problem
2. Deep Latent Variable Models for Text Generation
3. Monte-Carlo Methods for Constrained Text Generation
4. One model to acquire 4 language skills
   – Mirror Generative NMT [ICLR 20a]
5. mRASP: Multilingual Pretraining NMT
6. Summary

Modeling a Sequence

The quick brown fox jumps over the lazy dog .

$$x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$$

The central problem of *language modeling* is to find the *joint probability distribution*:

$$p_\theta(x) = p_\theta(x_1, \cdots, x_L)$$

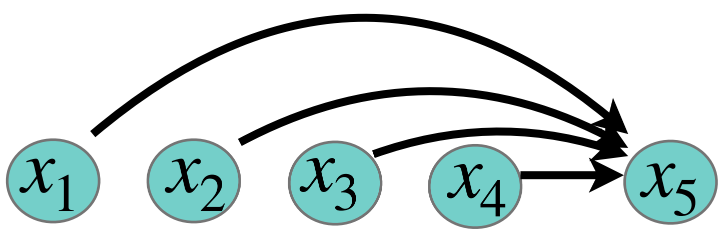
There are many ways to represent and learn the joint probability model.

Auto-Regressive Language Model

Decompose the joint distribution as a product of tractable conditional probabilities:

Given $x = [x_1, x_2, x_3 \ldots, x_n]$

$$p_\theta = \prod_{i=1}^{n} p_\theta(x_i \mid x_1, x_2, \ldots, x_{i-1}) = \prod_{i=1}^{n} p_\theta(x_i \mid x_{<i})$$



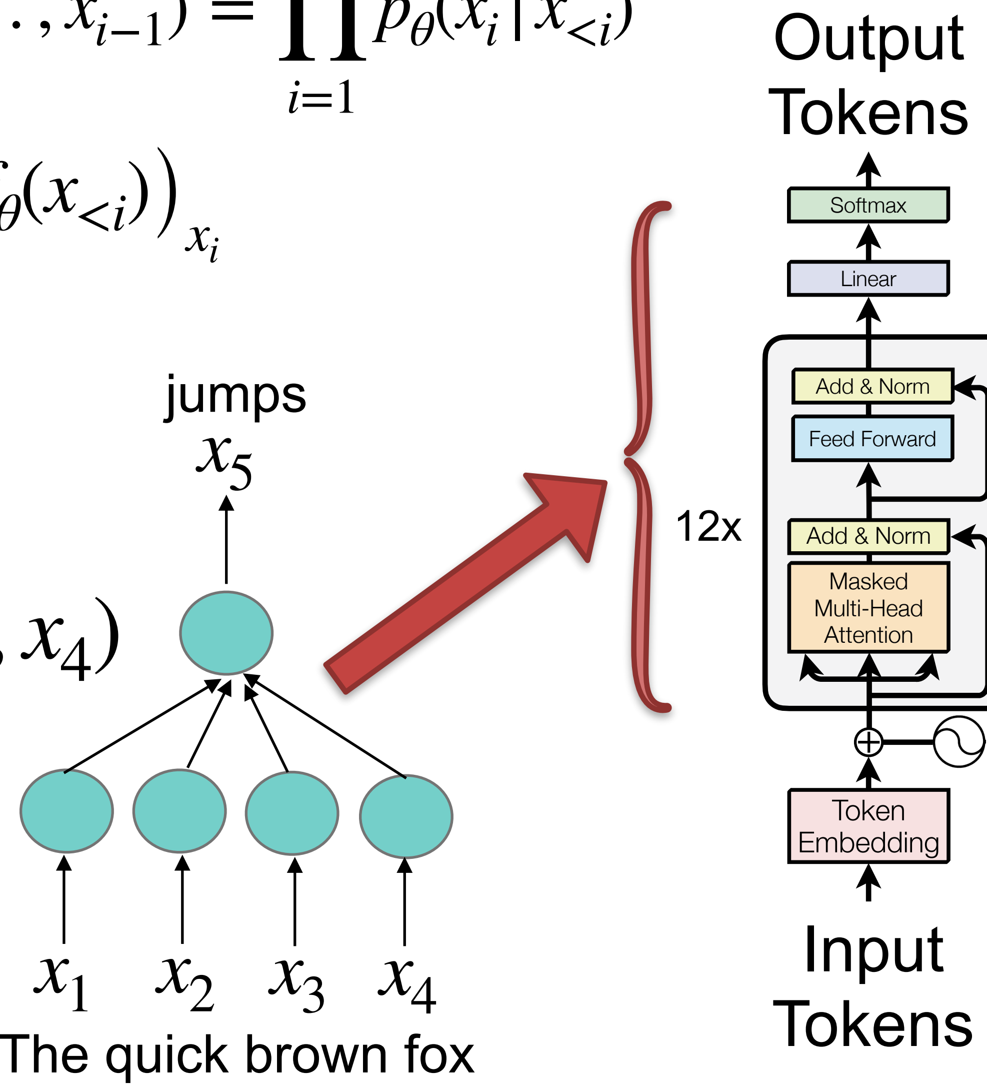
Auto-Regressive Factorization - Token Probability from a Neural Network

$$p_\theta = \prod_{i=1}^{n} p_\theta(x_i \,|\, x_1, x_2, \ldots, x_{i-1}) = \prod_{i=1}^{n} p_\theta(x_i \,|\, x_{<i})$$

$$p_\theta(x_i \,|\, x_{<i}) = \mathrm{Softmax}\left(f_\theta(x_{<i})\right)_{x_i}$$

$$\mathrm{Softmax}(\mathrm{x})_j = \frac{\exp x_j}{\sum_k \exp x_k}$$

$$p_\theta(x_5 \,|\, x_1, x_2, x_3, x_4)$$



Conditional Sequence Generation

aka. sequence-to-sequence generation

- Machine Translation
- Dialog Generation
- Question Answering
- …

敏捷的棕狐跳过懒狗

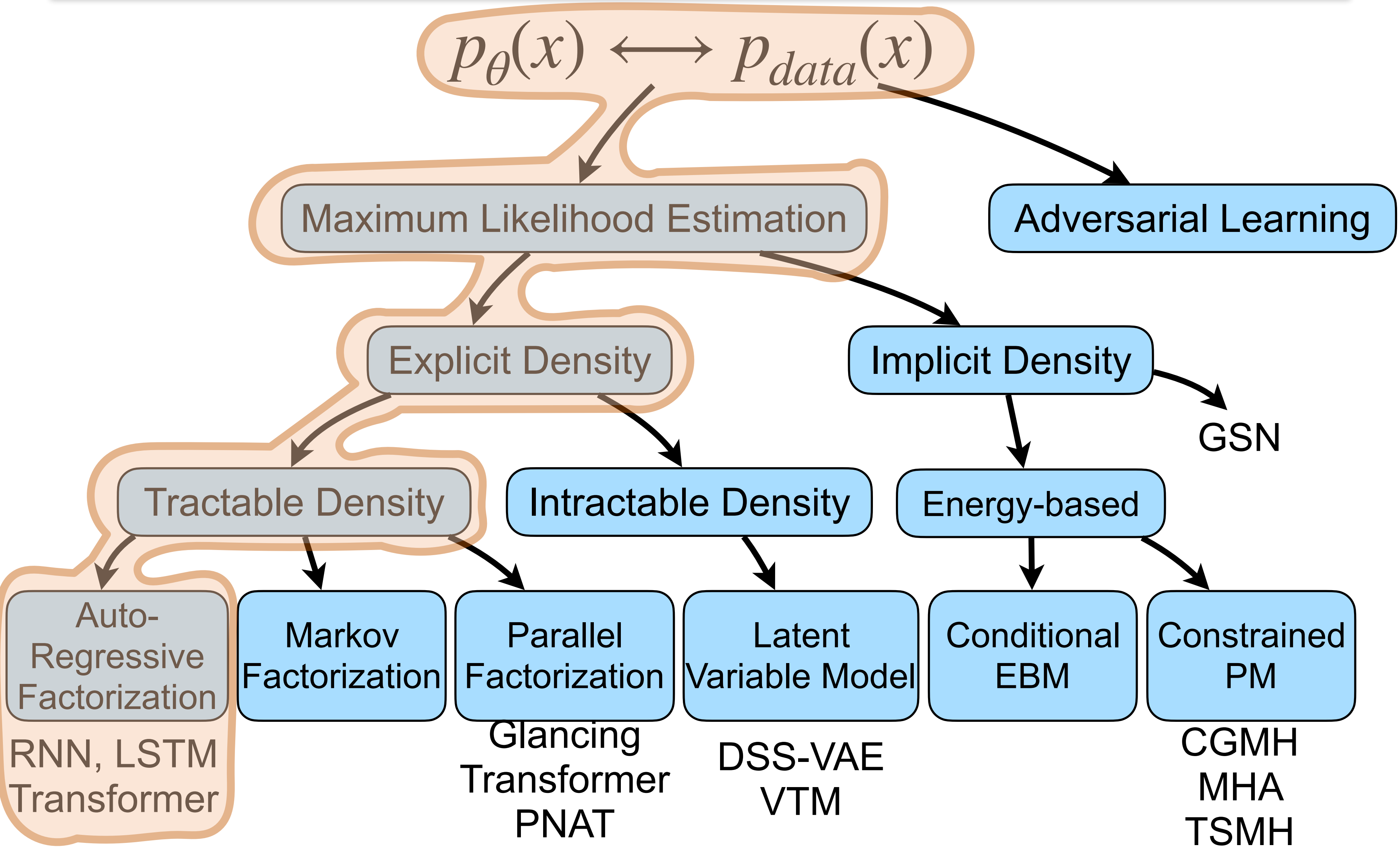
Output

$$p_\theta(\boxed{y} \mid \boxed{x})$$

Input

The quick brown fox jumps over the lazy dog .

