

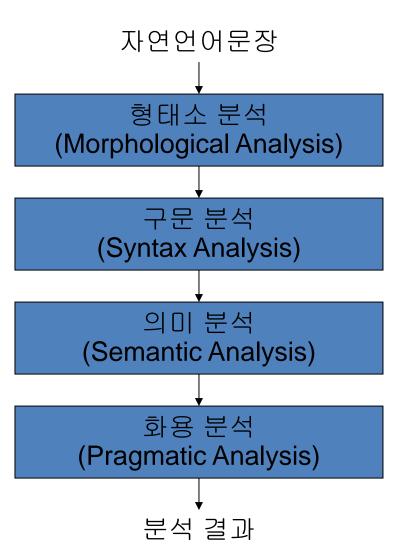
차례

- 자연어처리 소개
- 딥러닝 소개
- 딥러닝 기반의 자연어처리
 - Classification Problem
 - Sequence Labeling Problem
 - Sequence-to-Sequence Learning
 - Pointer Network
 - Machine Reading Comprehension

자연어처리

- 자연언어
 - 인공언어에 대응되는 개념
 - 인공언어: 특정 목적을 위해 인위적으로 만든 언어 (ex. 프로그래밍 언어) 자연언어에 비해 엄격한 구문을 가짐
 - 특정 집단에서 사용되는 모국어의 집합
 - 한국어, 영어, 불어, 독일어, 스페인어, 일본어, 중국어 등
- 자연언어처리 (Natural Language Processing)
 - 컴퓨터를 통하여 인간의 언어를 이해하고 처리하는 학문 분야
 - 기계번역, 자동통역, 정보검색, 질의응답, 문서요약, 철자오류 수 정 등
 - Google, Naver, IBM Watson, Apple Siri, ...

자연언어 분석 단계



- **형태소분석**: "감기는"의 결과
 - ➤ 감기(명사:cold) + 는(조사)
 - ▷ 감(동사 어간) + 기(명사화 어미) + 는(조사)
 - ▶ 감(동사 어간) + 기는(어미)
- 구문분석: Structural Ambiguities
 - ➤ Time flies like light →2가지 이상 tree
 - ➤ A man see a woman with a telescope → 2가지 이 상 tree

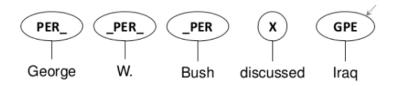


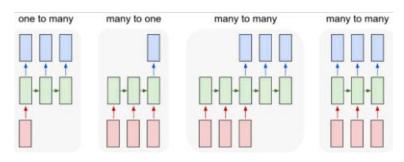
- **의미분석**: "말이 많다"
 - 말: horse or speech ?
- **화용분석**: "A씨는... B씨는 ... 그는 ..." ▶ 그: A or B?

자연어처리 특징

- Natural languages are **ambiguous**
 - Rule → Classification (Maximum Entropy, SVM) → Deep Learning
- NLP datasets are high dimensional
 - One-hot representation → Continuous representation (Word Embedding)
- Many NLP problems can be viewed as sequence labeling tasks
 - − Hidden Markov Model(HMM) → Conditional Random Fields (CRF) → Deep Learning (RNN)
- Many NLP problems can be posed as sequence-to-sequence tasks
 - Rule → Statistical Machine Translation → Neural MT

- "감기는" → 감기(명사) or 감다(동사) + 기 "말이 많다" → 말 = horse or speech ? "A씨는... B씨는 ... 그는 ..." → 그: A or B ?
- Ex. [0 0 0 0 0 **1** 0 0 0 0 0 0 0 0 0 0
- Dimensionality
 - 50K (PTB) 500K (big vocab) 3M ...



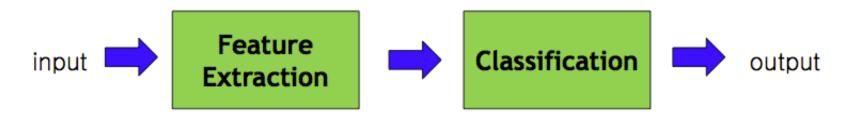


차례

- 자연어처리 소개
- 딥러닝 소개
- 딥러닝 기반의 자연어처리
 - Classification Problem
 - Sequence Labeling Problem
 - Sequence-to-Sequence Learning
 - Pointer Network
 - Machine Reading Comprehension

Why Deep Neural Networks?: Integrated Learning

- 기존 기계학습 방법론
 - Handcrafting features → time-consuming



Deep Neural Network: Feature Extractor + Classifier



Why Deep Neural Networks?: Unsupervised Feature Learning

- 기계학습에 많은 학습 데이터 필요
 - 소량의 학습 데이터
 - 학습 데이터 구축 비용/시간
 - 대량의 원시 코퍼스 (unlabeled data)
 - Semi-supervised, Unsupervised ...
- Deep Neural Network
 - Pre-training 방법을 통해 대량의 원시 코퍼스에서 자질 학습
 - Restricted Boltzmann Machines (RBM)
 - Stacked Autoencoder, Stacked Denosing Autoencoder
 - Word Embedding (for NLP)

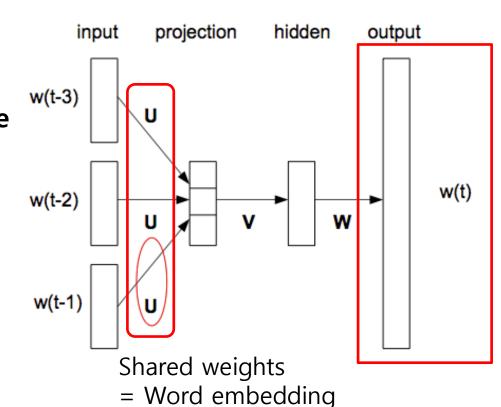
텍스트의 표현 방식

- NLP datasets are high dimensional
- One-hot representation (or symbolic)
 - Ex. [0 0 0 0 0 **1** 0 0 0 0 0 0 0 0 0]
 - Dimensionality
 - 50K (PTB) 500K (big vocab) 3M (Google 1T)
 - Problem
 - Motel [0 0 0 0 0 0 0 0 1 0 0] AND
 - Hotel [0 0 0 0 0 0 1 0 0 0 0] = 0
- Continuous representation
 - Latent Semantic Analysis, Random projection
 - Latent Dirichlet Allocation, HMM clustering
 - Neural Word Embedding
 - Dense vector
 - By adding supervision from other tasks → improve the representation

Neural Network Language Model (Bengio0 0,03)

LT: $|V|^*d$, Input(one hot): $|V|^*1 \rightarrow LT^TI$

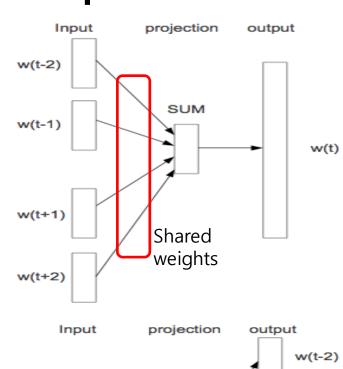
- Idea
 - A word and its context is a positive training sample
 - A random word in that same context → negative training sample
 - Score(positive) > Score(neg.)
- Training complexity is high
 - − Hidden layer → output
 - Softmax in the output layer
 - Hierarchical softmax
 - Negative sampling
 - Ranking(hinge loss)

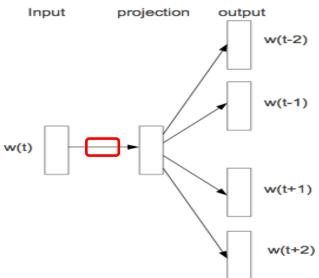


Input	Dim: 1	Dim: 2	Dim: 3	Dim: 4	Dim: 5
1 (boy)	0.01	0.2	-0.04	0.05	-0.3
2 (girl)	0.02	0.22	-0.05	0.04	-0.4

Word2Vec: CBOW, Skip-Gram

- Remove the hidden layer → Speedup 1000x
 - Negative sampling
 - Frequent word sampling
 - Multi-thread (no lock)
- Continuous Bag-of-words (CBOW)
 - Predicts the current word given the context
- Skip-gram
 - Predicts the surrounding words given the current word
 - CBOW + DropOut/DropConnect





