

A Neural Attention Model for Disfluency Detection

Shaolei Wang, Wanxiang Che, Ting Liu

**School of Computer Science and Technology
Harbin Institute of Technology, Harbin, China**

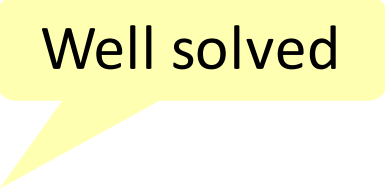
Disfluency Detection

- The transcribed speech text is mostly disfluent
- The goal of disfluency detection is to detect the disfluency in speech text

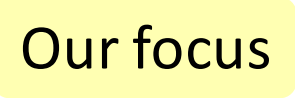
Disfluency Types

□ Disfluencies

- Filled pause
 - uh, oh, um...
- Explicit editing term
 - you know, excuse me, sorry...
- Discourse marker
 - well, so...
- Uncompleted word
 - Pre-...
- Reparandum (edited phrase)

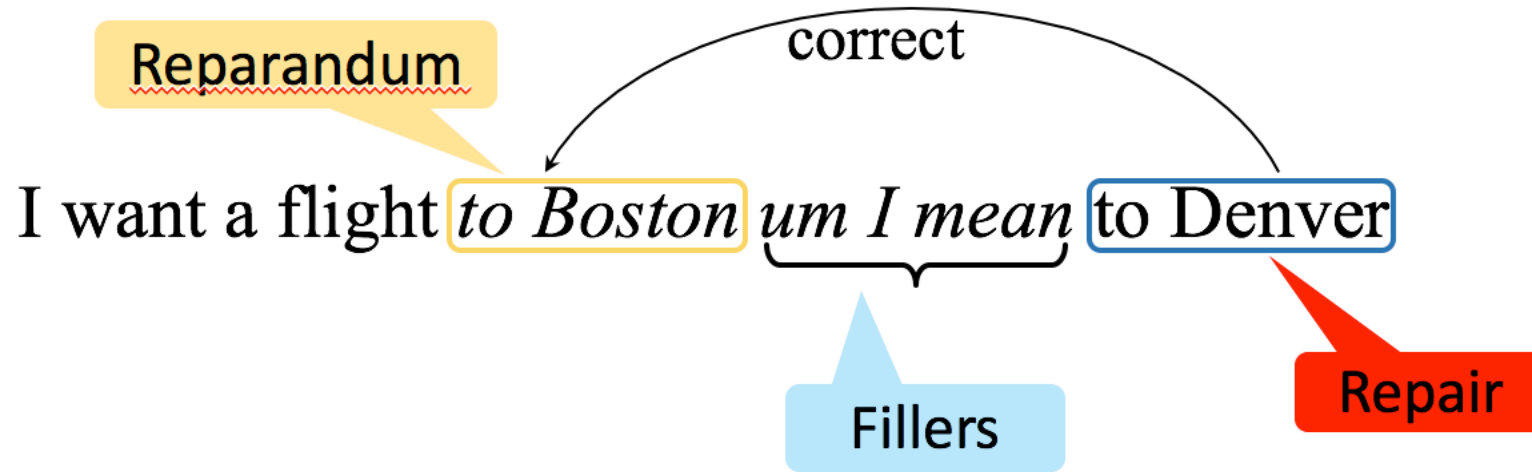


Well solved

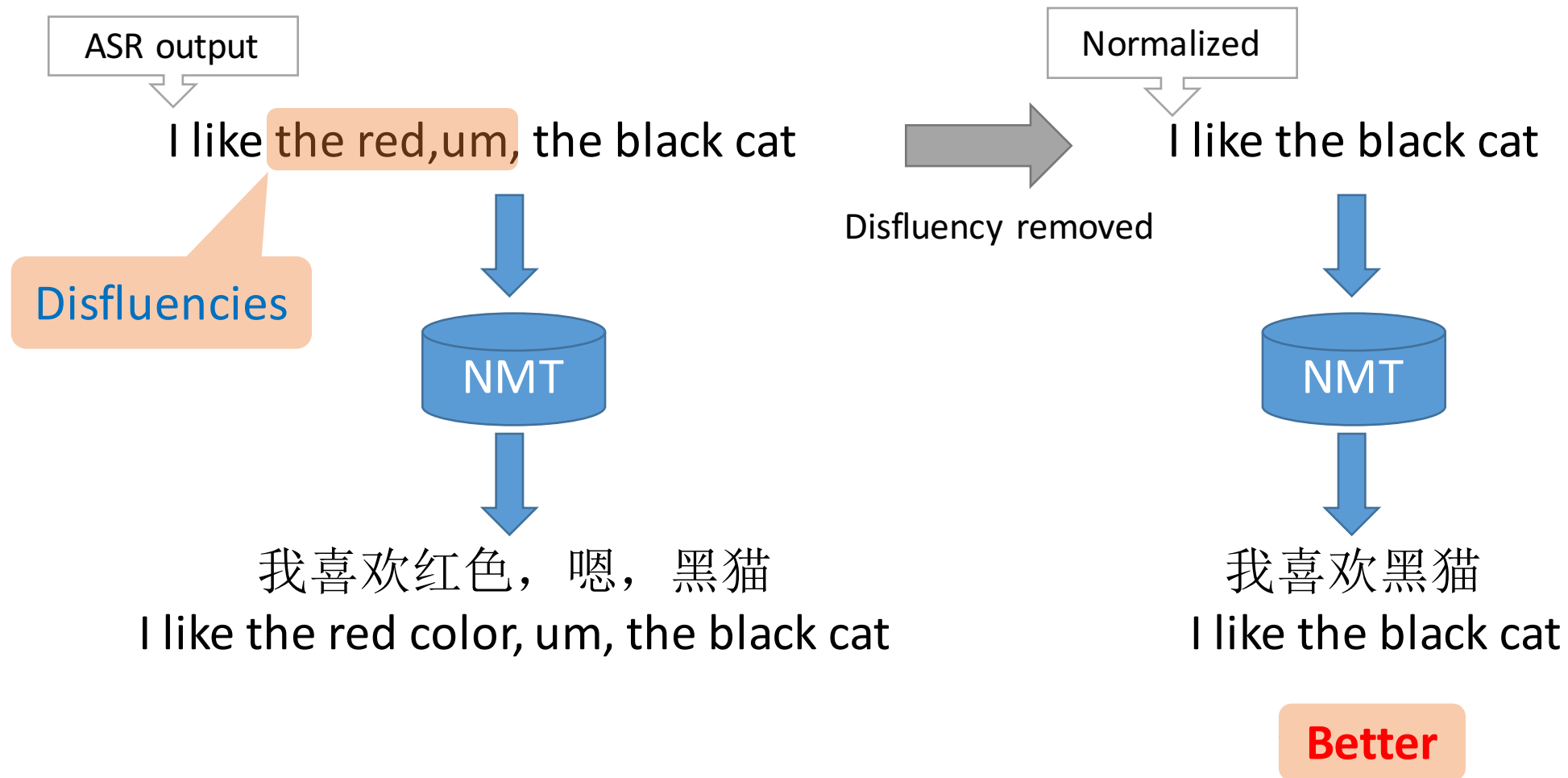


Our focus

Reparandum Disfluencies



Disfluency Effect on Machine Translation



Challenges of Disfluency Detection

- Vary in length and occur everywhere
- Long-range dependencies
- Keep the generated sentences grammatical