DGM Taxonomy



Outline

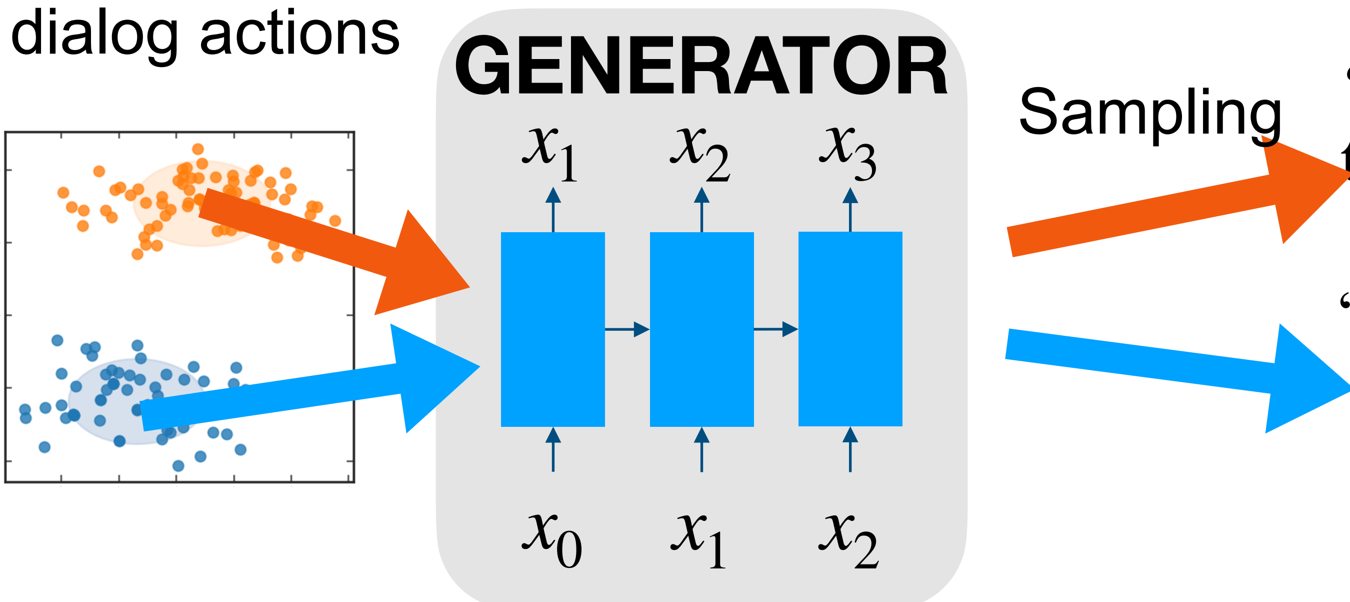
1. Sequence Generation Problem
2. Deep Latent Variable Models for Text Generation
3. Monte-Carlo Methods for Constrained Text Generation
4. One model to acquire 4 language skills
   – Mirror Generative NMT [ICLR 20a]
5. mRASP: Multilingual Pretraining NMT
6. Summary

Deep Latent Variable Models for Text

- Interpretable Deep Latent Representation from Raw Text
  - Learning Exponential Family Mixture VAE [ICML 20]
- Disentangled Representation Learning for Text Generation
  - Data to Generation: VTM [ICLR 20b]
  - Learning syntax-semantic representation [ACL 19c]

Learning Interpretable Latent Representation

Latent structure dialog actions



"Remind me about the football game."
[action=remind]

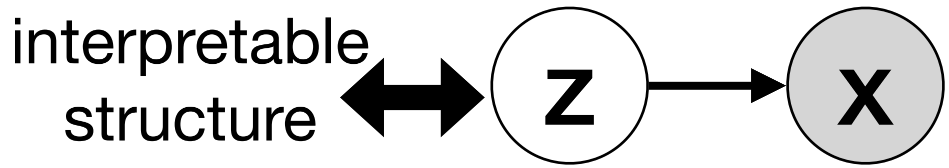
"Will it be overcast tomorrow?"
[action=request]

……

Generate Sentences with interpretable factors

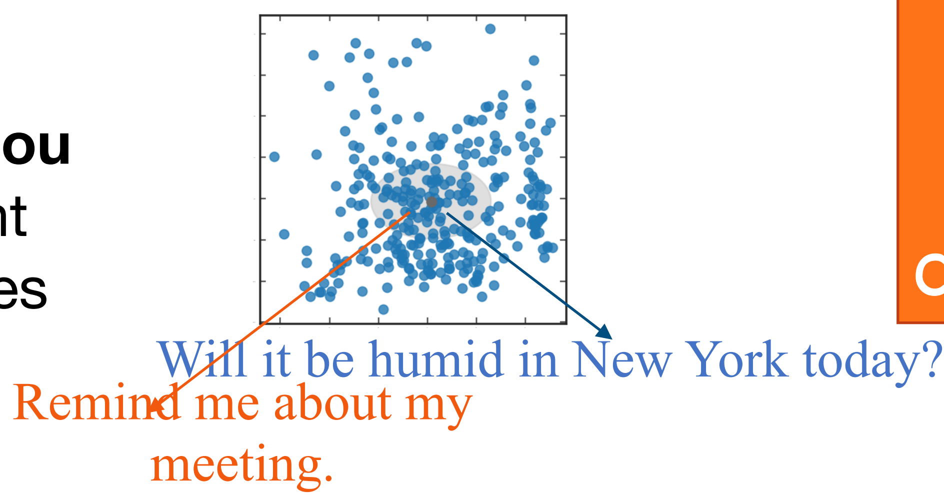
How to Interpret Latent Variables in VAEs?

Variational Auto-encoder (VAE)



(Kingma & Welling, 2013)

z: **continuous** latent variables

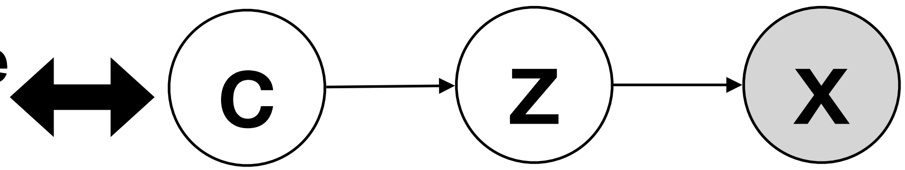


difficult to interpret discrete factors

Discrete Variables Could Enhance Interpretability - but one has to do it right!

Gaussian Mixture Variational Auto-encoder (GM-VAE)

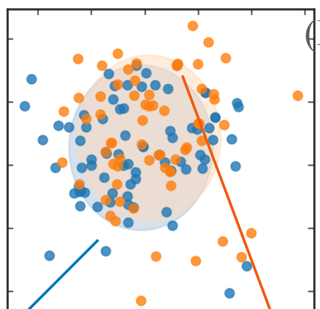
interpretable structure



(Dilokthanakul et al., 2016; Jiang et al., 2017)

c: **discrete** component

z: **continuous** latent variable



Will it be overcast tomorrow?

Remind me about the football game.

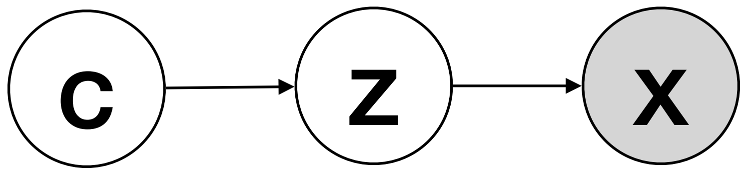
Why?

How to fix it?

mode-collapse

Do it right for VAE w/ hierarchical priors - Dispersed Exponential-family Mixture VAE

Exponential-family Mixture VAE



adding dispersion term in training

Dispersed EM-VAE

$$L(\theta; x) = \text{ELBO} + \beta \cdot \boxed{L_d,}$$

dispersion term

$$L_d = \mathbb{E}_{q_\phi(c|x)} A(\boldsymbol{\eta}_c) - A(\mathbb{E}_{q_\phi(c|x)} \boldsymbol{\eta}_c) \,.$$

DEM-VAE [W. Shi, H. Zhou, N. Miao, **Lei Li**, ICML 2020]

Latent Variables Learned by DEM-VAE are Semantically Meaningful

Example actions and corresponding utterances (classified by $q_\phi(c\,|\,x)$)

Inferred action=Inform-route/address
"There is a Safeway 4 miles away."
"There are no hospitals within 2 miles."
"There is Jing Jing and PF Changs."
…

Inferred action =Request-weather
"What is the weather today?"
"What is the weather like in the city?"
"What's the weather forecast in New York?"
…

Utterances of the same actions could be assigned with the same discrete latent variable c.

DEM-VAE [W. Shi, H. Zhou, N. Miao, **Lei Li**, ICML 2020]

Generate Sensible Dialog Response with DEM-VAE

Input Context
Sys: "Taking you to Chevron."

sampling different values of discrete latent variables

(action = thanks)

(action = request-address)

Predict
User: "Thank you car, let's go there!"

Predict
User: "What is the address?"

Responses with different actions are generated by sampling different values of discrete latent variables.

DEM-VAE [W. Shi, H. Zhou, N. Miao, **Lei Li**, ICML 2020]

Data-to-Text Generation

name	Sukiyaki
eatType	pub
food	Japanese
price	average
rating	good
area	seattle

Sukiyaki is a Japanese restaurant. It is a pub and it has a average cost and good rating. It is based in seattle.

Previous Idea: Templates

[name] is a [food] restaurant. It is a [eatType] and it has a [price] cost and [rating] rating. It is in [area].

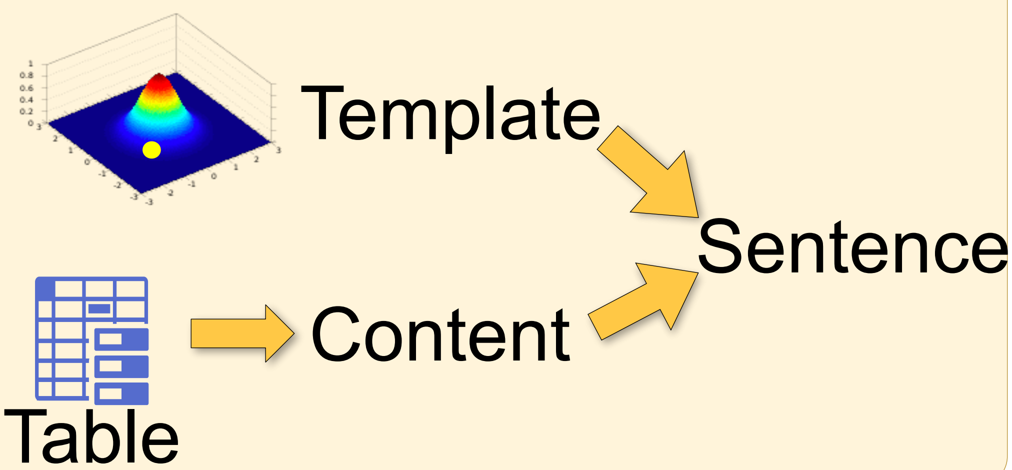
name	Sukiyaki
eatType	pub
food	Japanese
price	average
rating	good
area	seattle

Sukiyaki is a Japanese restaurant. It is a pub and it has a average cost and good rating. It is in seattle.