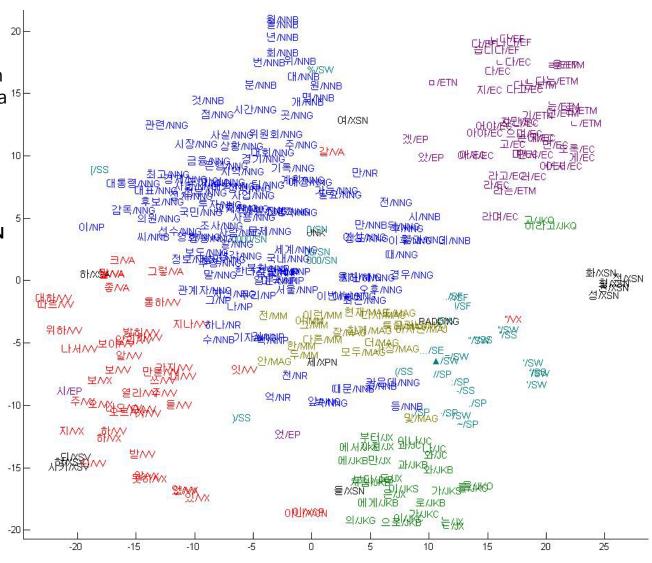
한국어 Word Embedding: NNLM

• Data

- 세종코퍼스 원문 + Korean Wiki abstract + News data 15
 - 2억 8000만 형태소
- Vocab. size: 60,000
 - 모든 형태소 대상 (기호, 숫자, 한자, 조사 포함)
 - 숫자 정규화 + 형태소 /POS: 정부/NNG, 00/SN

NNLM model

- Dimension: 50
- Matlab 구현
- 학습시간: 16일



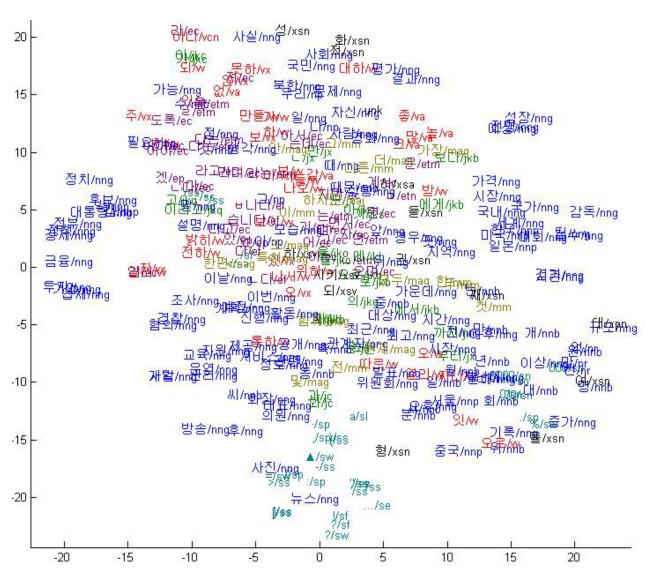
한국어 Word Embedding: Word2Vec(CBOW)

Data

- 2012-2013 News + Korean Wiki abstract:
 - 9GB raw text
 - 29억 형태소
- Vocab. size: 100,000
 - 모든 형태소 대상 (기호, 숫 자, 한자, 조사 포함)
 - 숫자 정규화 + 영어 소문자 + 형태소/POS

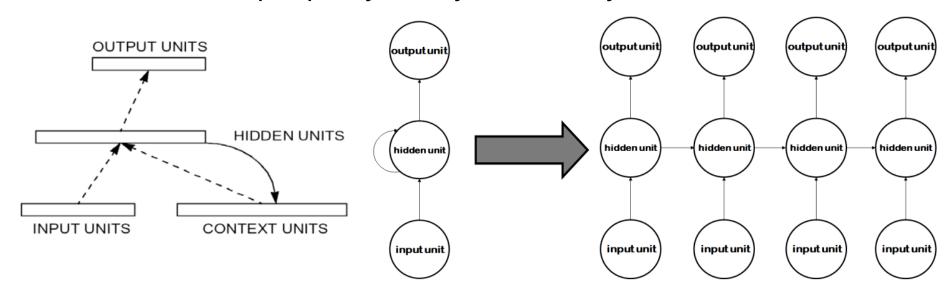
Word2Vec

- 모델: **CBOW** > SKIP-Gram
- 학습 시간: 24분



Recurrent Neural Network

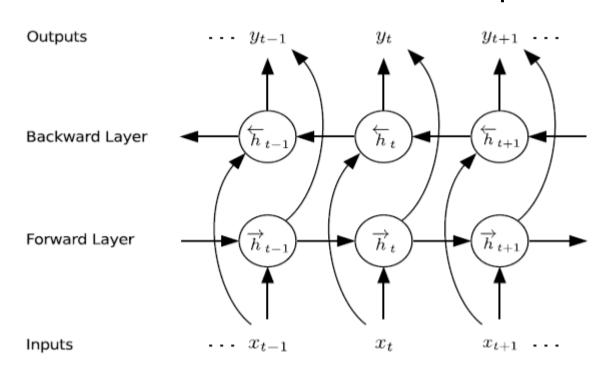
- Many NLP problems can be viewed as sequence labeling or sequence-to-sequence tasks
- "Recurrent" property → dynamical system over time



$$\mathbf{x} = (x_1, ..., x_T)$$
 $h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$
 $\mathbf{y} = (y_1, ..., y_T)$ $y_t = W_{hy}h_t + b_y$

Bidirectional RNN

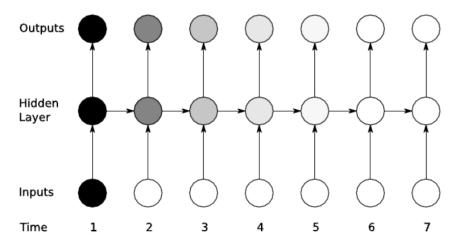
Exploit future context as well as past



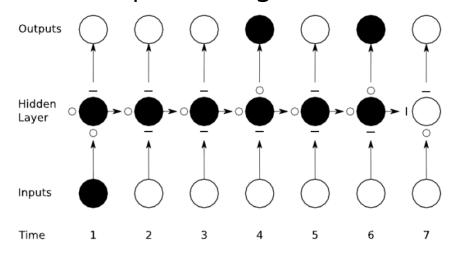
$$\begin{aligned} \overrightarrow{h_t} &= H \big(W_{x \overrightarrow{h}} x_t + W_{\overrightarrow{h} \overrightarrow{h}} \overrightarrow{h}_{t-1} + b_{\overrightarrow{h}} \big) \\ \overleftarrow{h_t} &= H \big(W_{x \overleftarrow{h}} x_t + W_{\overleftarrow{h} \overleftarrow{h}} \overleftarrow{h}_{t+1} + b_{\overleftarrow{t}} \big) \end{aligned} \qquad y_t = W_{\overrightarrow{h} y} \overrightarrow{h}_t + W_{\overleftarrow{h} y} \overleftarrow{h}_t + b_y$$

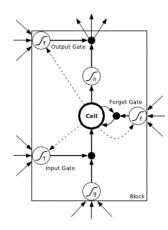
Long Short-Term Memory RNN

Vanishing Gradient Problem for RNN



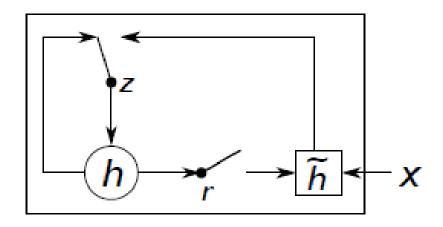
LSTM can preserve gradient information





- $i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$
- $f_t = \sigma (W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$
- $c_t = f_t c_{t-1} + i_t \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c)$
- $o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_0)$
- $h_t = o_t \tanh(c_t)$

Gated Recurrent Unit (GRU)



- $r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r)$
- $z_t = \sigma(W_{xx}x_t + W_{hz}h_{t-1} + b_z)$
- $\tilde{h}_t = \phi(W_{xh}x_t + W_{hh}(r_t \odot h_{t-1}) + b_h)$
- $h_t = z_t \odot h_t + (1 z_t) \odot \tilde{h}_t$
- $y_t = g(W_{hy}h_t + b_y)$

차례

- 자연어처리 소개
- 딥러닝 소개
- 딥러닝 기반의 자연어처리
 - Classification Problem
 - Sequence Labeling Problem
 - Sequence-to-Sequence Learning
 - Pointer Network
 - Machine Reading Comprehension

전이 기반의 한국어 의존구문분석

- **Transition-based(Arc-Eager): O(N)**
 - 의존구문분석 → 분류 문제

SBJ

MOD

OBJ

- 예: CJ그룹이 $_1$ 대한통운 $_2$ 인수계약을 $_3$ 체결했다 $_4$ ([..., w], B,A)
 - **[root]**, **[CJ그룹이**₁ 대한통운₂ ...**]**, {}
 - 1: **Shift**
 - [root **CJ그룹이₁**], [대한통운₂ 인수계약을₃ ...], {}
 - 2: **Shift**
 - [root CJ그룹이₁ 대한통운₂], [인수계약을 체결했다₄], {}
 - 3: Left-arc(NP_MOD)
 - [root CJ그룹이₁], [2←인수계약을₃ 체결했다₄], {(인수계약을₃→대한통운₂)}
 - 4: Shift
 - [root CJ그룹이₁ 2←인수계약을₃], [체결했다₄], {(인수계약을₃→대한통운₂)}
 - 5: Left-arc(NP OBJ)
 - [root CJ그룹이₁], [3←체결했다₄], {(체결했다₄→인수계약을₃), ...}
 - 6: Left-arc(NP SUB)
 - [root], [(1,3)←체결했다』], {(체결했다』→CJ그룹이₁), ...}
 - 7: Right-arc(VP)
 - [root→4 (1,3)←체결했다₄], [], {(root→체결했다₄), ...}