Related Work

- Sequence labeling
 - M³N labeling
 - (Qian et al., NAACL 2013)
 - Beam search decoding
 - (Wang et al., Coling 2014)
 - Semi-markov model
 - (Ferguson et al., NAACL 2015)

Related Work

- Joint parsing and disfluency
 - Left-to-right(L2R) parsing-based joint model
 - (Honnibal et al., TACL 2014)
 - Right-to-Left(R2L) parsing-based joint model
 - (Wu et al., ACL 2015)

Related Work

- Recurrent neural network (RNN)
 - RNN for incremental disfluency detection
 - (Hough and Schlangen, 2015)
 - Bidirectional LSTM
 - (Zayats et al., 2016)

Our Motivations

- Sequence-to-sequence method

Our Motivations

- Sequence-to-sequence method
 - utilize the global representation of the input sentence and may provide a good solution to the long-range dependencies question

Our Motivations

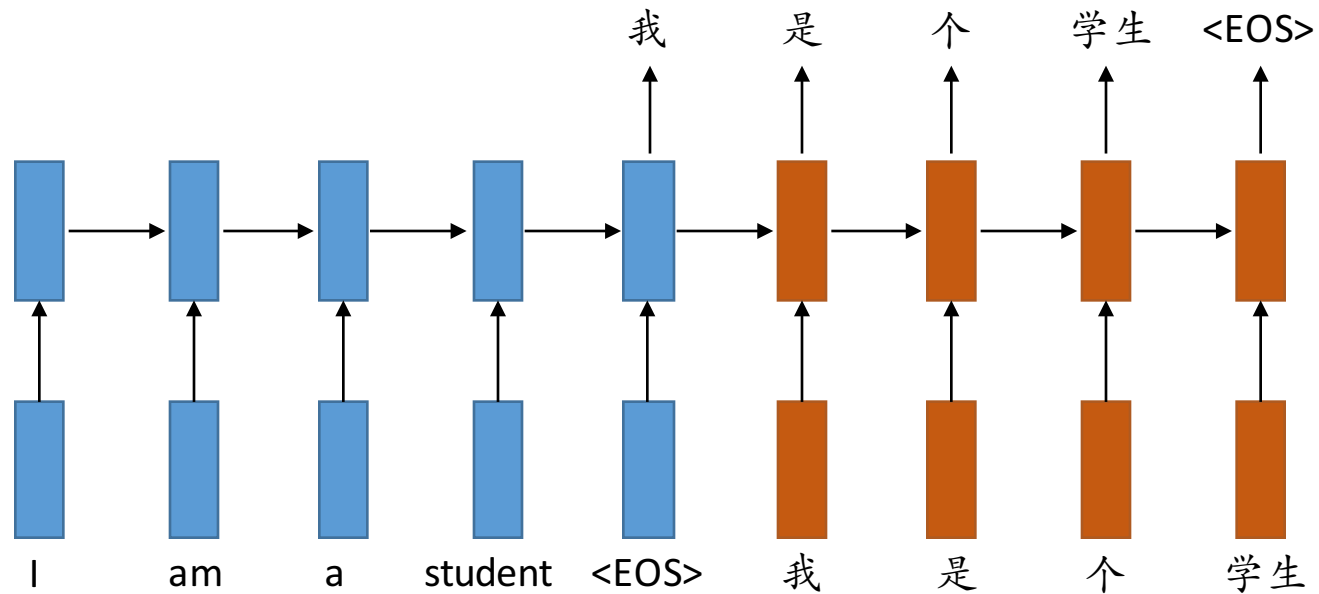
□ Sequence-to-sequence method

- utilize the global representation of the input sentence and may provide a good solution to the long-range dependencies question
- can be seen as a conditional language model and thus has the ability to keep the generated sentences grammatical

Background (Attention)

- Seq2Seq based MT model

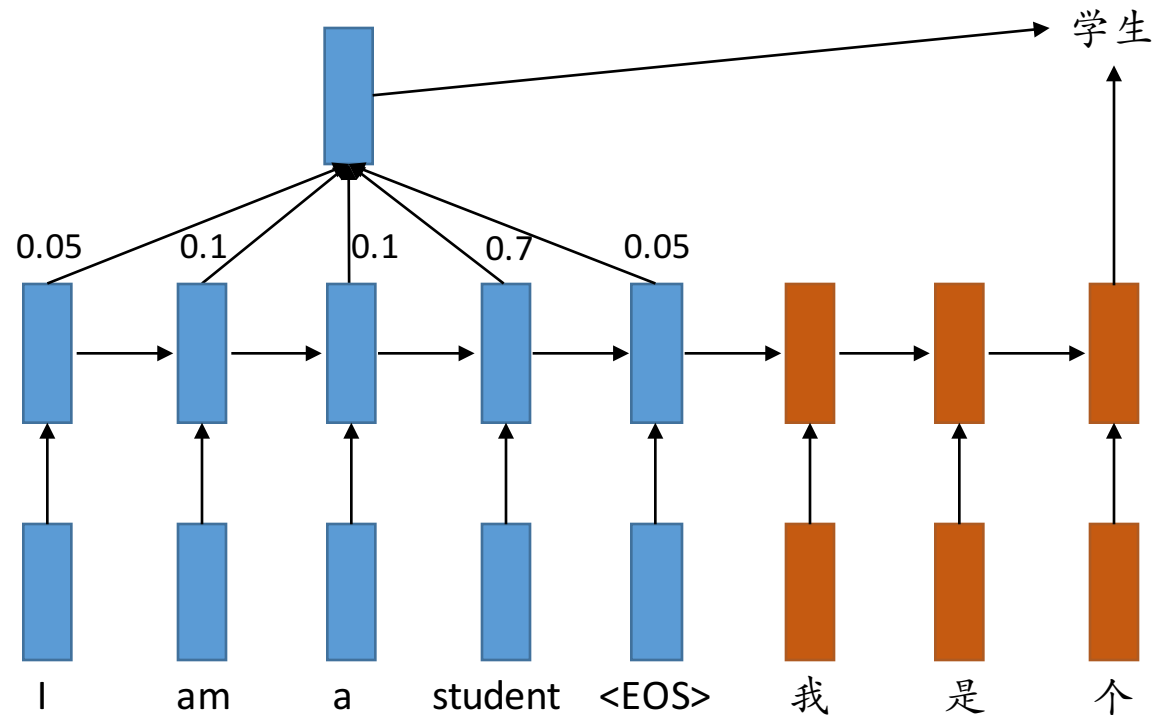
- Source: "I am a student" -> Target: " 我是个学生"



Background (Attention)

□ Attention-based MT model

- When predicting a target word, it first weighs every location in source sentence and then it calculates a weighted sum



Background (Attention)

- Disfluency detection requires that the output sentence should be an ordered subsequence of the input sentence

Background (Attention)

- Disfluency detection requires that the output sentence should be an ordered subsequence of the input sentence
- Limitations of the above attention network

Background (Attention)