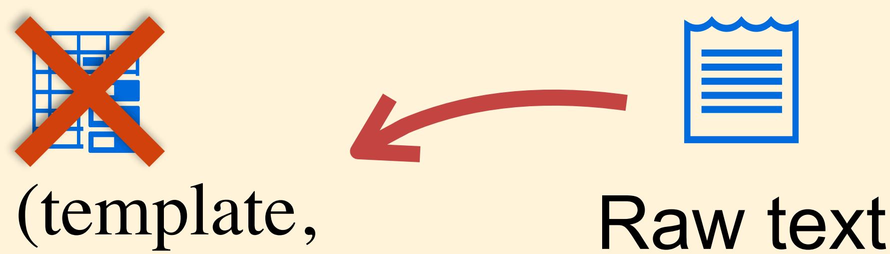
But manually creation of templates are tedious

Generating from Latent Factors

Motivation 1:
Continuous and disentangled representation for template and content



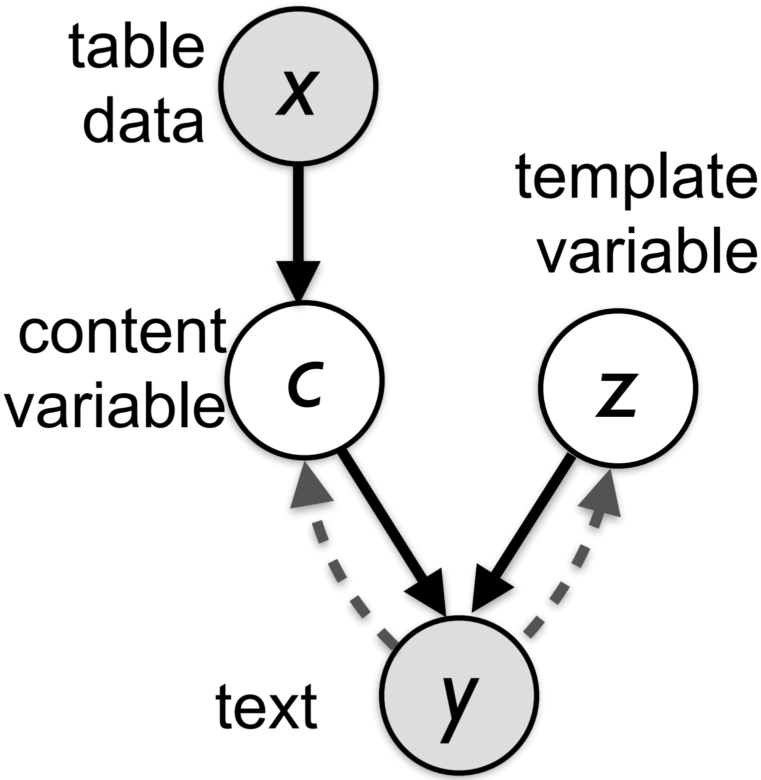
Motivation 2:
Incorporate raw text corpus to learn good representation.



q (template, content | sentence)

VTM [R. Ye, W. Shi, H. Zhou, Z. Wei, **Lei Li**, ICLR20b]

Variational Template Machine



Input: triples of <field_name, position, value>

$$\{x_k^f, x_k^p, x_k^v\}_{k=1}^K$$

1. $p(c \mid x) \sim$ Neural Net

$$\mathrm{maxpool}\Big(\tanh(W \cdot [x_f^k, x_p^k, x_v^k] + b)\Big)$$

2. Sample $z \sim p_0(z)$, e.g. Gaussian
3. Decode y from $[c, z]$ using another NN (e.g. Transformer)

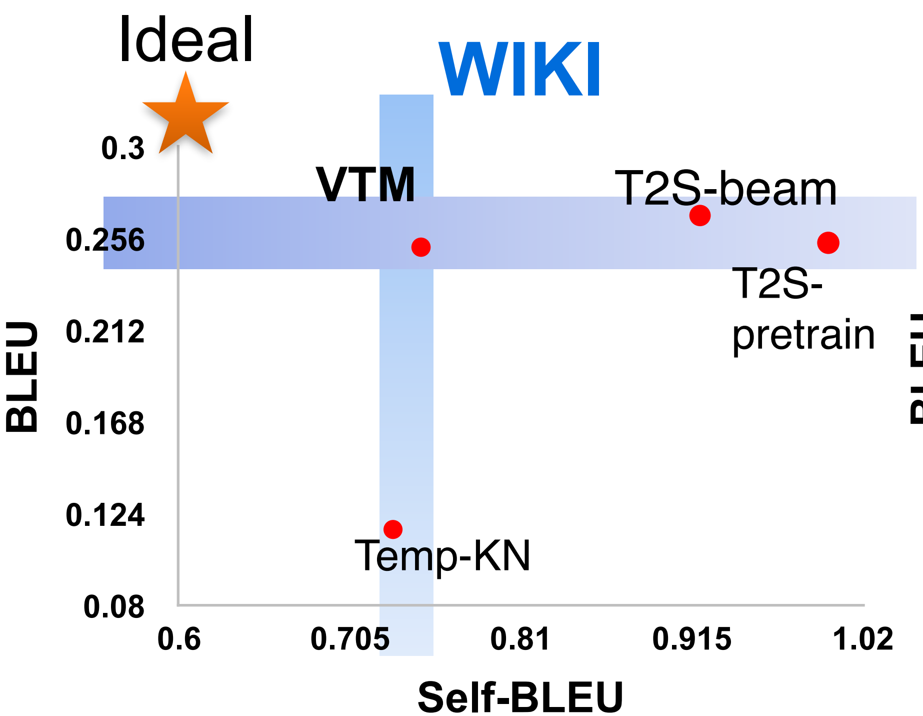
VTM [R. Ye, W. Shi, H. Zhou, Z. Wei, **Lei Li**, ICLR20b]

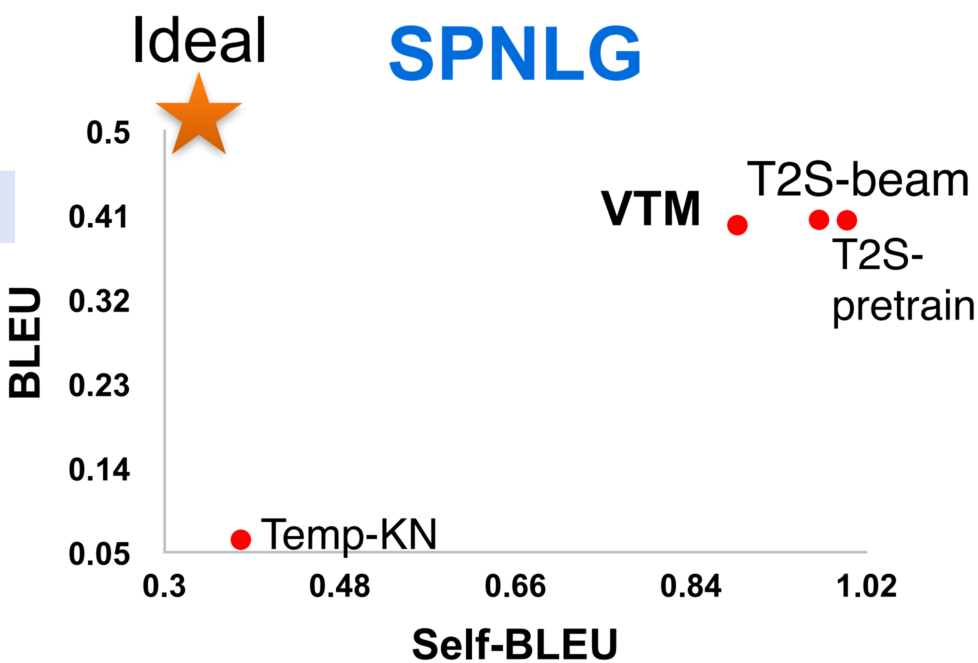
Learning with Raw Corpus

- Semi-supervised learning: “Back-translate” corpus to obtain pseudo-parallel pairs <table, text>, to enrich the learning

<table>
<tr><th colspan="2">Table</th><th>Text</th></tr>
<tr><td>name</td><td>Sukiyaki</td><td rowspan="6">Sukiyaki is a Japanese restaurant. It is a pub and it has a average cost and good rating. It is in seattle.</td></tr>
<tr><td>eatType</td><td>pub</td></tr>
<tr><td>food</td><td>Japanese</td></tr>
<tr><td>price</td><td>average</td></tr>
<tr><td>rating</td><td>good</td></tr>
<tr><td>area</td><td>seattle</td></tr>
<tr><td colspan="2">?
$q(<c,z> | y)$</td><td>Known for its creative flavours, Holycrab's signatures are the Hokkien crab.</td></tr>
</table>

VTM Produces High-quality and Diverse Text

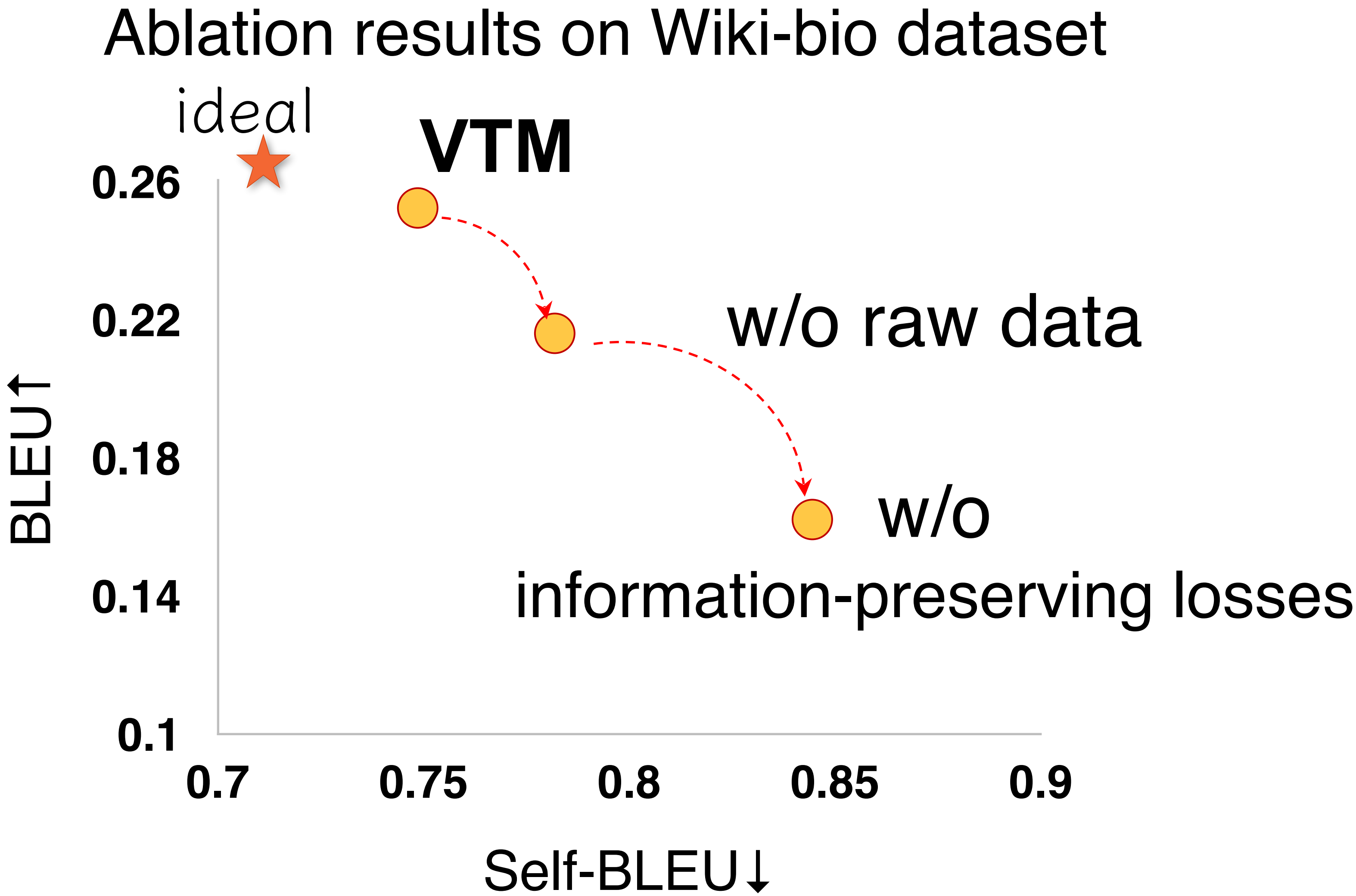




VTM uses beam-search decoding.