Configuration: (S, B, A) [S = Stack, B = Buffer, A = Arcs] $([ROOT], [w_1, ..., w_n], \{ \})$ Initial:

Terminal: (S,[],A)

Shift:

 $([\ldots, w_i], [w_i, \ldots], A) \implies ([\ldots, w_i, w_i], [\ldots], A)$

Reduce:

 \implies ([...], B, A)

Right-Arc(/):

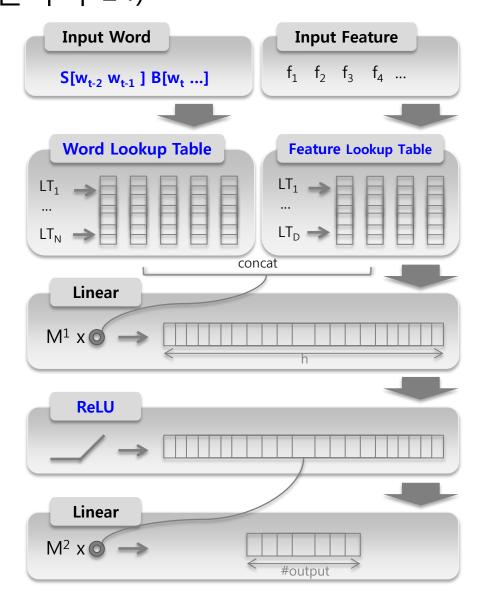
 $([\ldots,w_i],[w_j,\ldots],A) \implies ([\ldots,w_i,w_j],[\ldots],A\cup\{(w_i,w_j,l)\})$

Left-Arc(/):

 $([\ldots,w_i],[w_i,\ldots],A) \implies ([\ldots],[w_i,\ldots],A \cup \{(w_i,w_i,l)\})$

딥러닝 기반 한국어 의존구문분석 (한글 및 한국어 14)

- Transition-based + Backward
 - O(N)
 - 세종코퍼스 → 의존 구문 변환
 - 보조용언/의사보조용언 후처리
- Deep Learning 기반
 - ReLU(> Sigmoid) + Dropout
 - Korean Word Embedding
 - NNLM, Ranking(hinge, logit)
 - Word2Vec
 - Feature Embedding
 - POS (stack + buffer)
 - 자동 분석(오류 포함)
 - Dependency Label (stack)
 - Distance information
 - Valency information
 - Mutual Information
 - 대용량 코퍼스 → 자동 구문 분석



한국어 의존구문분석 실험 결과

기존 의존 구문 분석	UAS	LAS
이용훈[15]: 국어정보베이스	88.42	_
오진영[16]: 세종코퍼스	87.03	_
임수종[17]: 세종코퍼스	88.15	_
J.D. Choi[18]: 세종코퍼스	85.47	83.47
박정열[19]: 세종코퍼스	86.43	_
안광모[20]: 세종코퍼스	87.52	_
딥 러닝+전이기반 의존 구문 분석 (세종코퍼스, 자동 분석 형태소 이용)	UAS	LAS
ReLU+dropout	89.56	87.35
NNLM+ReLU+dropout	90.05	87.87
NNLM+ReLU+MI feat.	89.91	87.58
NNLM+ReLU+dropout+MI feat.	90.37	88.17
NNLM+sigmoid+MI feat.	89.94	87.64
NNLM+sigmoid+dropout+MI feat.	90.27	88.03
Ranking(hinge loss)+ReLU+dropout +MI feat.	90.19	88.01
Ranking(logit loss)+ReLU+dropout +MI feat.	90.31	88.11
Word2vec+ReLU+dropout+MI feat.	90.27	87.97

• 기존연구: UAS 85~88%

- Structural SVM 기반 성능:
 - UAS=89.99%
 - LAS=87.74%
- Pre-training > no Pre.
- Dropout > no Dropout
- ReLU > Sigmoid
- MI feat. > no MI feat.
- Word Embedding 성능 순위
 - 1. NNLM
 - 2. Ranking(logit loss)
 - 3. Word2vec
 - 4. Ranking(hinge loss)

한국어 상호참조해결

상호참조 (Coreference)

- 문서 내에서 이미 언급된 객체에 대하여 표현이 다른 단어로 다시 언급 하는 것
- Mention: 상호참조해결의 대상이 되는 모든 명사구(즉, 명사, 복합명사, 수식절을 포함한 명사구 등)를 의미
- Entity: 상호참조가 해결된 Mention들의 집합

Mention Detection 예제

- [[고양]에서 발생한 용오름]은 [토네이도]와 같은 것으로 [[[지상]의 뜨거운 공기]가 [[상층]의 찬 공기]와 갑자기 섞일 때] 발생합니다.
- [뜨거운 공기]가 빠르게 상승하고 [찬 공기]는 하강하면서 [[길다란 기 등] 모양의 구름]이 생겨나고 [[그] 안]에서 격렬한 [회오리바람]이 부는 겁니다.

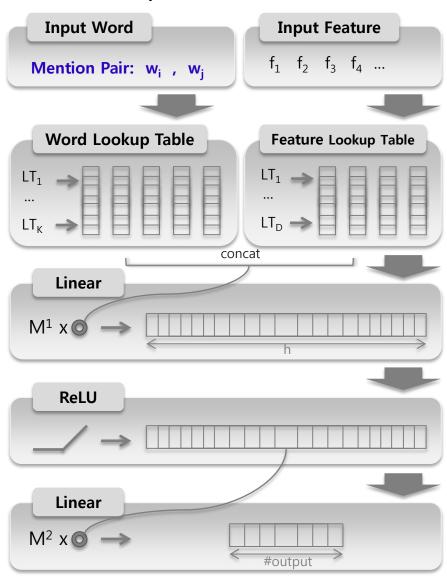
■ Entity 예제

- [지상의 뜨거운 공기], [뜨거운 공기]
- [상층의 찬 공기], [찬 공기]
- [길다란 기둥 모양의 구름], [그]

딥러닝 기반 한국어 상호참조해결

(KCC 15, ETRI Journal 16)

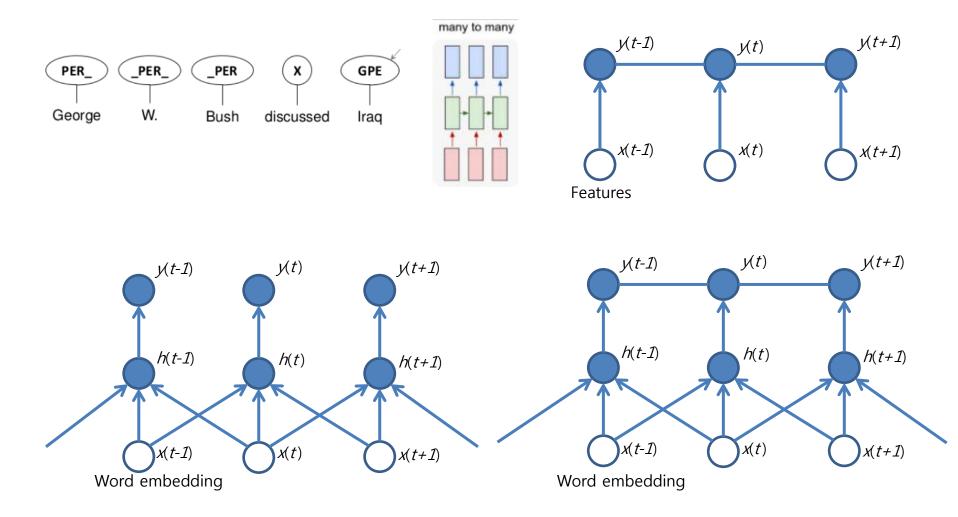
- 딥러닝 기반 Mention-Pair model
 - 상호참조해결 → 분류 문제
 - Korean word embedding + feature embedding (SVM과 동일)
 - Dropout + ReLU
- Data set
 - Test: 뉴스 1~20 문서
 - Training: 뉴스 21~100 문서, 장학퀴즈 QA 153개
- 성능
 - Deep Learning (MUC F1): 69.62%
 - No pre-training: F1 65.8%
 - SVM (MUC F1): 60.46%