- Disfluency detection requires that the output sentence should be an ordered subsequence of the input sentence
- Limitation of the above attention network
 - may generate a word not appearing in the input sentence

Background (Attention)

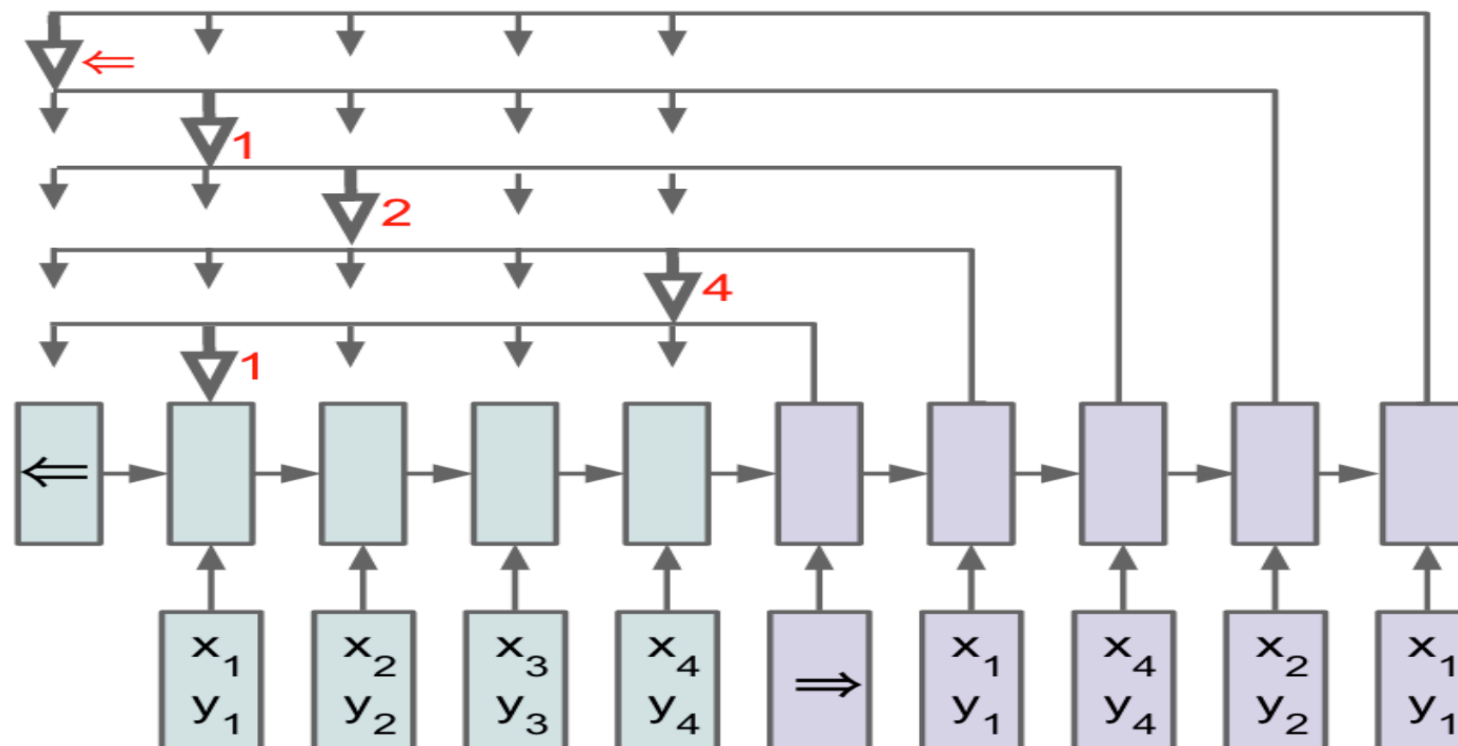
- Disfluency detection requires that the output sentence should be an ordered subsequence of the input sentence
- Limitation of the above attention network
 - may generate a word not appearing in the input sentence
 - can not generate the word not appearing in the fixed output dictionary

Background (Attention)

- Disfluency detection requires that the output sentence should be an ordered subsequence of the input sentence
- Limitation of the above attention network
 - may generate a word not appearing in the input sentence
 - can not generate the word not appearing in the fixed output dictionary
 - has no ability to model the order of the generated words

Background (Pointer network)

- Pointer network(Vinyals et al., NIPS 2015)
 - When predicting a target word, it first weighs every location in source sentence and then it select word with maximum weight



Background (Pointer network)

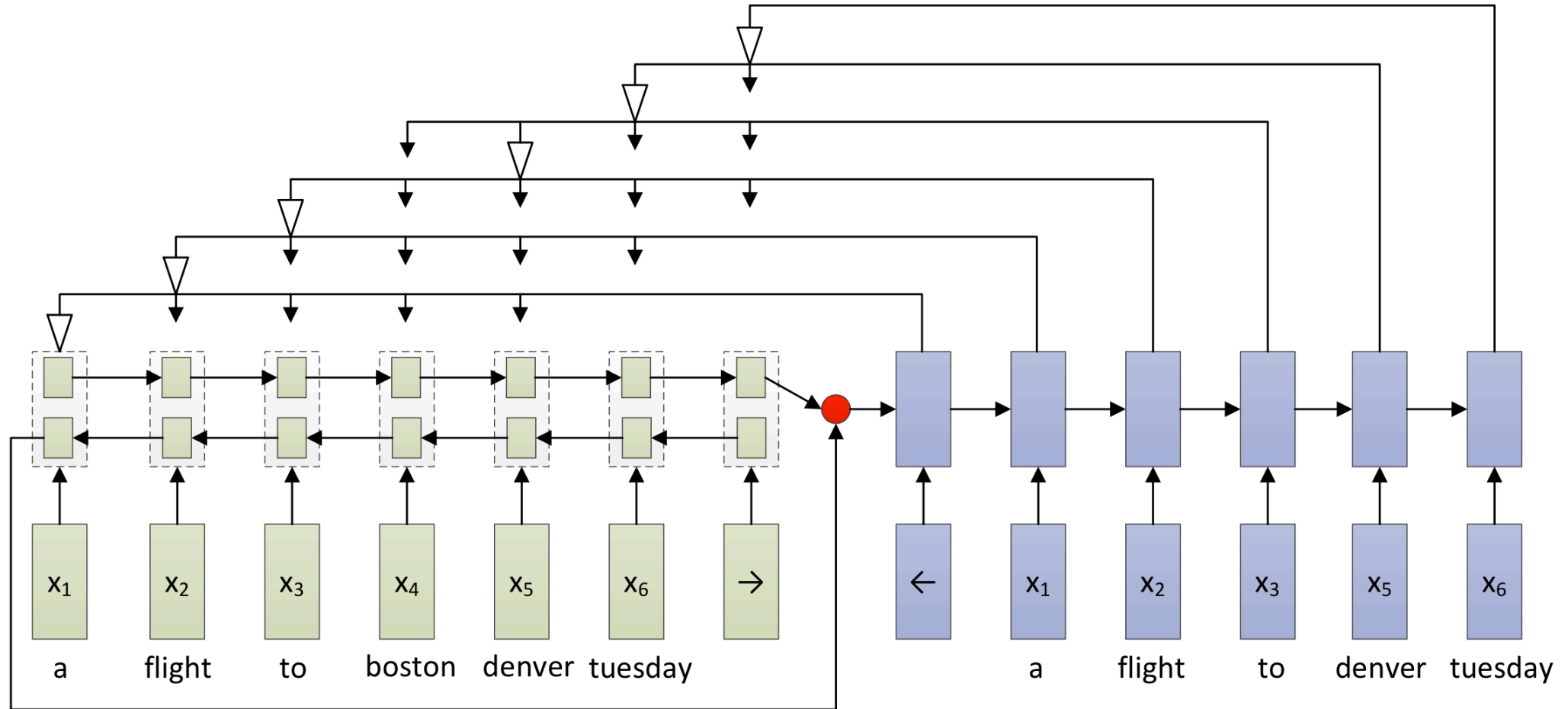
□ Solved:

- All of the word generated is in the input sentence
- Break the limit of the fixed output vocabulary

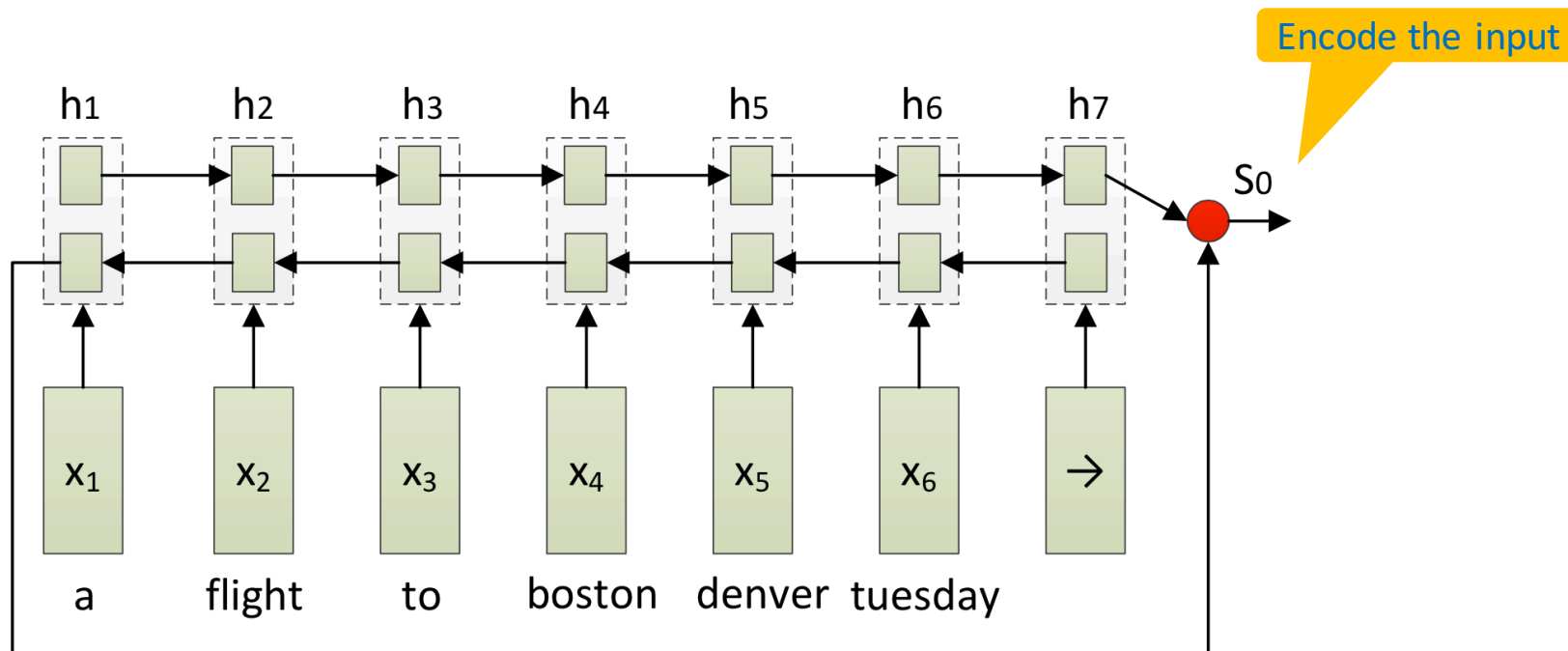
□ Unsolved

- has no ability to model the order of the generated words

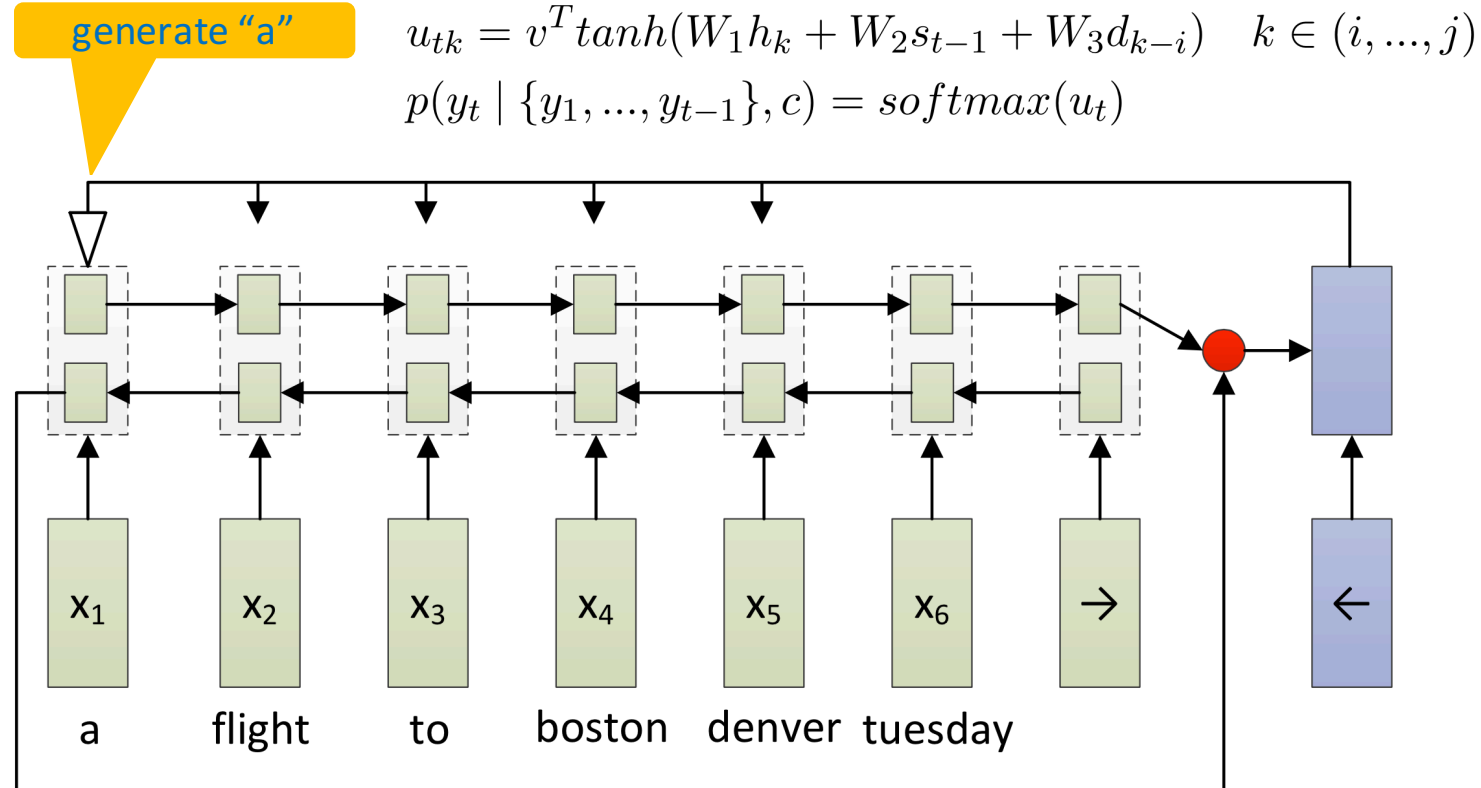