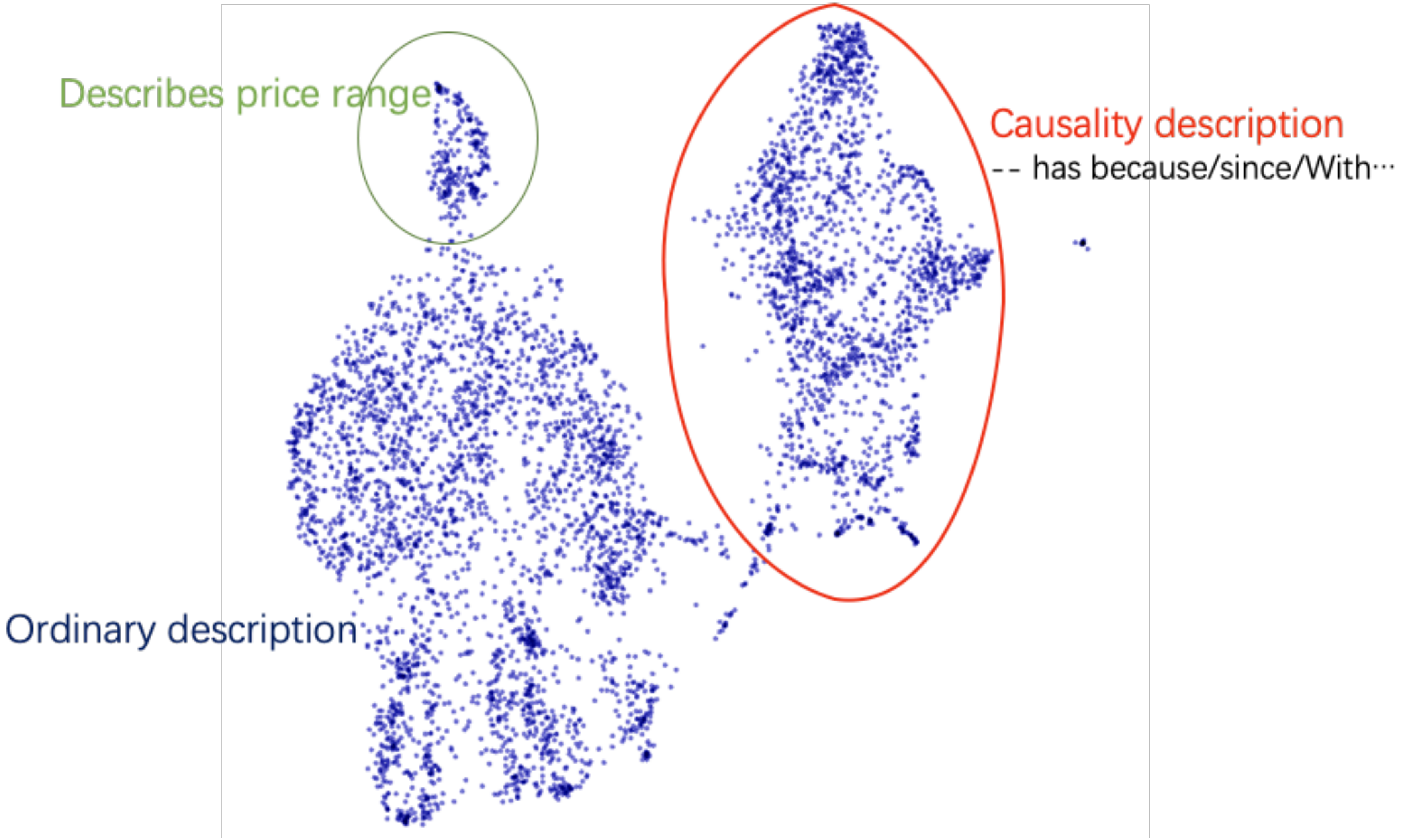
VTM [Ye, …, **Lei Li**, ICLR20b]

Raw data and loss terms are necessary



Interpreting VTM

Template variable project to 2D



VTM Generates Diverse Text

Input Data Table

Jack Ryder

Ryder in about 1930

Personal information	
Full name	John Ryder
Born	8 August 1889 Collingwood, Victoria, Australia
Died	3 April 1977 (aged 87) Fitzroy, Victoria, Australia
Nickname	The King of Collingwood
Height	1.85 m (6 ft 1 in)
Batting	Right-handed
Bowling	Right-arm medium pace
Role	All-rounder

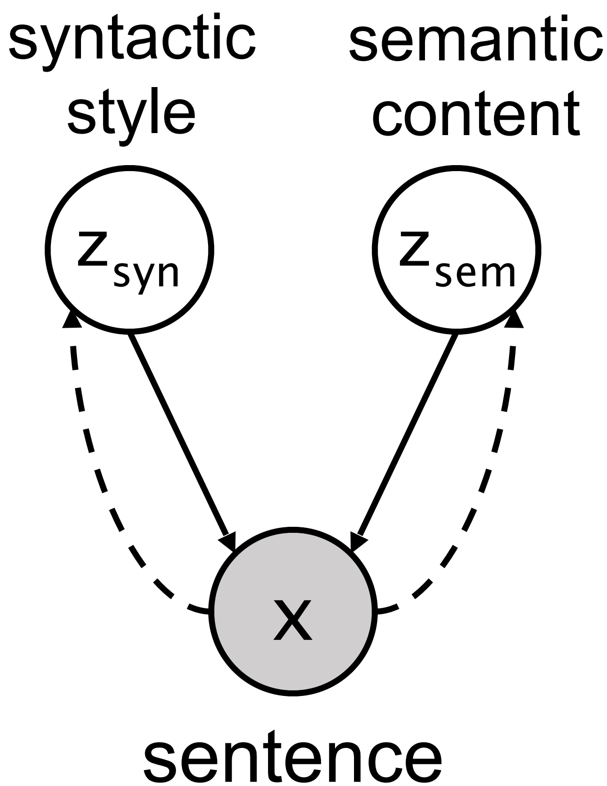
Generated Text

1: John Ryder (8 August 1889 – 4 April 1977) was an Australian cricketer.

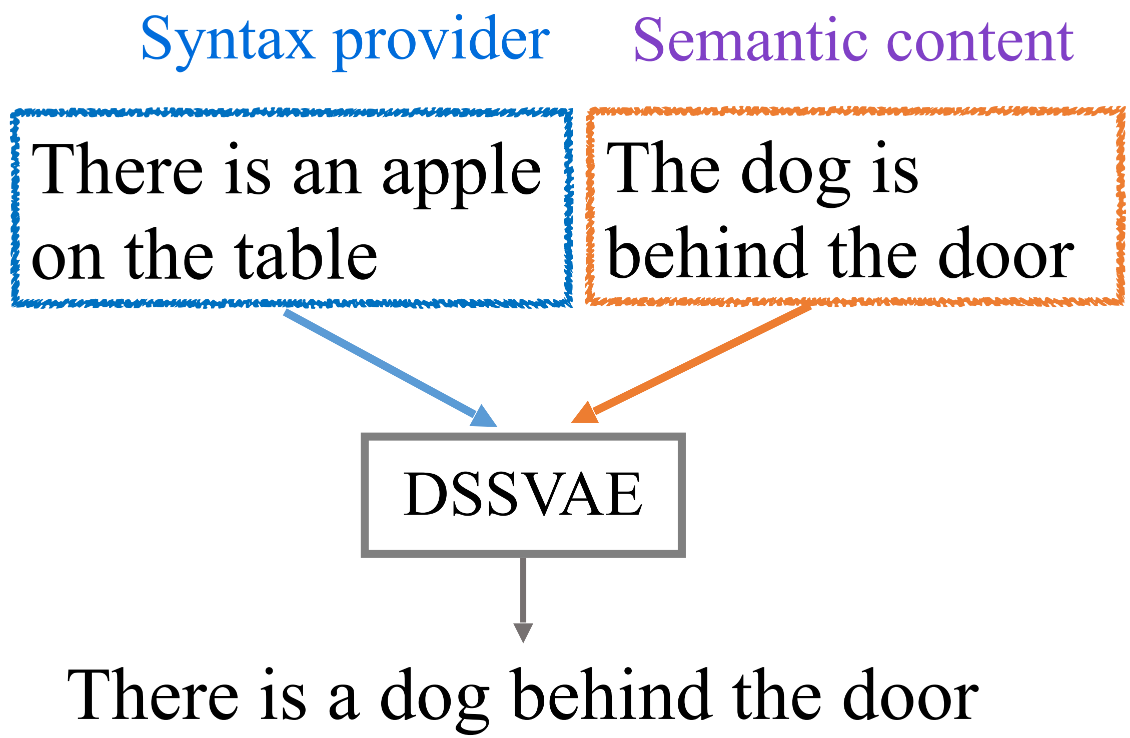
2: Jack Ryder (born August 9, 1889 in Victoria, Australia) was an Australian cricketer.

3: John Ryder, also known as the king of Collingwood (8 August 1889 – 4 April 1977) was an Australian cricketer.

Learning Disentangled Representation of Syntax and Semantics



DSSVAE enables learning and transferring sentence-writing styles



DSS-VAE [Y. Bao, H. Zhou, S. Huang, **Lei Li**, L. Mou, O. Vechtomova, X. Dai, J. Chen, ACL19c]

Outline

1. Sequence Generation Problem
2. Deep Latent Variable Models for Text Generation
3. Monte-Carlo Methods for Constrained Text Generation
4. One model to acquire 4 language skills
   – Mirror Generative NMT [ICLR 20a]
5. mRASP: Multilingual Pretraining NMT
6. Summary

Constrained Text Generation

To generate sentences that are:

- Fluent
- Constraint-satisfying
  - e.g. keyword-occurrence constraint

“Autumn”
“Sports shoes”

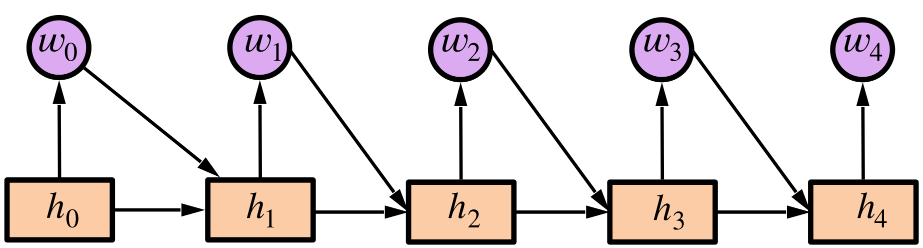
Comfortable **sports shoes**, a breathing pair of man's shoes, accompanying you in **autumn**

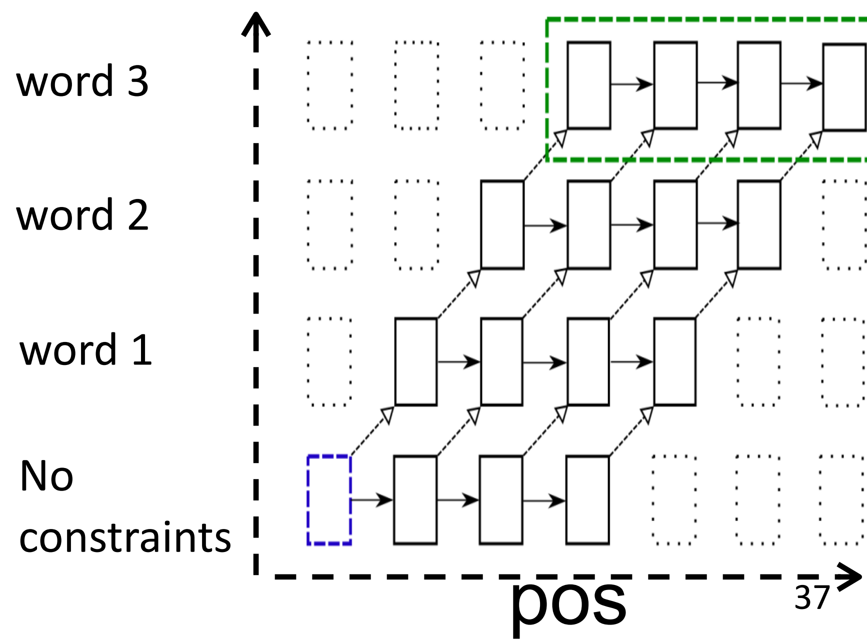
Why is Constrained Text Generation difficult?