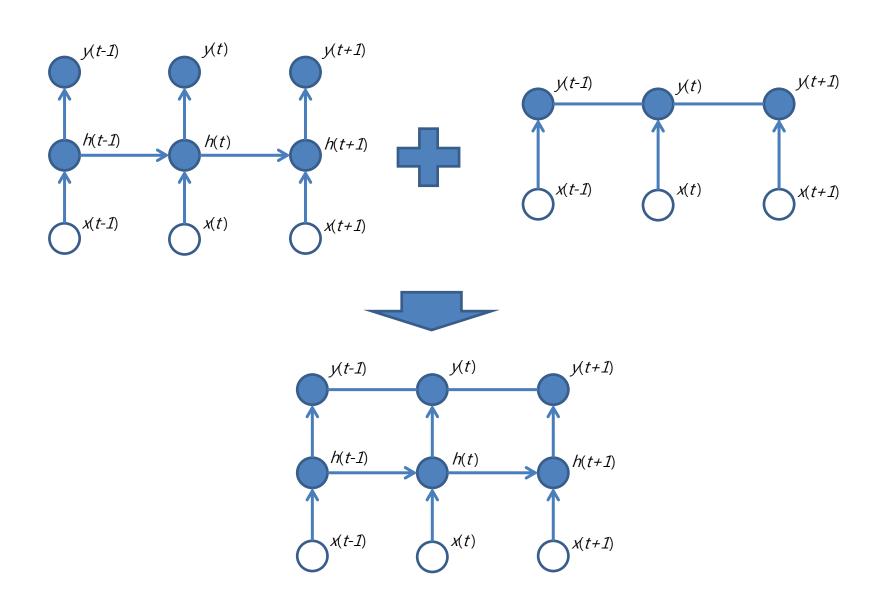
차례

- 자연어처리 소개
- 딥러닝 소개
- 딥러닝 기반의 자연어처리
 - Classification Problem
 - Sequence Labeling Problem
 - Sequence-to-Sequence Learning
 - Pointer Network
 - Machine Reading Comprehension

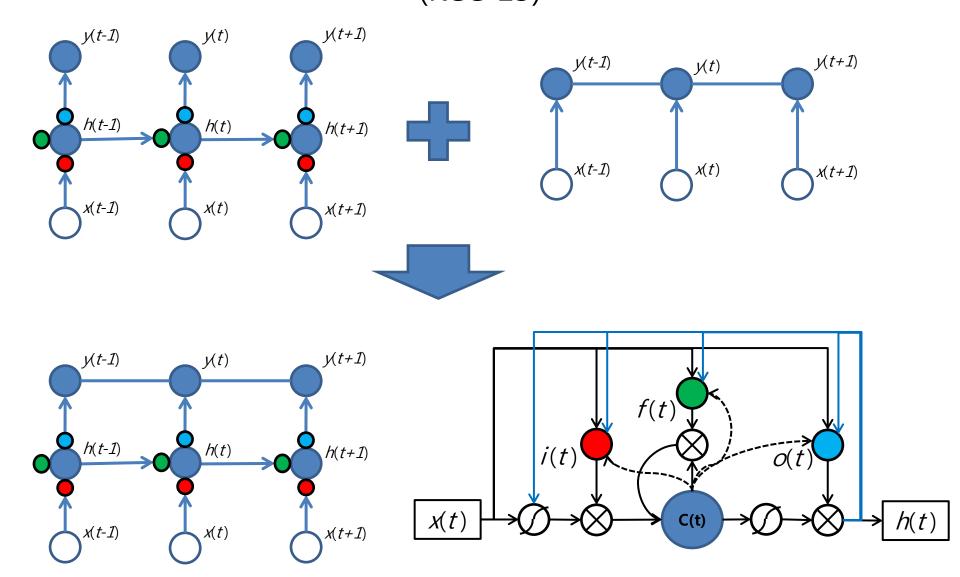
Sequence Labeling Tasks: CRF, FFNN(or CNN), CNN+CRF (SENNA)



RNN + CRF → Recurrent CRF



LSTM RNN + CRF → LSTM-CRF (KCC 15)

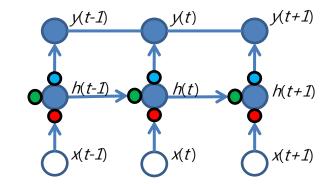


LSTM-CRF

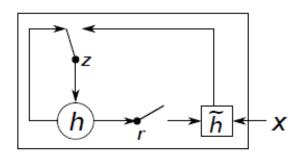
- $i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$
- $f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$

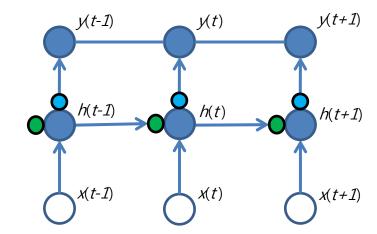


- $o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)$
- $h_t = o_t \odot \tanh(c_t)$
- $y_t = g(W_{hy}h_t + b_y)$ > 단어 단위로 학습
- $\bullet \quad y_t = W_{hy}h_t + b_y$
- $s(\mathbf{x}, \mathbf{y}) = \sum_{t=1}^{T} \{A(y_{t-1}, y_t) + y_t\}$
- $\log P(\mathbf{y}|\mathbf{x}) = s(\mathbf{x},\mathbf{y}) \log \sum_{\mathbf{y}'} \exp(s(\mathbf{x},\mathbf{y}'))$ \rightarrow 문장 단위로 학습



GRU-CRF

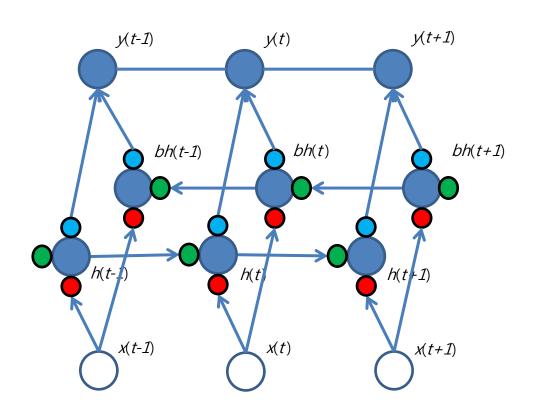




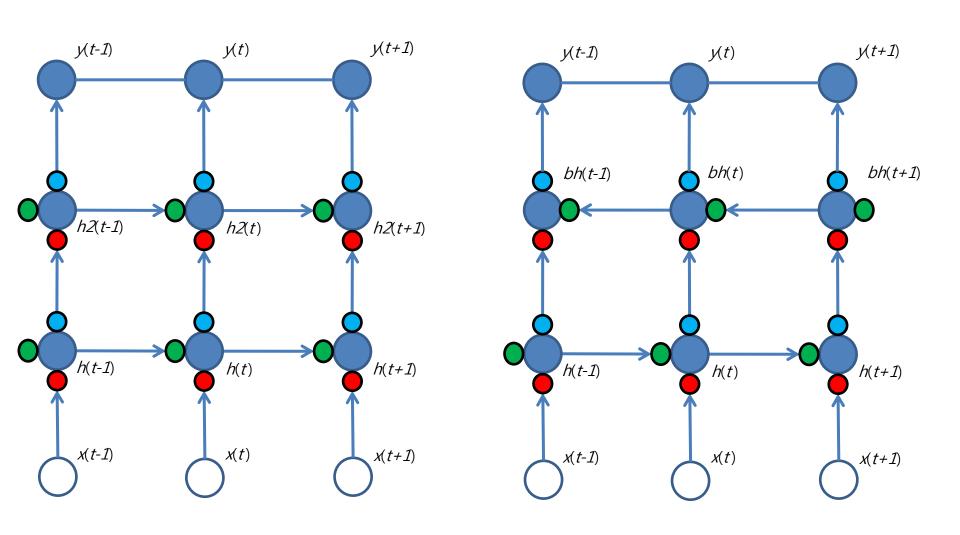
- $\bullet \quad r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r)$
- $\bullet \quad z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z)$
- $\tilde{h}_t = \phi(W_{xh}x_t + W_{hh}(r_t \odot h_{t-1}) + b_h)$
- $h_t = z_t \odot h_{t-1} + (\mathbf{1} z_t) \odot \tilde{h}_t$
- $y_t = g(W_{hy}h_t + b_y)$ > 단어 단위로 학습
- $y_t = W_{hy}h_t + b_y$
- $s(\mathbf{x}, \mathbf{y}) = \sum_{t=1}^{T} \{A(y_{t-1}, y_t) + y_t\}$
- $\log P(\mathbf{y}|\mathbf{x}) = s(\mathbf{x},\mathbf{y}) \log \sum_{\mathbf{y}'} \exp(s(\mathbf{x},\mathbf{y}'))$ \rightarrow 문장 단위로 학습

BI-LSTM CRF

- Bidirectional LSTM+CRF
- Bidirectional GRU+CRF
- Stacked Bi-LSTM+CRF ...