Our Model



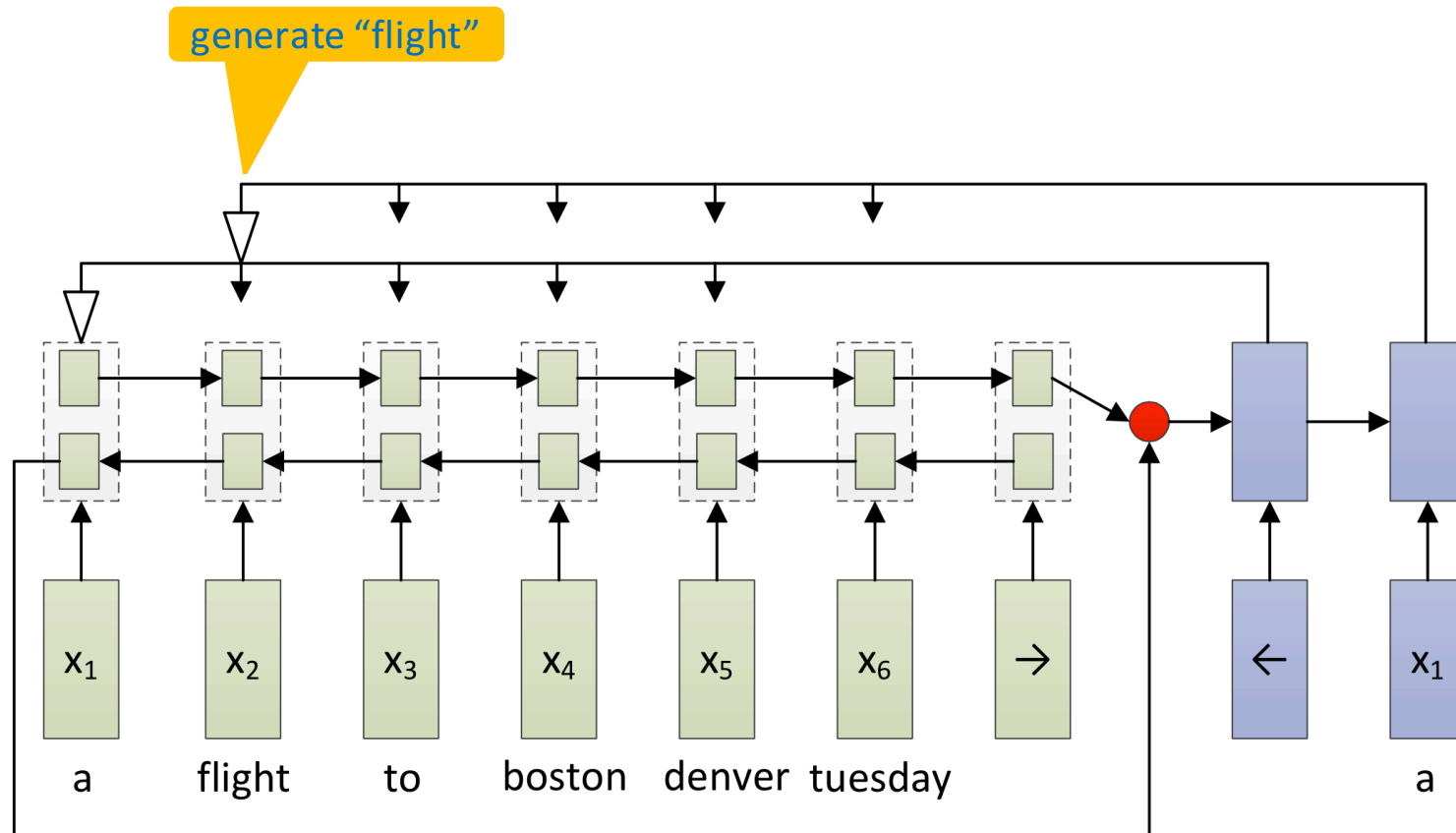
Our Model



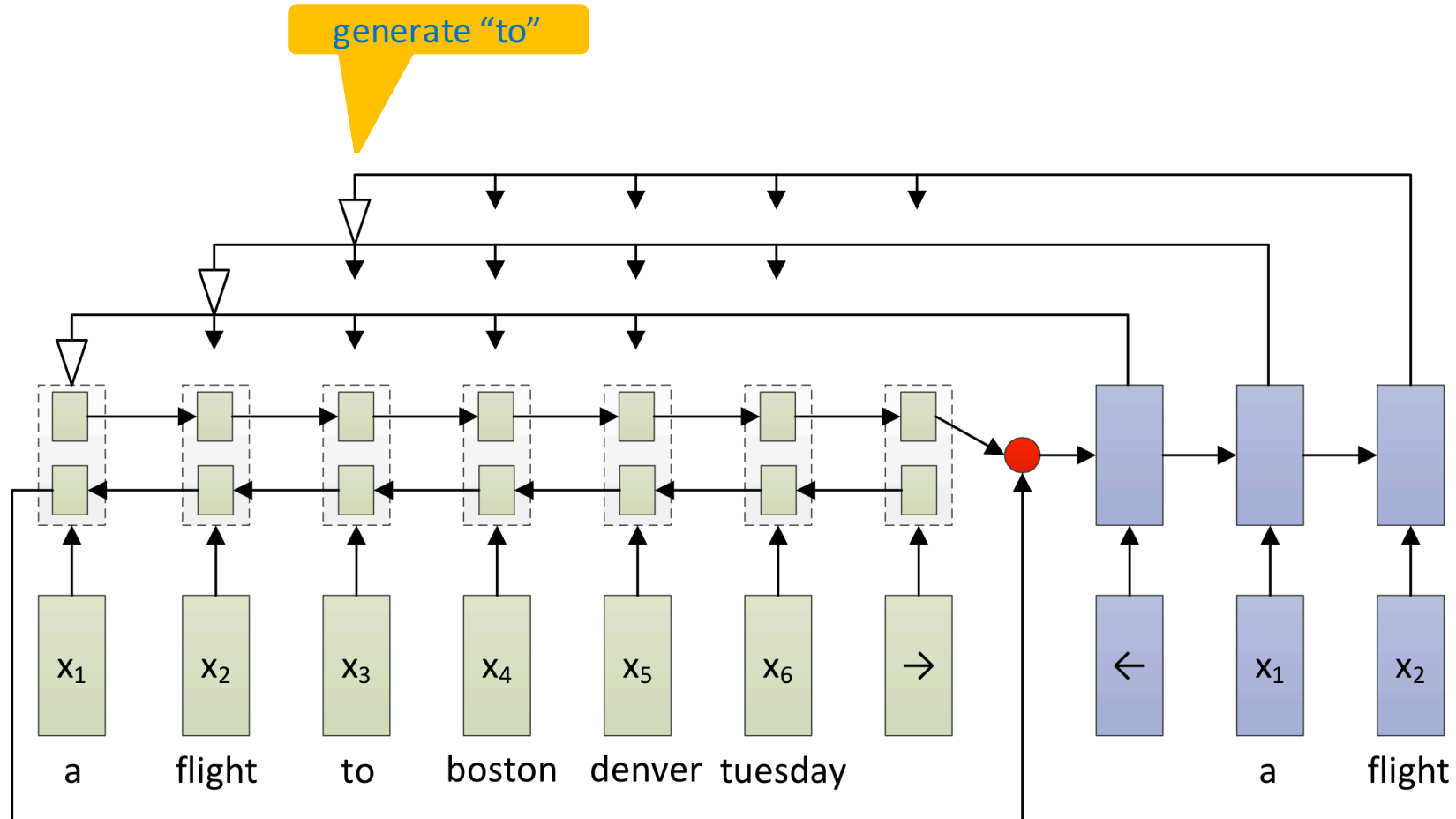
Our Model



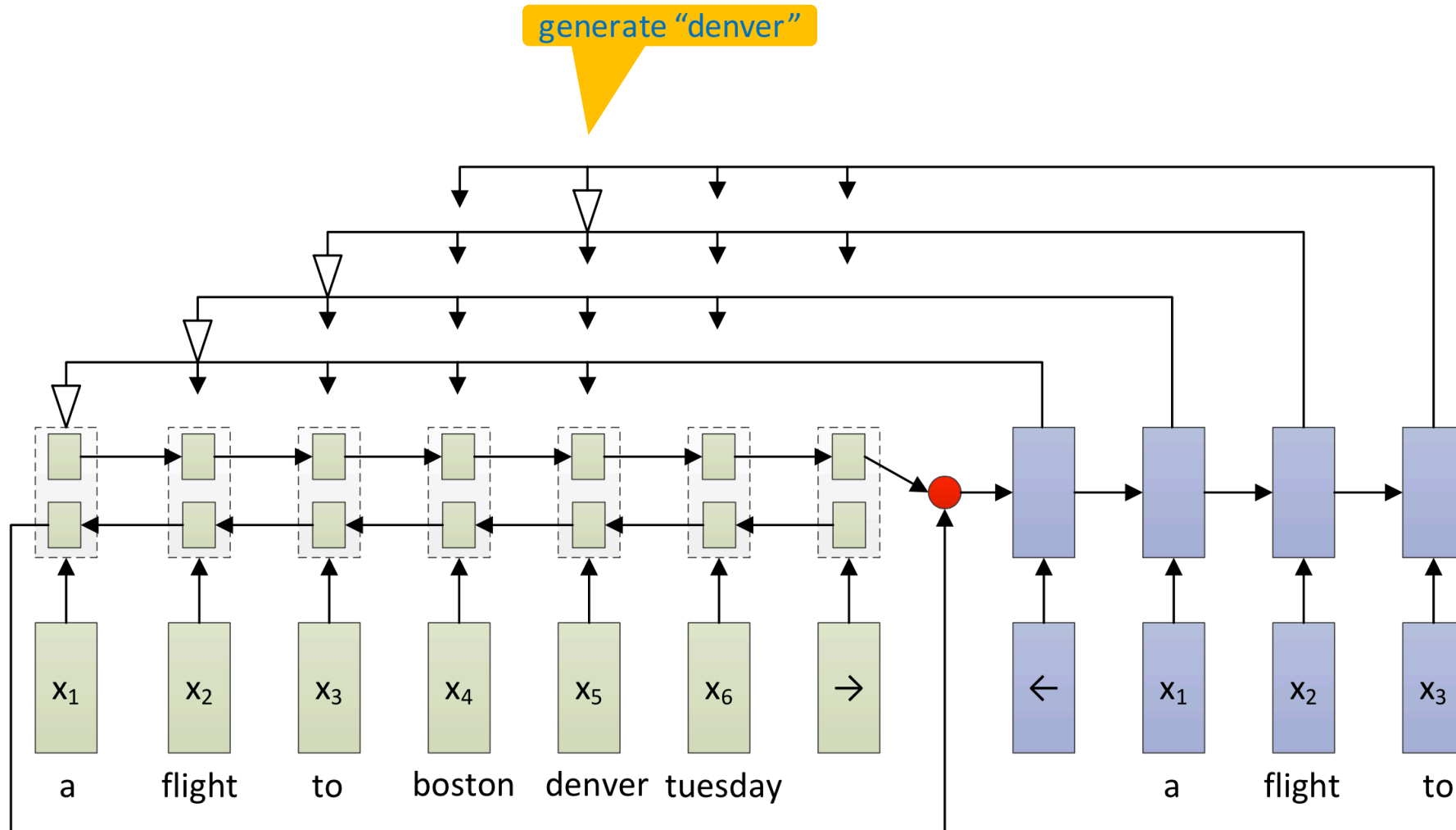
Our Model



Our Model

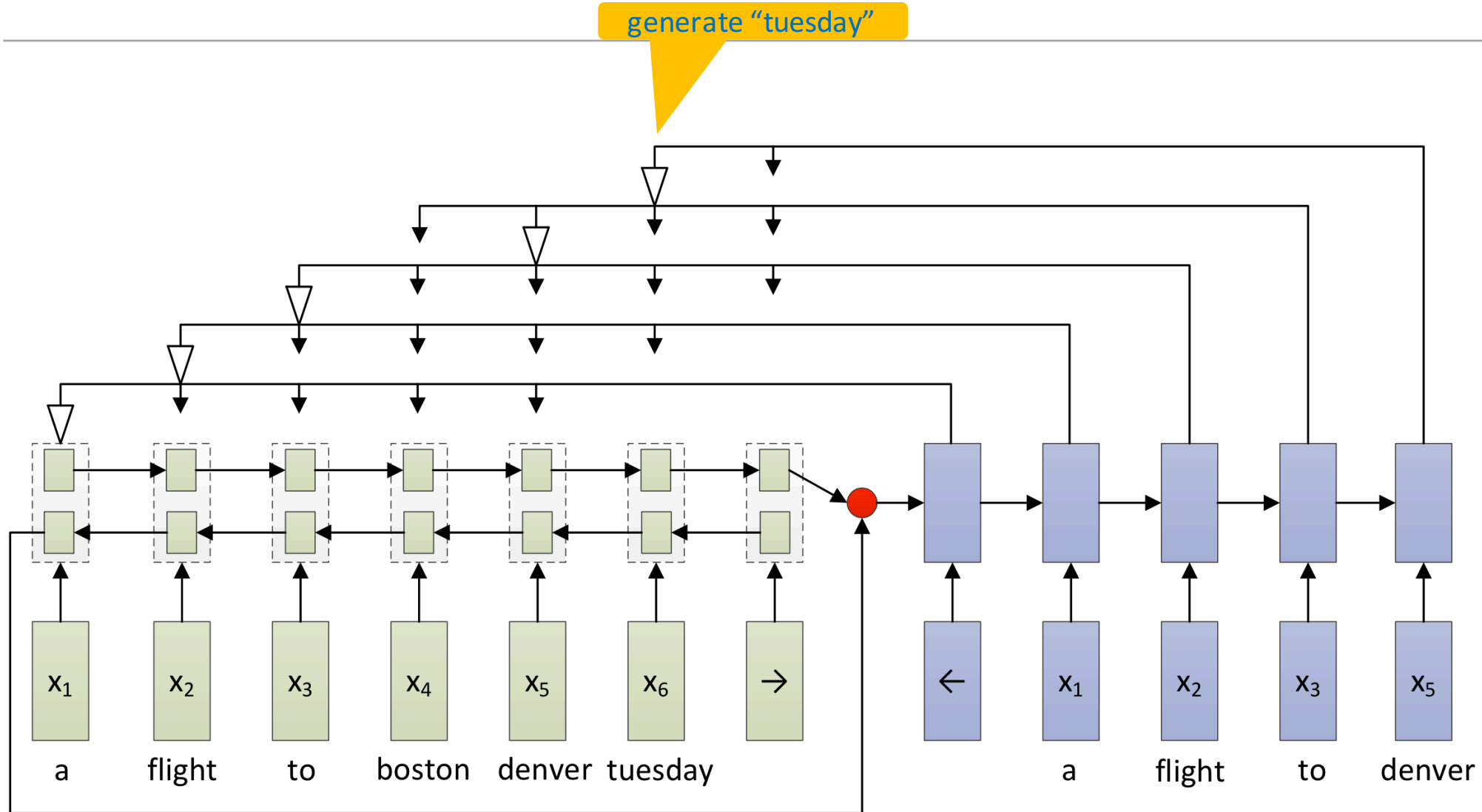


Our Model



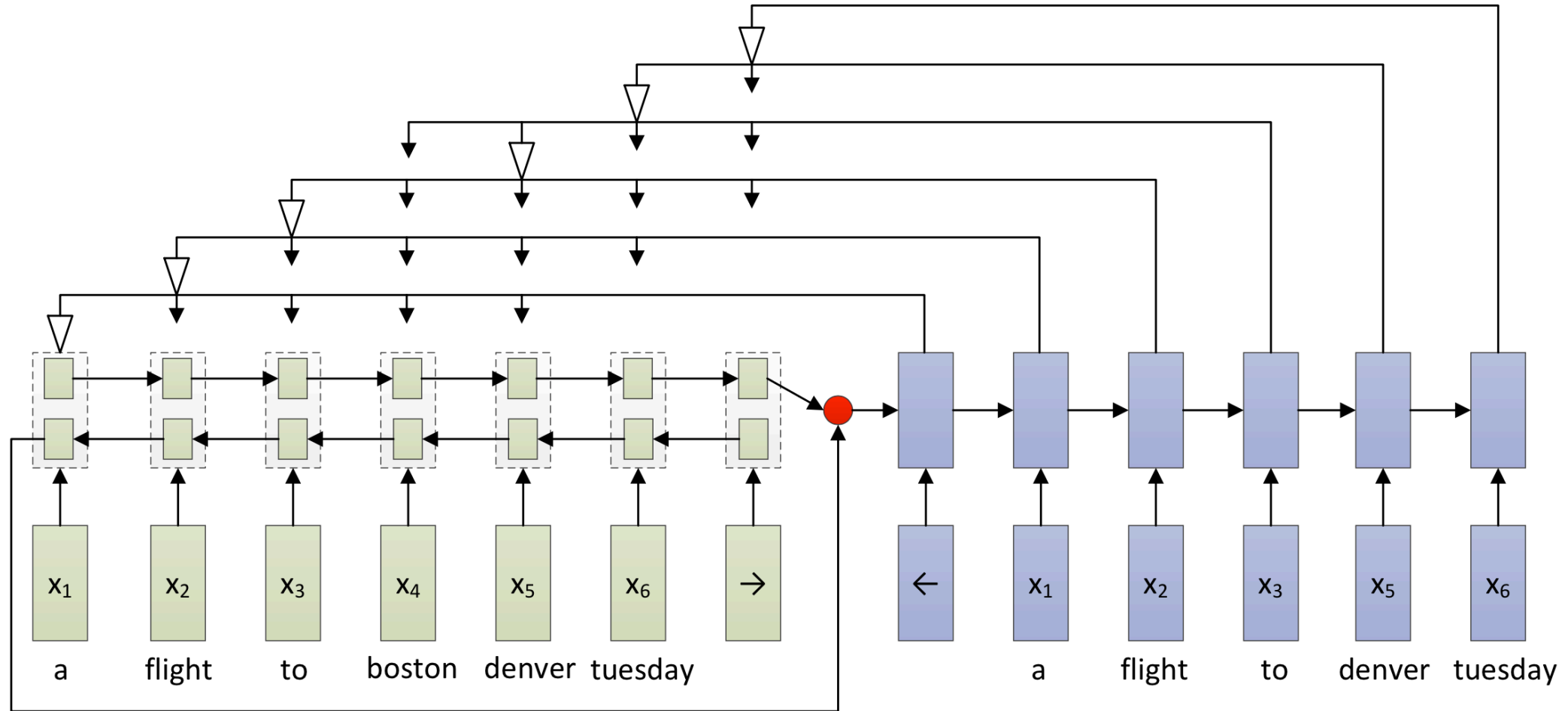