Exponential search space, $O((N\text{-}k)^V)$

RNN grid beam search [Hokamp & Liu 2017]

does not usually produce high quality sentences





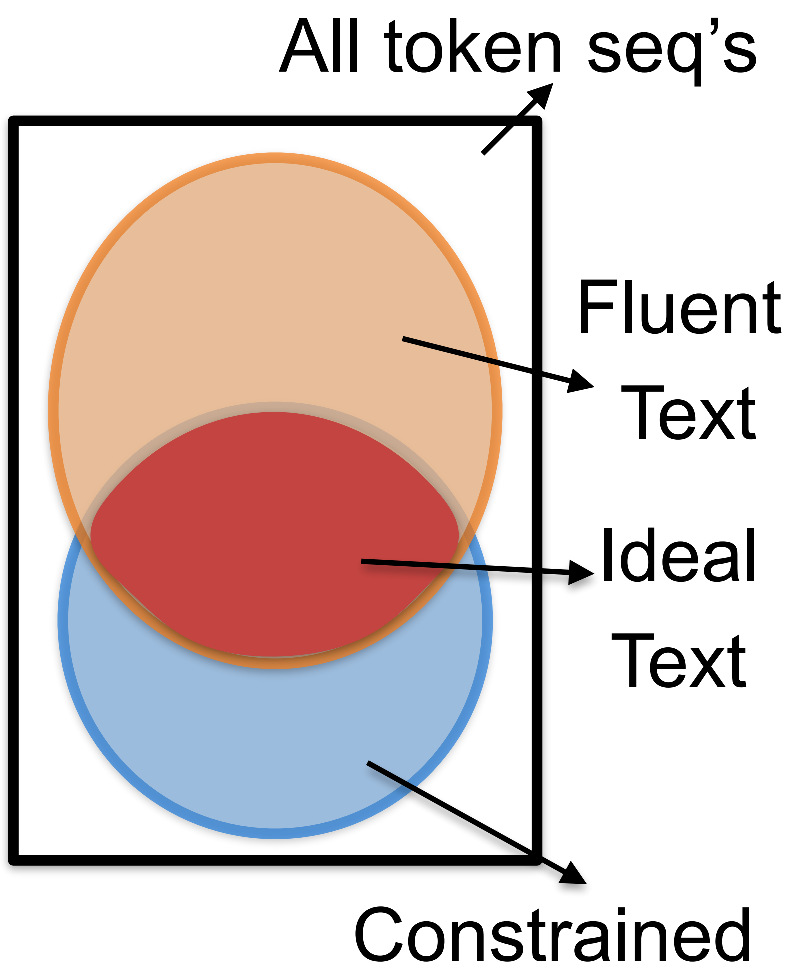
Constrained Sentence Generation via Metropolis-Hastings Sampling

- Key idea: To generation samples from the *implicit* distribution by iterative editing (MH sampling)

$$\pi(x) = \prod_i P(x_i \mid x_{0:i-1}) \cdot \prod_j P_C^j(x)$$

pre-trained language model prob.

indicator (0-1) function for constraints



CGMH [N. Miao, H. Zhou, L. Mou, R. Yan, **Lei Li**, AAAI19]

CGMH: Main Idea

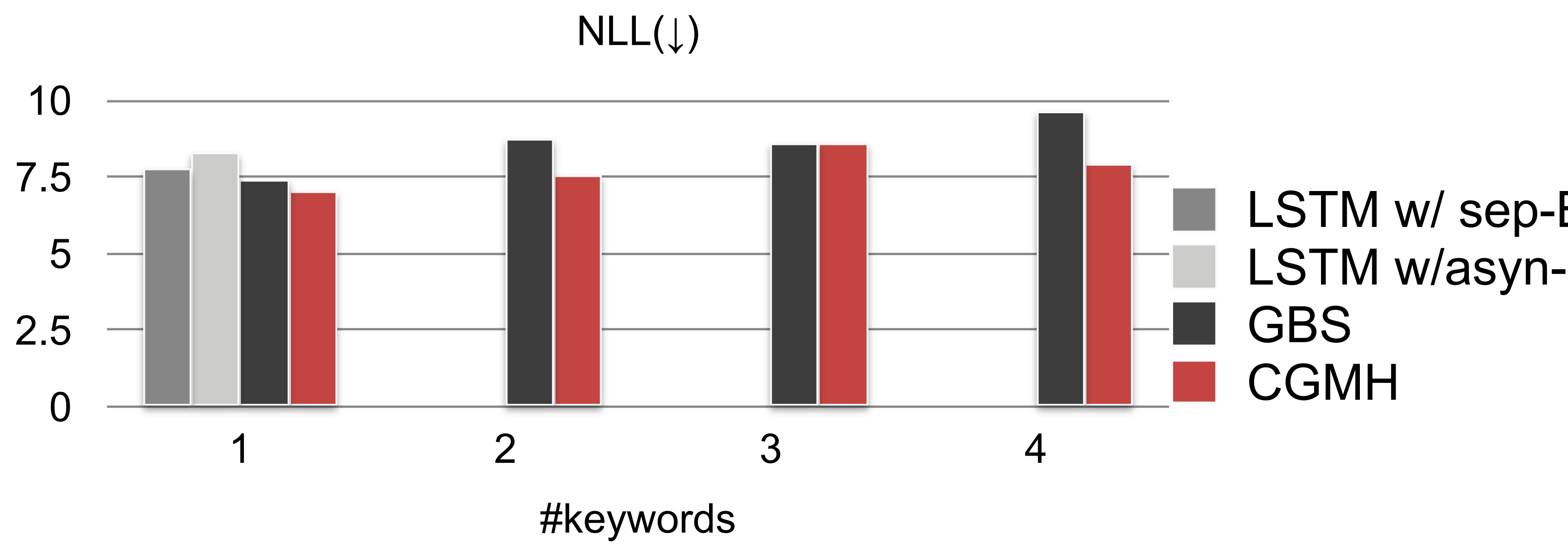
- CGMH performs constrained generation by:
  1. Pretrain Neural Language Model (e.g. GPT2);
  2. Iterative Editing:
     1) Start from a initial sentence x_0;
     2) Propose a new sentence x_t from x_{t-1}, and accept/ reject the action. Action proposal include:
        I. Replacement: change a word to another one
        II. Insertion: add a word
        III. Deletion: remove a word

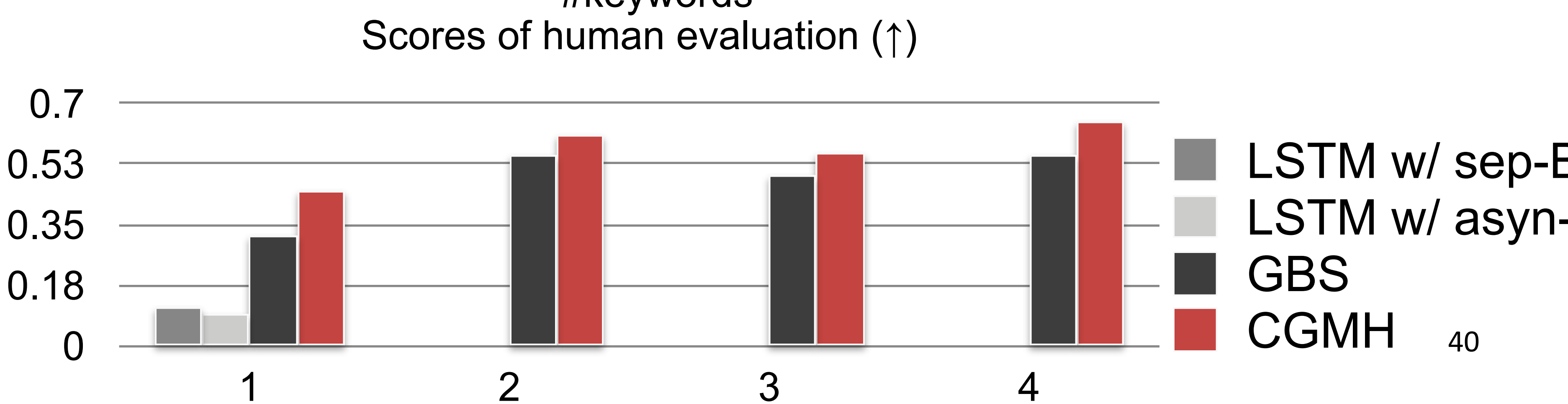
BMW, the sports car of daily life

BMW, the sports car of today's life
BMW, the sports car of future life
BMW, the sports car of new life
BMW, the sports car of happy life
…

CGMH [N. Miao, H. Zhou, L. Mou, R. Yan, **Lei Li**, AAAI19]

CGMH generates better sentences from keywords





Impact

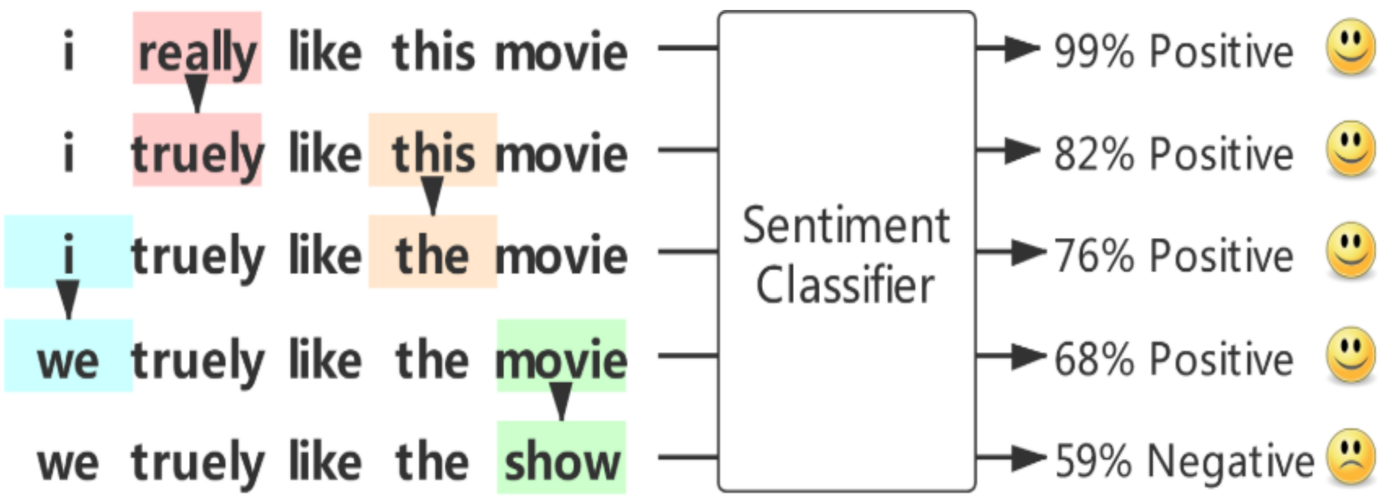
- CGMH is deployed in a large-scale online ads creation platform
- Active used by 100,000 merchants and organizations
- Adoption rate: ~75%

“Autumn”
“Sports shoes”

Comfortable **sports shoes**, a breathing pair of man's shoes, accompanying you in **autumn**

Generating Adversarial Fluent Sentence Generation

- Machine learning models are vulnerable to noises and attacks.
- Generating fluent adversarial text is challenging, due to the discreteness in text! (Ebrahimi et al., 2018; Alzantot et al., 2018)
- Our MHA achieves higher attack success rate



MHA [H. Zhang, N. Miao, H. Zhou, **Lei Li**, ACL19a]

Generation under Combinatorial Constraints

- Logical and Combinatorial constraints
- E.g. generating a question for the following statement.
  - Paris is located in France.
  - ==> Is Paris located in France?
  - ==> Which country is Paris located in?