Stacked LSTM CRF



영어 개체명 인식

(KCC 15, IEICE Journal 17)

영어 개체명 인식 (CoNLL03 data set)	F1(dev)	F1(test)
SENNA (Collobert)	-	89.59
Structural SVM (baseline + Word embedding feature)	-	85.58
FFNN (Sigm + Dropout + Word embedding)	91.58	87.35
RNN (Sigm + Dropout + Word embedding)	91.83	88.09
LSTM RNN (Sigm + Dropout + Word embedding)	91.77	87.73
GRU RNN (Sigm + Dropout + Word embedding)	92.01	87.96
CNN+CRF (Sigm + Dropout + Word embedding)	93.09	88.69
RNN+CRF (Sigm + Dropout + Word embedding)	93.23	88.76
LSTM+CRF (Sigm + Dropout + Word embedding)	93.82	90.12
GRU+CRF (Sigm + Dropout + Word embedding)	93.67	89.98

한국어 개체명 인식

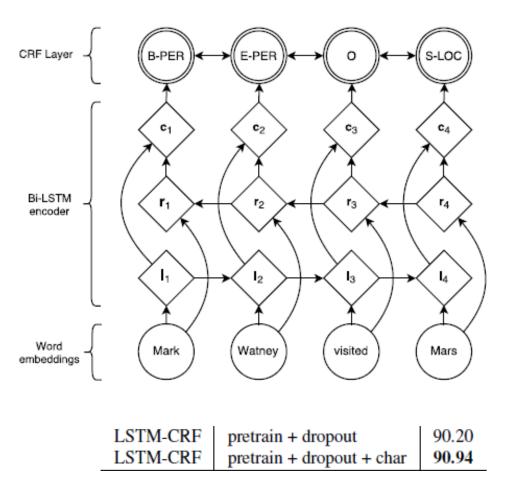
(IEICE Journal 17)

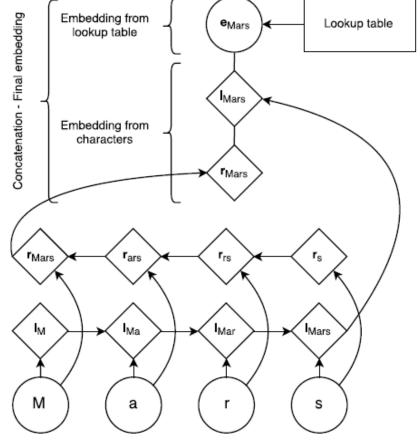
한국어 개체명 인식 (TV domain)	F1(test)
Structural SVM (baseline) (basic + NE dic. + word cluster feature + morpheme feature)	89.03
FFNN (ReLU + Dropout + Word embedding)	87.70
RNN (Tanh + Dropout + Word embedding)	88.93
LSTM RNN (Tanh + Dropout + Word embedding)	89.38 (+0.35)
Bi-LSTM RNN (Tanh + Dropout + Word embedding)	89.21 (+0.18)
CNN+CRF (ReLU + Dropout + Word embedding)	90.06 (+1.03)
RNN+CRF (Sigm + Dropout + Word embedding)	90.52 (+1.49)
LSTM+CRF (Sigm + Dropout + Word embedding)	91.04 (+2.01)
GRU+CRF (Sigm + Dropout + Word embedding)	91.02 (+1.99)

Neural Architectures for NER

(CMU Arxiv 16)

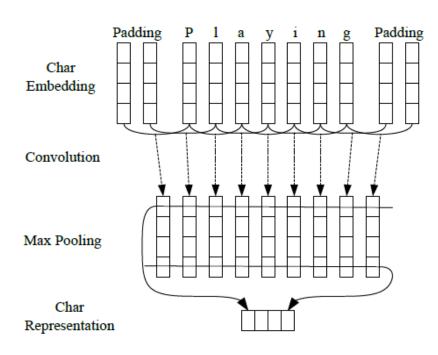
- LSTM-CRF model + Char-based Word Representation
 - Char: Bi-LSTM RNN



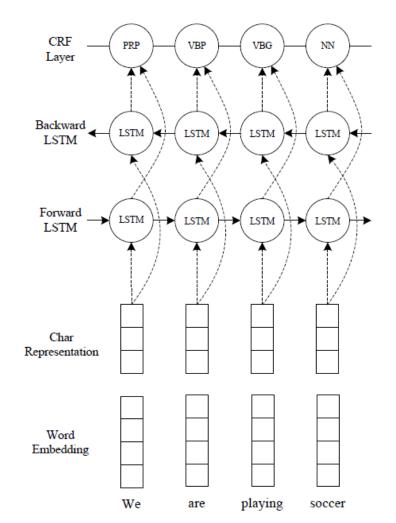


End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF (CMU ACL16)

- LSTM-CRF model + Char-level Representation
 - Char: CNN

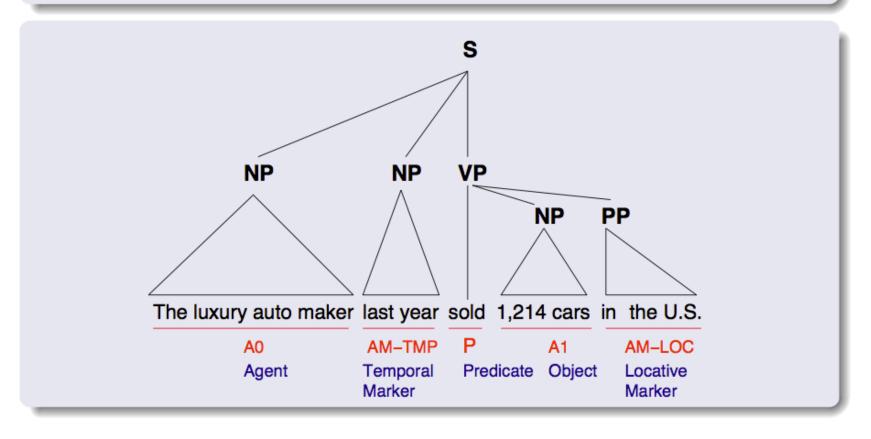


	NER					
	Dev			Test		
Model	Prec.	Recall	F1	Prec.	Recall	F1
BRNN					83.88	
BLSTM	92.31	90.85	91.57	87.77	86.23	87.00
BLSTM-CNN	92.52	93.64	93.07	88.53	90.21	89.36
BRNN-CNN-CRF	94.85	94.63	94.74	91.35	91.06	91.21



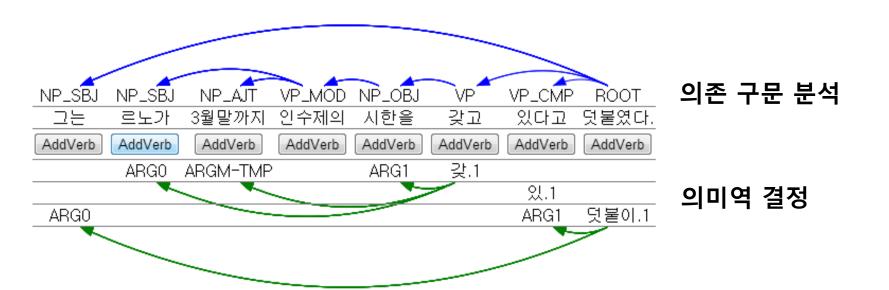
의미역 결정 (Semantic Role Labeling)

SRL ^{def} detecting basic event structures such as *who* did *what* to *whom*, *when* and *where* [IE point of view]



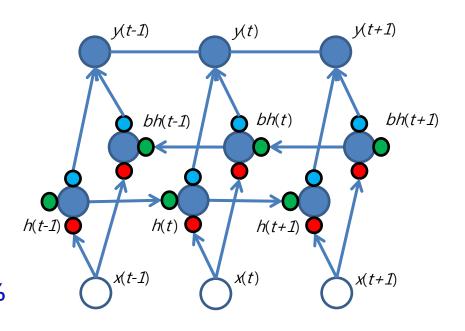
한국어 의미역 결정 (SRL)

- 서술어 인식(PIC)
 - 그는 르노가 3월말까지 인수제의 시한을 [갖고]_{갖.1} 있다고 [덧붙였다]_{덧붙.1}
- 논항 인식(AIC)
 - 그는 [르노가]_{ARG0} [3월말까지]_{ARGM-TMP} 인수제의 [시한을]_{ARG1} [갖고]_{갖.1} [있다고]_{AIIX} 덧붙였다
 - [그는]_{ARG0} 르노가 3월말까지 인수제의 시한을 갖고 [있다고]_{ARG1} [덧붙였다] 덧붙.1

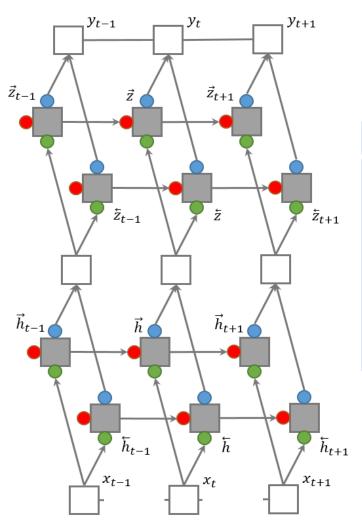


딥러닝 기반 한국어 의미역 결정 (한글 및 한국어 정보처리 15, 동계학술대회 15)

- Bidirectional LSTM+CRF
- Korean Word embedding
 - Predicate word, argument word
 - NNIM
- Feature embedding
 - POS, distance, direction
 - Dependency path, LCA
- Bi-LSTM+CRF 성능 (AIC)
 - F1: 78.2% (+1.2)
 - Backward LSTM+CRF: F1 77.6%
 - S-SVM 성능 (KCC14)
 - 기본자질: F1 74.3%
 - 기본자질+word cluster: 77.0%
 - 정보과학회 논문지 2015.02