Our Model

generate "tuesday"



Our Model

generate "->"



Our Model

□ Input representation:

$$x = \max\{0, V[\tilde{w}; w; p; d] + b\}$$

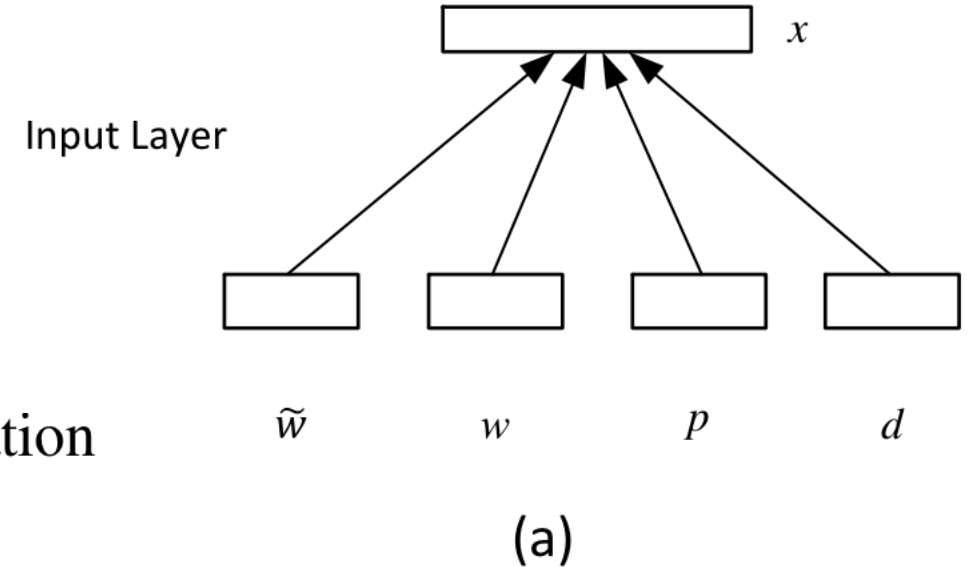
Where

w : a learned word embedding

p : a learned POS-tag embedding

d : a hand-crafted feature representation

\tilde{w} : a fixed word embedding



Our Model

□ The hand-crafted features:

duplicate features

$Duplicate(i, w_{i+k}), -15 \leq k \leq +15$ and $k \neq 0$: if w_i equals w_{i+k} , the value is 1, others 0