TSMH [M. Zhang, N. Jiang, **Lei Li**, Yexiang Xue, EMNLP20e]

Generation under Combinatorial Constraints

- Logical and Combinatorial constraints

$$\pi(x) = \underbrace{P_{\mathrm{LM}}(x;\theta)}_{\text{Language Model}} \cdot \underbrace{\phi(x)}_{\text{Constraint}}$$

$$\phi(x) = \beta^{M - \sum_i c_i(x)}, \quad 0 < \beta < 1$$

$c_i(x)$ is a formula or logical constraint. e.g. the first word must be Wh- words.

Method: Tree search enhanced Metropolis-Hastings

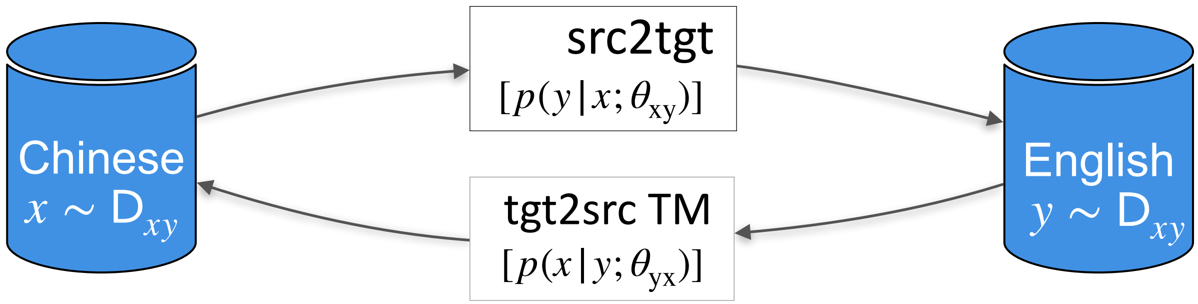
details in TSMH [M. Zhang, N. Jiang, **Lei Li**, Yexiang Xue, EMNLP20e]

Mirror Generative Model for Neural Machine Translation

MGNMT [Z. Zheng, H. Zhou, S. Huang, **Lei Li**, X. Dai, J. Chen, ICLR 2020a]

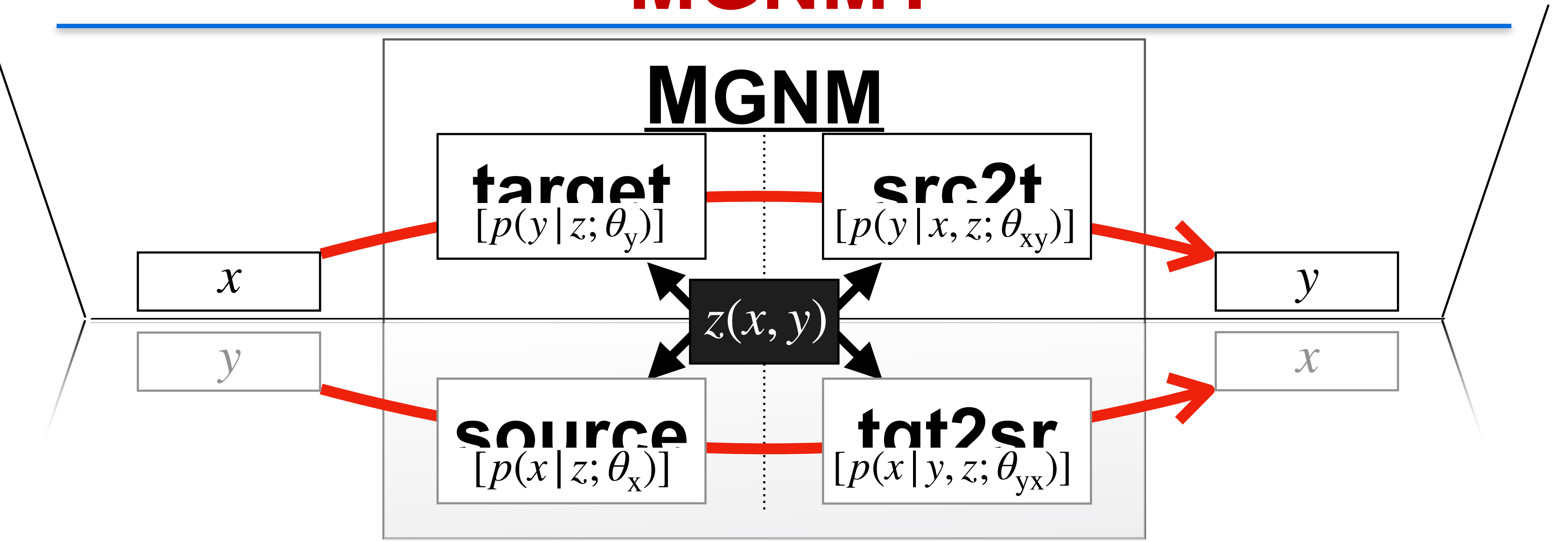
Neural Machine Translation

- Neural machine translation (NMT) systems are super good when you have large amount of **parallel bilingual data**



- **BUT**, very expensive/non-trivial to obtain
  - Low resource language pairs (*e.g.*, English-to-Tamil)
  - Low resource domains (*e.g.*, social network)
- Large-scale mono-lingual data are not fully utilized

Integrating Four Language Skills with MGNMT



1. composing sentence in Source lang
2. composing sentence in Target lang
3. translating from source to target
4. translating from target to source

Benefits utilizing both parallel bilingual data and non-parallel corpus

MGNMT [Z. Zheng, H. Zhou, S. Huang, **L. Li**, X. Dai, J. Chen, ICLR 2020a]

Approach: Mirror-Generative NMT

- The **mirror** property to decompose

$$\log p(x,y|z) = \log p(x|z) + \log p(y|x,z) = \log p(y|z) + \log p(x|y,z)$$
$$= \frac{1}{2}[\underbrace{\log p(y|x,z)}_{\textit{src2tgt}\ \mathrm{TM}_{\mathrm{x}\to\mathrm{y}}} + \underbrace{\log p(y|z)}_{\textit{target}\ \mathrm{LM}_{\mathrm{y}}} + \underbrace{\log p(x|y,z)}_{\textit{tgt2src}\ \mathrm{TM}_{\mathrm{y}\to\mathrm{x}}} + \underbrace{\log p(x|z)}_{\textit{source}\ \mathrm{LM}_{\mathrm{x}}}]$$

$$p(x,y\,|\,z) = p(y\,|\,x,z)p(x\,|\,z) = p(x\,|\,y,z)p(x\,|\,z)$$

- **Relevant** TMs & LMs under a **unified probabilistic framework!**
  - Enables the **aforementioned advantages**

MGNMT makes better use of non-parallel data

- Low resource results

Model	Low-Resource		Cross-Domain			
	WMT16 En↔Ro		In-Domain (TED)		Out-Domain (News)	
	En-Ro	Ro-En	En-De	De-En	En-De	De-En
Transformer (Vaswani et al., 2017)	32.1	33.2	27.5	32.8	17.1	19.9
GNMT (Shah & Barber, 2018)	32.4	33.6	28.0	33.2	17.4	20.1
GNMT-M-SSL + *non-parallel* (Shah & Barber, 2018)	34.1	35.3	28.4	33.7	22.0	24.9
Transformer+BT + *non-parallel* (Sennrich et al., 2016b)	33.9	35.0	27.8	33.3	20.9	24.3
Transformer+JBT + *non-parallel* (Zhang et al., 2018)	34.5	35.7	28.4	33.8	21.9	25.1
Transformer+Dual + *non-parallel* (He et al., 2016a)	34.6	35.7	28.5	34.0	21.8	25.3
MGNMT	32.7	33.9	28.2	33.6	17.6	20.2
MGNMT + *non-parallel*	**34.9**	**36.1**	28.5	34.2	**22.8**	**26.1**

MGNMT makes better use of non-parallel data

- High resource results

Model	WMT14		NIST	
	EN-DE	DE-EN	EN-ZH	ZH-EN
Transformer (Vaswani et al., 2017)	27.2	30.8	39.02	45.72
GNMT (Shah & Barber, 2018)	27.5	31.1	40.10	46.69
GNMT-M-SSL + *non-parallel* (Shah & Barber, 2018)	29.7	33.5	41.73	47.70
Transformer+BT + *non-parallel* (Sennrich et al., 2016b)	29.6	33.2	41.98	48.35
Transformer+JBT + *non-parallel* (Zhang et al., 2018)	30.0	33.6	42.43	48.75
Transformer+Dual + *non-parallel* (He et al., 2016b)	29.6	33.2	42.13	48.60
MGNMT	27.7	31.4	40.42	46.98
MGNMT + *non-parallel*	30.3	33.8	42.56	49.05

  - Non-parallel data is helpful
  - MGNMT works well especially on low resource sett