Stacked BI-LSTM CRF 기반 한국어의미역결정 (정보과학회지 17)



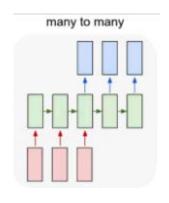
Syntactic information	w/	w/o
Structural SVM	76.96	74.15
FFNN	76.01	73.22
Backward LSTM CRFs	76.79	76.37
Bidirectional LSTM CRFs	78.16	78.17
Stacked Bidirectional LSTM CRFs (2 layers)	78.12	78.57
Stacked Bidirectional LSTM CRFs (3 layers)	78.14	78.36

차례

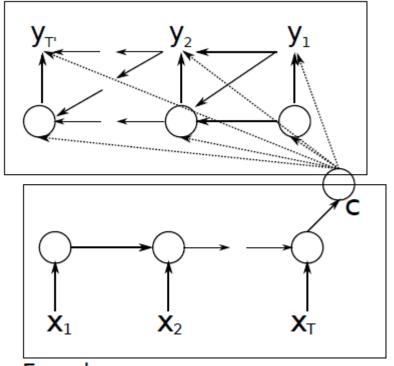
- 자연어처리 소개
- 딥러닝 소개
- 딥러닝 기반의 자연어처리
 - Classification Problem
 - Sequence Labeling Problem
 - Sequence-to-Sequence Learning
 - Pointer Network
 - Machine Reading Comprehension

Recurrent NN Encoder–Decoder for SMT

(Montreal U. EMNLP 14)



Decoder



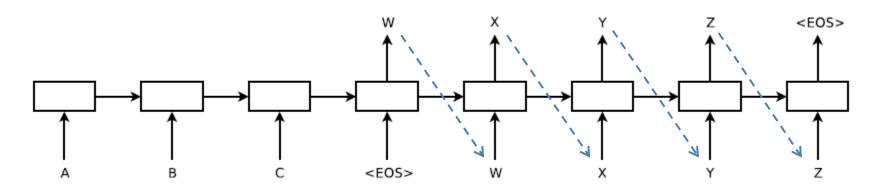
Encoder

Models	BLEU		
Wiodels	dev	test	
Baseline	30.64	33.30	
RNN	31.20	33.87	
CSLM + RNN	31.48	34.64	
CSLM + RNN + WP	31.50	34.54	

GRU RNN → Encoding GRU RNN → Decoding Vocab: 15,000 (src, tgt)

Source	Translation Model	RNN Encoder–Decoder
at the end of the	[a la fin de la] [f la fin des années] [être sup- primés à la fin de la]	[à la fin du] [à la fin des] [à la fin de la]
for the first time	[r © pour la premirère fois] [été donnés pour la première fois] [été commémorée pour la première fois]	

Sequence to Sequence Learning with Neural Networks (Google NIPS 14)



Method	test BLEU score (ntst14)
Bahdanau et al. [2]	28.45
Baseline System [29]	33.30
Single forward LSTM, beam size 12	26.17
Single reversed LSTM, beam size 12	30.59
Ensemble of 5 reversed LSTMs, beam size 1	33.00
Ensemble of 2 reversed LSTMs, beam size 12	33.27
Ensemble of 5 reversed LSTMs, beam size 2	34.50
Ensemble of 5 reversed LSTMs, beam size 12	34.81

Source Voc.: 160,000 **Target Voc.: 80,000**

Deep LSTMs with 4 layers

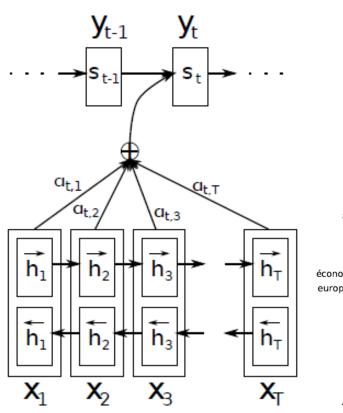
Train: 7.5 epochs

(12M sentences, 10 days with 8-

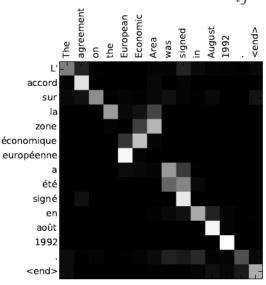
GPU machine)

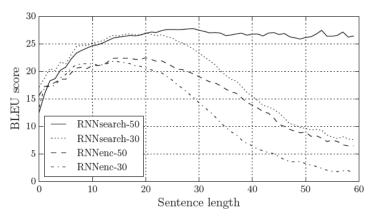
Method	test BLEU score (ntst14)
Baseline System [29]	33.30
Cho et al. [5]	34.54
State of the art [9]	37.0
Rescoring the baseline 1000-best with a single forward LSTM	35.61
Rescoring the baseline 1000-best with a single reversed LSTM	35.85
Rescoring the baseline 1000-best with an ensemble of 5 reversed LSTMs	36.5
Oracle Rescoring of the Baseline 1000-best lists	~45

Neural MT by Jointly Learning to Align and Translate (Jacobs U., Montreal U. ICLR 15)



$$c_{i} = \sum_{j=1}^{T_{x}} \alpha_{ij} h_{j}. \quad \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_{x}} \exp(e_{ik})},$$
$$e_{ij} = a(s_{i-1}, h_{i})$$





GRU RNN + **Attention** → Encoding

GRU RNN → Decoding

Vocab: 30,000 (src, tgt)

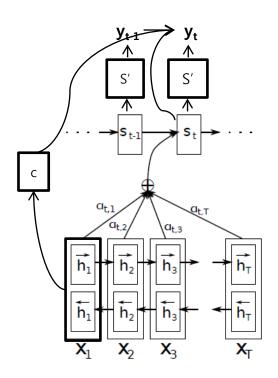
Train: 5 days

Model	All	No UNK°
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

문자 단위의 NMT

(WAT 15, 한글 및 한국어 15)

- 기존의 NMT: 단어 단위의 인코딩-디코딩
 - **미등록어** 후처리 or NMT 모델의 수정 등이 필요
- 문자 단위의 NMT
 - 입력 언어는 단어 단위로 인코딩
 - 출력 언어는 문자 단위로 디코딩
 - 단어 단위: その/UN 結果/NCA を/PS 詳細/NCD ...
 - 문자 단위: そ/B の/I 結/B 果/I を/B 詳/B 細/I ...
- 문자 단위 NMT의 장점
 - 모든 문자를 사전에 등록 **→ 미등록어 문제해결**
 - 기존 NMT 모델의 수정이 필요 없음
 - 미등록어 후처리 작업이 필요 없음



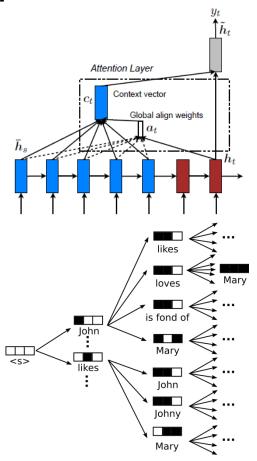
ASPEC E-to-J 실험 (WAT 15)

- ASPEC E-to-J data
- 성능 (Juman 이용 BLEU)
 - PB SMT: 27.48
 - HPB SMT: 30.19
 - Tree-to-string SMT: 32.63
 - NMT (Word-level decoding): 29.78
 - NMT (Character-level decoding): 33.14 (4위)
 - RIBES 0.8073 (2위)
 - Tree-to-String + NMT(Character-level) reranking
 - BLEU 34.60 (2위)
 - Human 53.25 (2위)

- This/DT:0 paper/NN:1 explaines/NNS:2 experimenta l/JJ:3 result/NN:4 according/VBG:5 to/TO:6 the/DT:7 model/NN:8 ./.:9 </s>:10
- **こ/B**:0 **の/I**:1 モ/B:2 デ/I:3 ル/I:4 に/B:5 よ/B:6 る/I:7 実/B:8 験/I:9 結/B:10 果/I:11 を/B:12 説/B:13 明/I:14 し/B:15 た/B:16 。/B:17 </s>:18

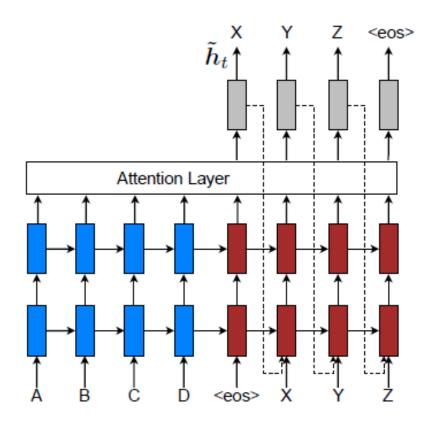
```
TIS
                                                 10
                 0.3
                                        0.7
 0:
 1: 0.2 0.1
                 0.1
                                   0.2 0.2
                                                 0.2
 2: 0.3 0.1
                                   0.1 0.2
                                                 0.2
 3:
                 0.1
                                                 0.3
                                   0.2 0.4
 4:
                                        0.9 0.1
 5:
                                            0.8
             0.1 0.2
 6: 0.1
                          0.2 0.2
                                                 0.1
 7: 0.2
             0.1 0.2
                          0.1 0.1
                                                 0.3
 8: 0.1
                 0.8
                                                 0.1
 9: 0.1
                 0.4 0.1
                                                 0.2
10:
                      1.0
11: 0.1
             0.5 0.1 0.1
                                                 0.1
12:
                          0.1
                                                 0.1
13:
             0.7 0.1
                                                 0.1
14:
             0.6 0.1
                                        0.1
                                                 0.1
                                                 0.5
15:
                                                 0.6
16: 0.1 0.1 0.2
17: 0.1 0.1 0.1
                                   0.1 0.1
                                                 0.5
18: 0.1 0.1 0.1
```

Input-feeding Approach (Stanford EMNLP 15)



The attentional decisions are made independently, which is suboptimal.