$Duplicate(p_i, p_{i+k}), -15 \leq k \leq +15$ and $k \neq 0$: if p_i equals p_{i+k} , the value is 1, others 0

$Duplicate(w_i w_{i+1}, w_{i+k} w_{i+k+1}), -4 \leq k \leq +4$ and $k \neq 0$: if $w_i w_{i+1}$ equals $w_{i+k} w_{i+k+1}$,
the value is 1, others 0

$Duplicate(p_i p_{i+1}, p_{i+k} p_{i+k+1}), -4 \leq k \leq +4$ and $k \neq 0$: if $p_i p_{i+1}$ equals $p_{i+k} p_{i+k+1}$,
the value is 1, others 0

similarity features

$fuzzyMatch(w_i, w_{i+k}), k \in \{-1, +1\}$:

$similarity = num_same_letters / (len(w_i) + len(w_{i+k}))$.

if $similarity > 0.8$, the value is 1, others 0

Model Training

- minimize the negative log-probability of the output sequence over the input:

$$-\sum_{i=1}^N \log(p(y_i|x_i)) = -\sum_{i=1}^N \log\left(\prod_{t=1}^T p(y_t | \{y_1, \dots, y_{t-1}\}, c)\right)$$

Experimental Setting

□ Dataset

□ English Switchboard corpus

- (Following Charniak and Johnson, 2001)
- Training data: Switchboard sw[23] files
- Dev data: Switchboard sw4[5-9] files
- Test data (only used once): Switchboard sw[0-1] files

□ Evaluation metric

$$Prec. = \frac{\#Correctly\ Predicted}{\#Predicted}$$

$$Rec. = \frac{\#Correctly\ Predicted}{\#Total}$$

$$F1 = \frac{2 * Prec.* Rec.}{(Rec.+Prec.)}$$

Experiment results

- Experiment results on the development and test data of English Switchboard data

Method	Dev			Test		
	P	R	F1	P	R	F1
CRF	93.8%	77.7%	85.0%	92.0%	74.5%	82.3%
Attention-based method	93.0%	81.6%	86.9%	91.6%	82.3%	86.7%

Experiment results

- Comparison with the previous state-of-the-art methods

Method	P	R	F1
Attention-based	91.6%	82.3%	86.7%
M ³ N (Qian and Liu, 2013)	-	-	84.1%
Joint Parser (Honnibal and Johnson, 2014)	-	-	84.1%
semi-CRF (Ferguson et al., 2015)	90.1%	80.0%	84.8%
UBT (Wu et al., 2015)	90.3%	80.5%	85.1%

Experiment results

- Comparison with the previous state-of-the-art methods

Method	P	R	F1
Attention-based	91.6%	82.3%	86.7%
M ³ N (Qian and Liu, 2013)	-	-	84.1%
Joint Parser (Honnibal and Johnson, 2014)	-	-	84.1%
semi-CRF (Ferguson et al., 2015)	90.1%	80.0%	84.8%
UBT (Wu et al., 2015)	90.3%	80.5%	85.1%

+1.9%

Experiment results

- Comparison with the previous state-of-the-art methods

Method	P	R	F1
Attention-based	91.6%	82.3%	86.7%
M ³ N (Qian and Liu, 2013)	-	-	84.1%
Joint Parser (Honnibal and Johnson, 2014)	-	-	84.1%
semi-CRF (Ferguson et al., 2015)	90.1%	80.0%	84.8%
UBT (Wu et al., 2015)	90.3%	80.5%	85.1%

+1.6%

Chinese Experimental Setting

- In-house annotated Chinese corpus
 - 200k spoken sentences from minutes of meetings
 - Training data: 160k sentences
 - Dev data: 20k sentences
 - Test data: 20k sentences

Chinese Experiment results

- performance on Chinese annotated data

Method	Dev			Test		
	P	R	F1	P	R	F1
CRF	76.5%	42.0%	54.2%	75.9%	41.6%	53.8%
Attention-based method	83.7%	50.6%	63.1%	82.4%	48.9%	61.4%

Chinese Experiment results

- performance on Chinese annotated data

Method	Dev			Test		
	P	R	F1	P	R	F1
CRF	76.5%	42.0%	54.2%	75.9%	41.6%	53.8%
Attention-based method	83.7%	50.6%	63.1%	82.4%	48.9%	61.4%

- <http://www.iflyrec.com/>
 - The online disfluency detection has a much better F1-score

Conclusion

- We try to use the sequence-to-sequence framework for the problem of disfluency detection
- Propose a novel attention-based model for disfluency detection

Conclusion

- We try to use the sequence-to-sequence framework for the problem of disfluency detection
- Propose a novel attention-based model for disfluency detection
 - Utilize the global representation of the sentence
 - Take into account the language model
 - Achieve the state-of-art results on both English Switchboard corpus and in-house annotated Chinese corpus

Future work

- Greedy search is used in our work and can result in serious error propagation

- Explore the beam search method on this neural structure



Research Center for Social Computing and Information Retrieval

理解语言 认知社会

Thank you!



Research Center for Social Computing and Information Retrieval

理解语言 认知社会

语音转写产品优势_讯飞听见

www.iflyrec.com/help/product.jsp

应用 reinforcement lear... quora google ACL Wiki ACL Anthology 哈尔滨工业大学社会... 首页 - 王少磊 - 赛尔... 机器学习 论文写作 公开课 论文 NLP

讯飞听见 www.iflyrec.com

首页 | 录音宝 | 听见录音笔 | 智能会议系统 | 小音智能速记 | 充值卡商城

产品优势

讯飞听见网站是专为习惯于PC端操作的用户设计，是一个音频一键转文字的快速服务平台。网站以语音转文字为核心功能，提供便捷高效的机器转写服务和专业精准的人工转写服务。网站支持多种音频、视频格式上传，方便您将录音整理成文字，可解决企事业单位日常会议、媒体发布会、教育培训、媒体传播等各种场景下的音频转写问题，让各行各业的人不再为速记费用昂贵、整理录音复杂、查找重点困难、角色辨认模糊、录音质量低劣等方方面面的录音及整理问题而苦恼。

机器转写

- 
- 
- 

双引擎转写 文本顺滑

网站针对VIP客户会调用更好的转写

话标准

在线咨询