Multilingual Pretraining NMT

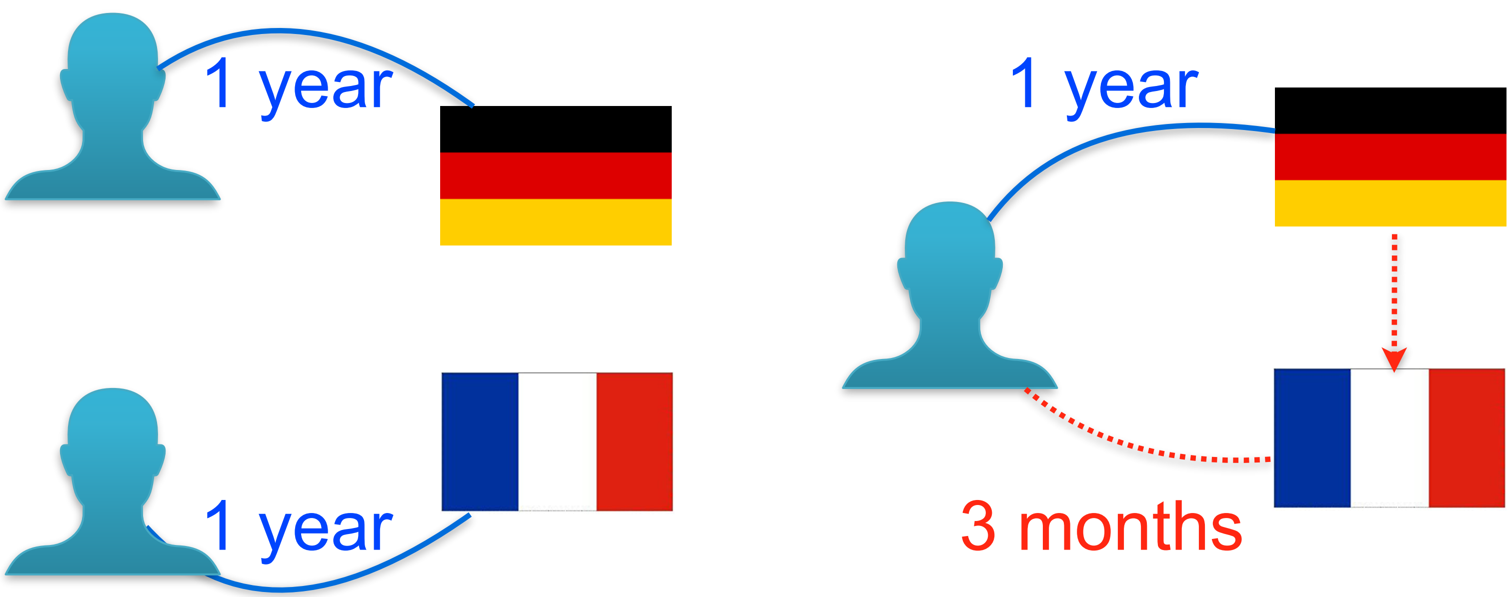
mRASP [Zehui Lin, Xiao Pan, Mingxuan Wang, Xipeng Qiu, Jiangtao Feng, Hao Zhou, Lei Li, EMNLP 2020]

The Ultimate Quest of Machine Translation

- # of human languages: >6900.
- How to build a universal MT system that is capable of translating any source language into a target one?

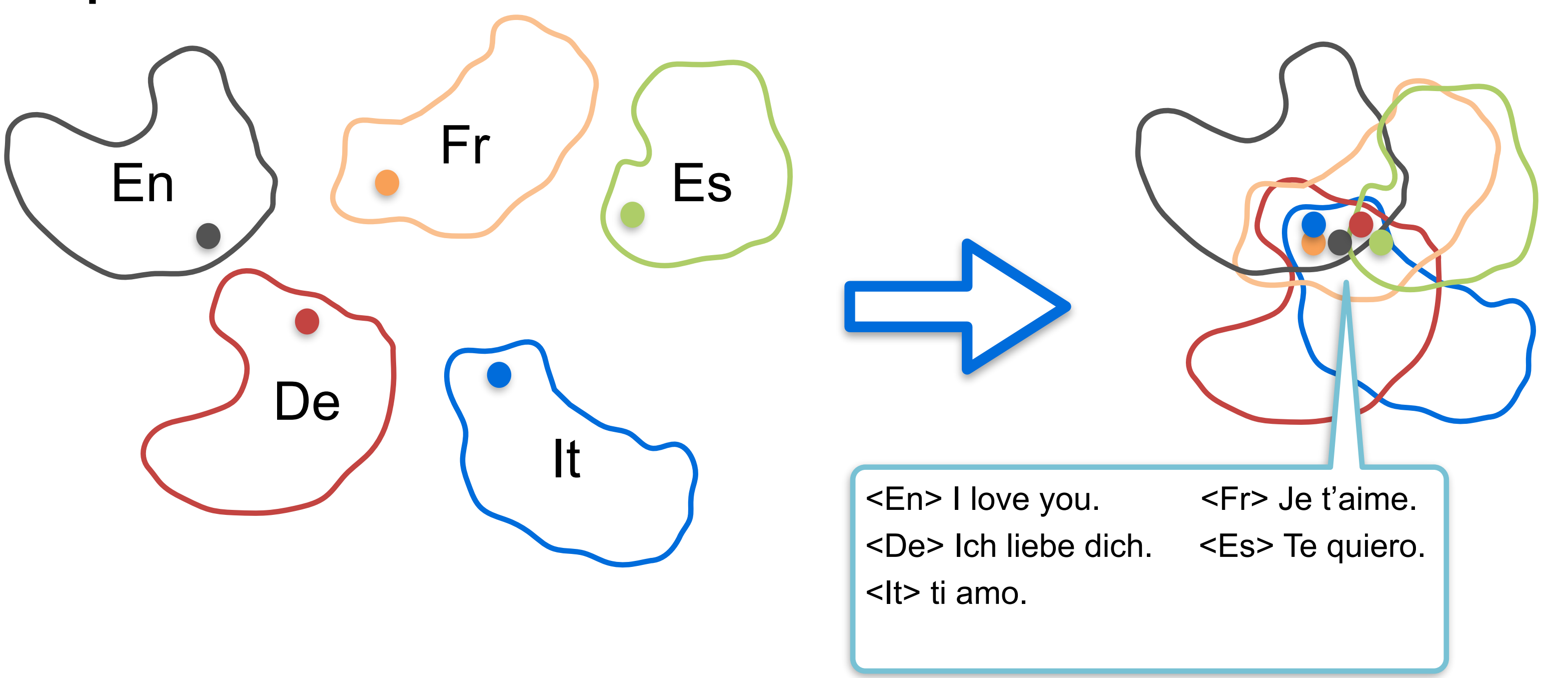
Why Training Multilingual MT Jointly?

- Data scarcity for low/zero resource languages.
- Transfer knowledge between languages.

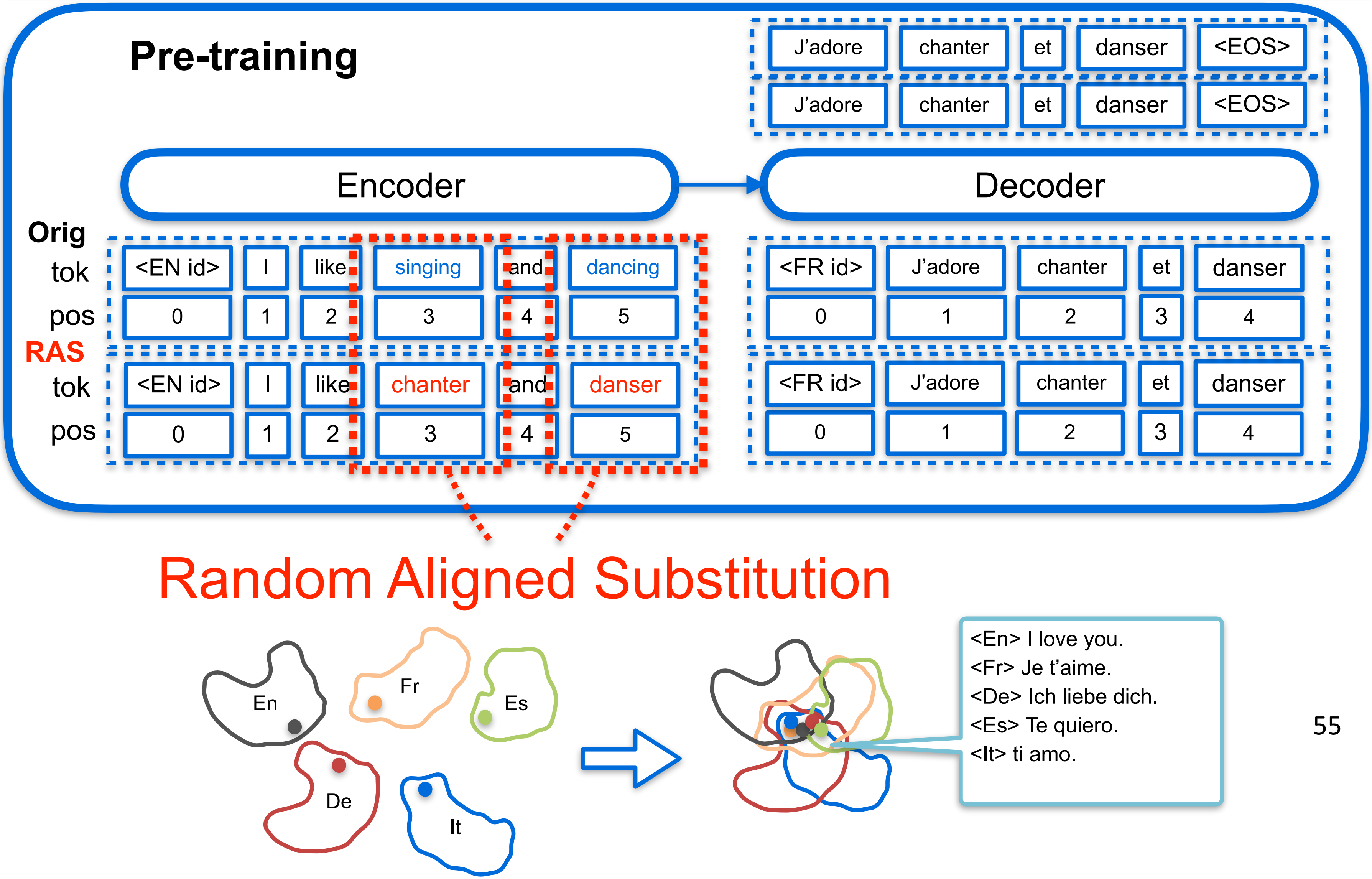


Further Pursuit: Unified Multilingual Representation

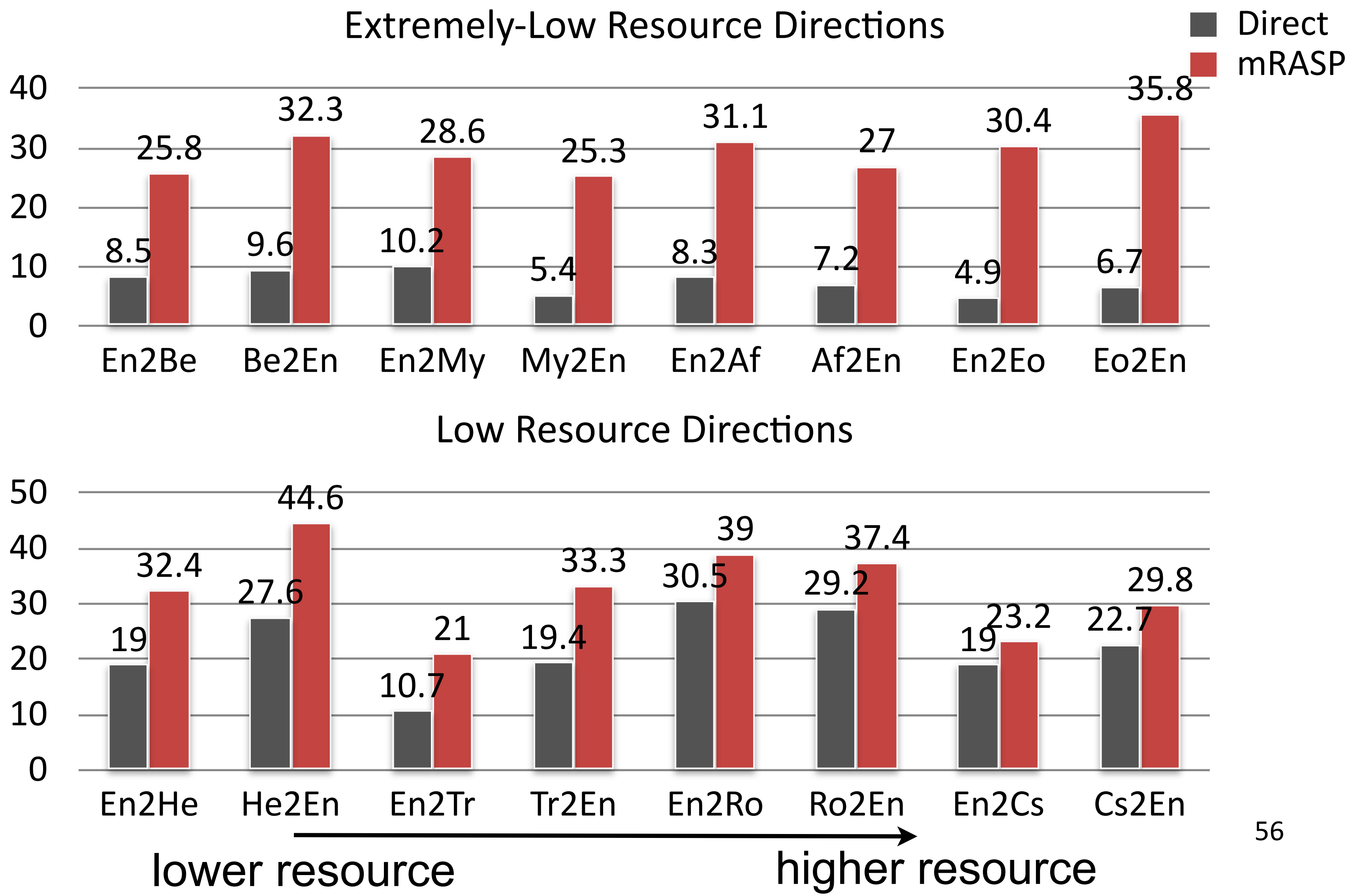
- Further: It is expected to bridge distributional representation of different languages.
- Utterances in different languages with the same semantics will be mapped to adjacent embedding spaces.