In standard MT, a coverage set is often maintained during the translation process to keep track of which source words have been translated.

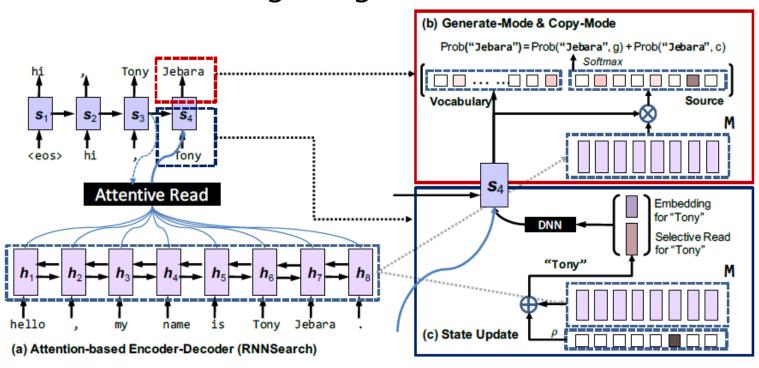


Effect:

- We hope to make the model fully aware of previous alignment choices
- We create a very deep network spanning both horizontally and vertically

Copying Mechanism or CopyNet

(Hong Kong U. ACL16)



|: Hello Jack, my name is Chandralekha.

R: Nice to meet you, Chandralekha.

I: This new guy doesn't perform exactly as we expected.

R: What do you mean by "doesn't perform exactly as we expected"?

$$\begin{split} p(y_t|\mathbf{s}_t, y_{t-1}, \mathbf{c}_t, \mathbf{M}) &= p(y_t, \mathbf{g}|\mathbf{s}_t, y_{t-1}, \mathbf{c}_t, \mathbf{M}) \\ &+ p(y_t, \mathbf{c}|\mathbf{s}_t, y_{t-1}, \mathbf{c}_t, \mathbf{M}) \quad (4) \\ p(y_t, \mathbf{g}|\cdot) &= \begin{cases} \frac{1}{Z} e^{\psi_g(y_t)}, & y_t \in \mathcal{V} \\ 0, & y_t \in \mathcal{X} \cap \bar{V} \quad (5) \quad \psi_g(y_t = v_i) = \mathbf{v}_i^\top \mathbf{W}_o \mathbf{s}_t, & v_i \in \mathcal{V} \cup \mathbf{UNK} \\ \frac{1}{Z} e^{\psi_g(\mathbf{UNK})} & y_t \notin \mathcal{V} \cup \mathcal{X} \end{cases} \\ p(y_t, \mathbf{c}|\cdot) &= \begin{cases} \frac{1}{Z} \sum_{j: x_j = y_t} e^{\psi_c(x_j)}, & y_t \in \mathcal{X} \\ 0 & \text{otherwise} \end{cases} \quad (6) \quad \psi_c(y_t = x_j) = \sigma\left(\mathbf{h}_j^\top \mathbf{W}_c\right) \mathbf{s}_t, \quad x_j \in \mathcal{X} \end{split}$$

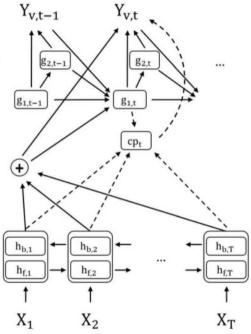
Abstractive Text Summarization

(한글 및 한국어 16)

지난 16일 트위터에는 로드킬로 숨진 친구를 안전한 곳에 옮긴 뒤 묵묵히 자리를 지킨 길고양이의 사진이 게재되며 많은 이들의 가슴을 먹먹하게 했다. 공개된 사진 속 고양이는 도로에서 차에 치여 처참히 쓰러진 친구를 발견하고 그 자리에 멈췄다. 앙상한 팔다리로 길고양이는 사력을 다해 친구를 안전한 곳으로 옮겼다. 고양이의 놀라운 행동은 여기서 끝이 아니었다. 고양이는 마치 떠난 친구를 애도하듯 하염없이 사체를 바라봤다. 이내 바로 옆에 자리를 잡고 묵묵히 친구의 결을 지켰다. 그 모습이 혼자 떠난 친구의 마지막 가는 길을 지켜주는 것 같아 보는 이들의 눈시울을 붉게 만들었다. 사진을 본 누리꾼들은 "비록 말 못하는 동물이지만 친구를 잃은 슬픔과 고통은 사람과 다르지 않은 것 같다"는 반응을 보였다.



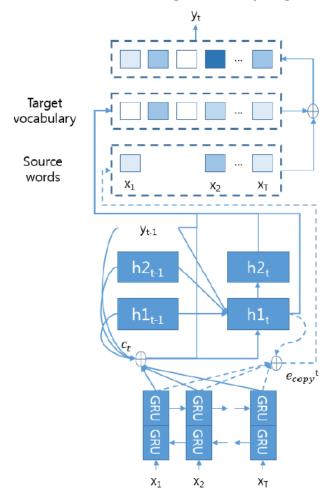
<u>로드킬로 숨진 친구의 곁을 지키는 길고양이의 모습이</u> 포착 되었다.



End-to-End 한국어 형태소 분석

(동계학술대회 16)

Attention + Input-feeding + Copying mechanism



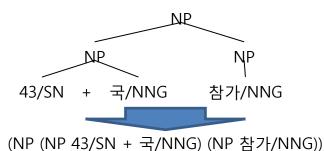
입력	오 늘 <sp> 경 기 에 서 도 <sp> 정 말 <sp> 잘</sp></sp></sp>
	했 다 .
	오 늘 <nng> <sp> 경 기 <nng> 에 서 <jkb> 도</jkb></nng></sp></nng>
출력	<jx><sp> 정 말 <mag> <sp> 잘 <mag> 하</mag></sp></mag></sp></jx>
	<xsv> 았 <ep> 다 <ef>. <sf></sf></ef></ep></xsv>

표 2. 모델 별 형태소 분석 성능(*은 평가셋이 동일하지 않음)

	개발데이터	평가데이터
CRF기반[5]*		97.65%*
S-SVM 기반[6]		98.03%
Sequence-to-sequence 모델 기반[8]*		97.15%*
RNN-search[2]	96.56%	95.92%
RNN-search[2]+Input- feeding	97.50%	96.87%
RNN-search[2]+Input- feeding+Copying mechanism	97.64%	97.08%

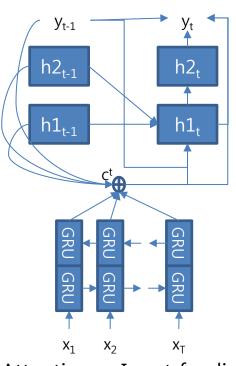
	형태소 분석 결과
정답	··· 도쿄/NNP ./SP 삿포로/NNP
0 0	등/NNB 의/JKG …
Copying	··· 도쿄/NNP ./SP 삿포로/NNP
적용	등/NNB 의/JKG …
Copying	··· 도쿄/NNP ./SP 엑포로/NNP
미적용	등/NNB 의/JKG …

Sequence-to-sequence 기반 한국어 구구조 구문 분석 (한글 및 한국어 16)



입력	선 생 <nng> 님 <xsn> 의 <jkg> <sp> 이 야 기 <nng> <sp> 끝 나 <vv> 자 <ec> <sp> 마 치 <vv> 는 <etm> <sp> 종 <nng> 이 <jks> <sp> 울 리 <vv> 었 <ep> 다 <ef> . <sf></sf></ef></ep></vv></sp></jks></nng></sp></etm></vv></sp></ec></vv></sp></nng></sp></jkg></xsn></nng>
정답	(S (S (NP_SBJ (NP_MOD XX) (NP_SBJ XX)) (VP XX)) (S (NP_SBJ (VP_MOD XX) (NP_SBJ XX)) (VP XX)))
RNN-search[7]	(S (VP (NP_OBJ (NP_MOD XX) (NP_OBJ XX)) (VP XX)) (S (NP_SBJ (VP_MOD XX) (NP_SBJ XX)) (VP XX)))
RNN-search + Input-feeding + Dropout	(S (S (NP_SBJ (NP_MOD XX) (NP_SBJ XX)) (VP XX)) (S (NP_SBJ (VP_MOD XX) (NP_SBJ XX)) (VP XX)))