Overview of mRASP

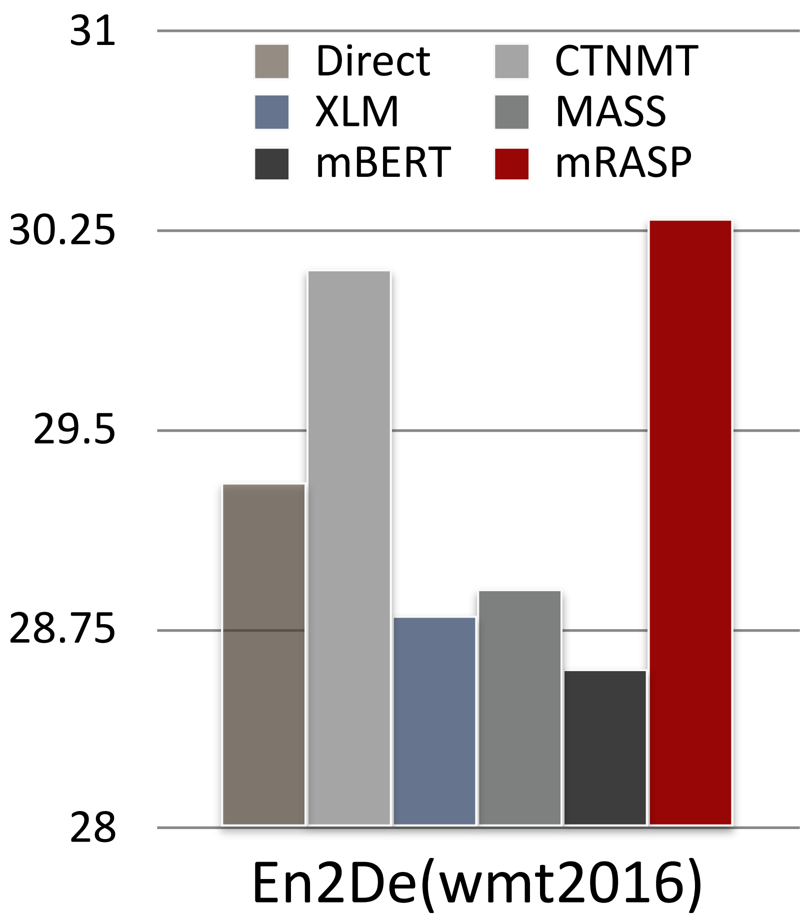


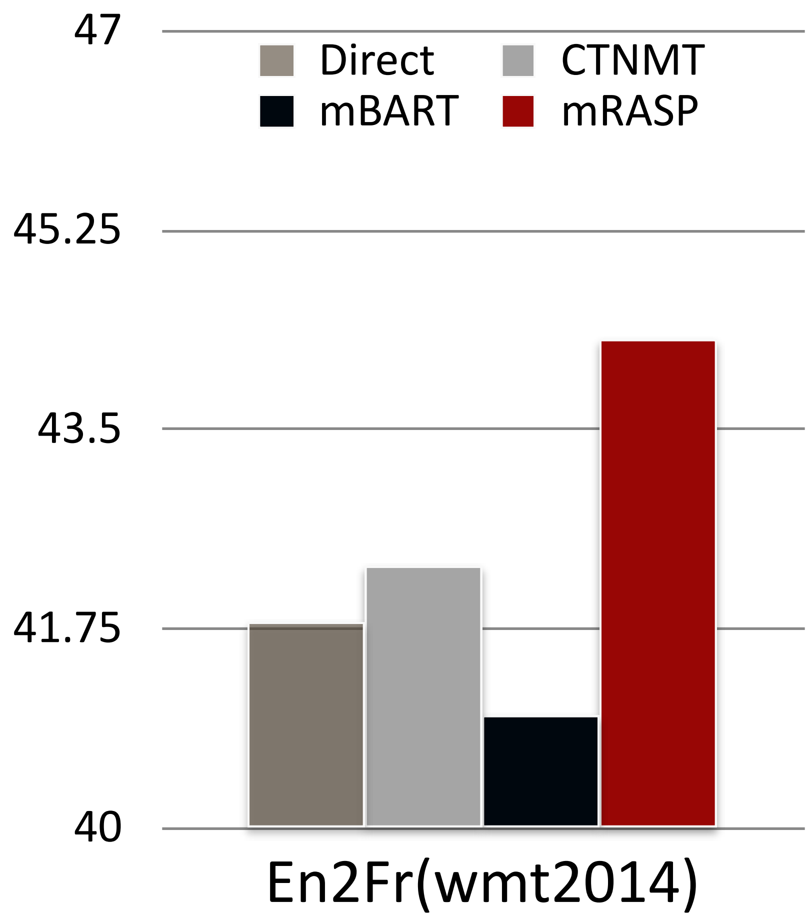
(Extremely) Low Resource



Medium & Rich Resource (Popular Benchmark)

- **Rich resource** benchmarks can be further improved (En->Fr +1.1BLEU).





Does mRASP boost MT performance for Exotic Languages?

- mRASP generalizes on all exotic scenarios.

		Fr-Zh(20K)		De-Fr(9M)	
		—>	<—	—>	<—
Exotic Pair	Direct	0.7	3	23.5	21.2
	mRASP	25.8	26.7	29.9	23.4
		Nl-Pt(12K)		Da-El(1.2M)	
		—>	<—	—>	<—
Exotic Full	Direct	0.0	0.0	14.1	16.9
	mRASP	14.1	13.2	17.6	19.9
		En-Mr(11K)		En-Gl(1.2M)	
		—>	<—	—>	<—
Exotic Source/Target	Direct	6.4	6.8	8.9	12.8
	mRASP	22.7	22.9	32.1	38.1
		En-Eu(726k)		En-Sl(2M)	
		—>	<—	—>	<—
	Direct	7.1	10.9	24.2	28.2
	mRASP	19.1	28.4	27.6	29.5

12k: **Direct** not work **VS** **mRASP** achieves 10+ BLEU!!

Summary

- Multimodal Machine Writing
  - Xiaomingbot system: 600k articles and 150k followers
- Disentangled Latent Representation
  - VTM: Learning Latent Templates in Variational Space
  - DSS-VAE: Disentangled syntax and semantic representation
- DEM-VAE: Self identifying meaningful clusters with corpus
- Bayesian approach to constrained text generation
  - CGMH: generic framework to specify constraints and generate
  - MHA, TSMH
- MGNMT:
  - integrate four language capabilities together
  - Utilize both parallel and non-parallel corpus
- mRASP: a new pre-trained model for many translation directions

For the Community

mRASP Multilingual MT Pretraining
https://github.com/linzehui/mRASP

LightSeq
A high performance sequence processing lib
https://github.com/bytedance/lightseq

https://translate.volcengine.cn

火山翻译

Thanks

- ByteDance AI Lab MLNLC Group and many collaborators
- Contact: lileilab@bytedance.com

Reference

1. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention Is All You Need. NeurIPS 2017.
2. Ning Miao, Hao Zhou, Lili Mou, Rui Yan, Lei Li. “CGMH: Constrained Sentence Generation by Metropolis-Hastings Sampling”. In: the 33rd AAAI Conference on Artificial Intelligence (AAAI). Jan. 2019.
3. Huangzhao Zhang, Ning Miao, Hao Zhou, Lei Li. “Generating Fluent Adversarial Examples for Natural Languages”. In: the 57th Annual Meeting of the Association for Computational Linguistics (ACL) - short papers. July 2019.
4. Yu Bao, Hao Zhou, Shujian Huang, Lei Li, Lili Mou, Olga Vechtomova, Xinyu Dai, Jiajun Chen. “Generating Sentences from Disentangled Syntactic and Semantic Spaces”. In: the 57th Annual Meet- ing of the Association for Computational Linguistics (ACL). July 2019.
5. Ning Miao, Hao Zhou, Chengqi Zhao, Wenxian Shi, Lei Li. “Kernelized Bayesian Softmax for Text Generation”. In: the 33rd Conference on Neural Information Processing Systems (NeurIPS). Dec. 2019.
6. Zaixiang Zheng, Hao Zhou, Shujian Huang, Lei Li, Xinyu Dai, Jiajun Chen. “Mirror Generative Models for Neural Machine Translation”. In: International Conference on Learning Representations (ICLR). Apr. 2020.
7. Rong Ye, Wenxian Shi, Hao Zhou, Zhongyu Wei, Lei Li. “Variational Template Machine for Data- to-Text Generation”. In: International Conference on Learning Representations (ICLR). Apr. 2020.
8. Ning Miao, Yuxuan Song, Hao Zhou, Lei Li. “Do you have the right scissors? Tailoring Pre-trained Language Models via Monte-Carlo Methods”. In: the 58th Annual Meeting of the Association for Computational Linguistics (ACL) - short papers. July 2020.
9. Wenxian Shi, Hao Zhou, Ning Miao, Lei Li. “Dispersing Exponential Family Mixture VAEs for Interpretable Text Generation”. In: Proceedings of the 37th International Conference on Machine learning (ICML). July 2020.
10. Maosen Zhang, Nan Jiang, Lei Li, Yexiang Xue. “Constraint Satisfaction Driven Natural Language Generation: A Tree Search Embedded MCMC Approach”. In: the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings. Nov. 2020.
11. Zehui Lin, Xiao Pan, Mingxuan Wang, Xipeng Qiu, Jiangtao Feng, Hao Zhou, Lei Li. Pre-training Multilingual Neural Machine Translation by Leveraging Alignment Information. EMNLP 2020.