	모델	F1
	스탠포드 구문분석기[13]	74.65
	버클리 구문분석기[13]	78.74
형태소의 음절 + 품사 태그 + <sp></sp>	RNN-search[7] (Beam size 10)	88.00
	RNN-search + Input-feeding	88.68
-11-4 · 15 P	RNN-search + Input-feeding + Dropout	89.03



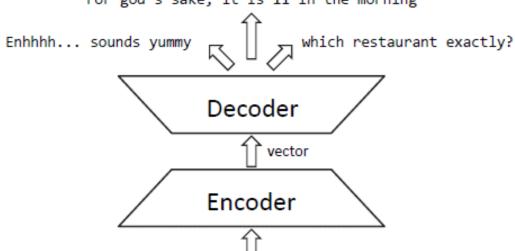
Attention + Input-feeding

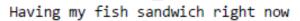
Neural Responding Machine for Short-Text Conversation (Huawei ACL 15)

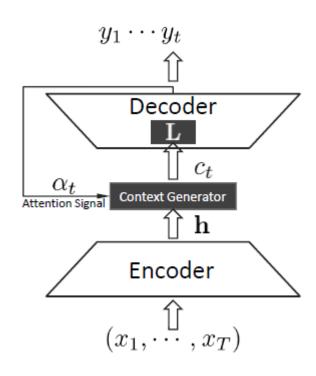
Post	Having my fish sandwich right now
UserA	For god's sake, it is 11 in the morning
UserB	Enhhhh sounds yummy
UserC	which restaurant exactly?

	#posts	219,905
Training	#responses	4,308,211
	#pairs	4,435,959
Test Data	#test posts	110

For god's sake, it is 11 in the morning

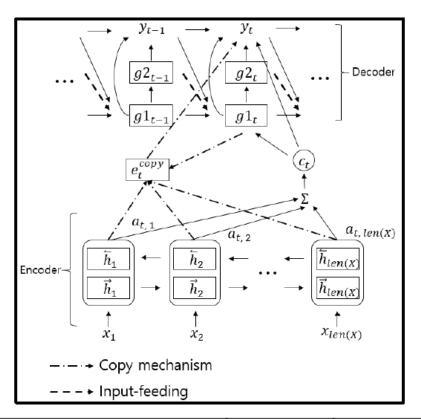






Short-Text Conversation

(동계학술대회 16)



모델	BLEU-1	ROUGE-1
GRU-Search (baseline)	1.8	20.10
GRU-Search + input-feeding	3.59 (+1.79)	21.40 (+1.30)
GRU-Search + input-feeding + copy mechanism	3.02 (+1.22)	21.74 (+1.64)

- Data: 클리앙 '아무거나 질문 게시판'
 - 77,346 질문-응답 쌍
 - 학습:개발:평가 = 8:1:1

질의: 종합소득세 문의(프리랜서 디자이너) 안녕하세요 저는 디자이너 프리랜서로 일하고 있습니다.보통은 일한 내역을 업체에서 3.3% 제외코 입금 받는데 자동차 명의는 아내의 이름으로 되어있는데 자동차 보험료를 제 <u>카드로 지불</u>하면 이것도 <u>공제 대상</u>이 되는지 궁금합니다.

사람이 단 응답: 사업자 내시고 비용처리하는거 아닌이상, 카드 사용 금액은 하나도 도움이 안됩니다.

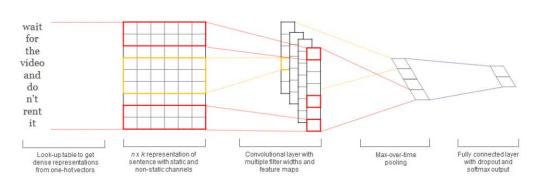
GRU search: 네 맞습니다. 제가 아는 것이 맞는 걸로 알고 있습니다. 자세한 것은 고객센터에 문의해보시는 것이 정확할 것 같습니다.

GRU search + input-feeding: 제가 알기로는 해주지 않는 걸로 알고 있습니다. 자세한 건 고객센터에 문의해보시는 것이 좋을 것같습니다.

GRU search + input-feeding + copy mechanism: <u>공제대상</u>이 아 니라 <u>카드</u>로 <u>지불</u>하는 것이 아닌 걸로 알고 있습니다. 자세한 것 은 고객 센터에 문의해보시는 것이 좋을 것 같습니다.

CNN 기반 한국어 감성 분석 (KCC 16)

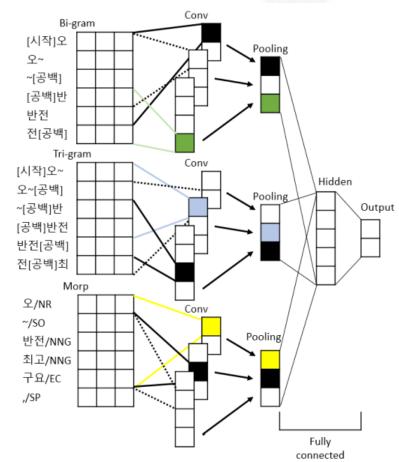
- 한국어 감성 분석: 문장 → 긍정 or 부정 (분류 문제)
- EMNLP14 CNN 모델 확장
 - 한국어 특징 반영



<한국어 영화평 감성 분석 데이터 구축>

	Emb_drop	Hidden	Hid_act	Hid_drop	Accuracy
Bi(baseline)	0.5	50	sigmoid	0.5	90.96%
Tri(baseline)	0.5	100	sigmoid	0.5	90.67%
Morp(baseline)	0.5	250	sigmoid	0	91.35%
Bi + Tri	0.5	250	sigmoid	0.5	90.84%
Bi + Morp	0.5	25	relu	0.25	91.52%
Tri + Morp	0.5	250	sigmoid	0.25	91.44%
Bi + Tri + Morp	0.5	25	relu	0	91.52%

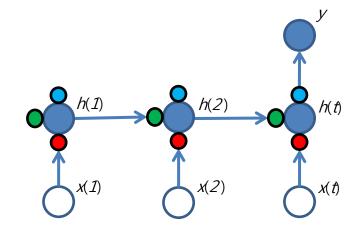




many to one

LSTM RNN 기반 한국어 감성분석

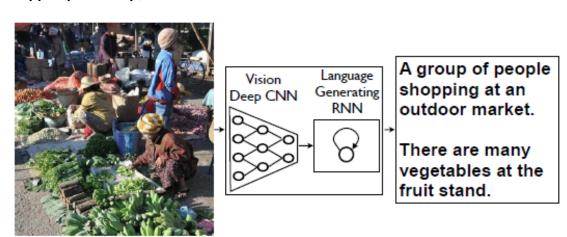
- LSTM RNN-based encoding
 - Sentence embedding → 입력
 - Fully connected NN → 출력
 - GRU encoding 도 유사함

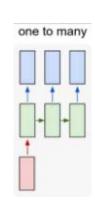


Data set	Model	Accuracy
Mobile	SVM (word feature)	85.58
Train: 4543 Test: 500	CNN(EMNLP14: relu,kernel3,hid50)	91.20
	GRU encoding + Fully connected NN	91.12
	LSTM RNN encoding + Fully connected NN	90.93

이미지 캡션 생성

- 이미지 내용 이해 → 이미지 내용을 설명하는 캡션 자동 생성
 - 이미지 인식(이해) 기술 + 자연어처리(생성) 기술
- 활용 분야
 - 이미지 검색
 - 맹인들을 위한 사진 설명, 네비게이션
 - 유아 교육, ...

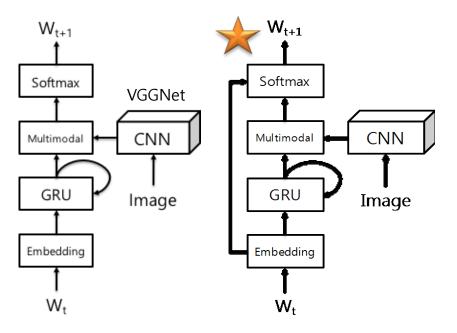


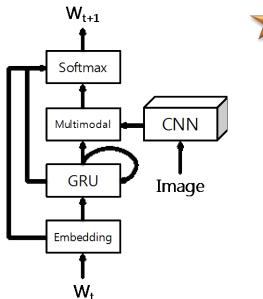


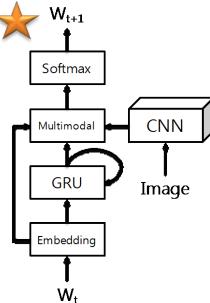
RNN을 이용한 이미지 캡션 생성 (동계학술대회 15)

Flickr 8K	B-1	B-2	B-3	B-4
m-RNN (Baidu)[2]	56.5	38.6	25.6	17.0
DeepVS (Stanford)[5]	57.9	38.3	24.5	16.0
NIC (Google)[4]	63.0	41.0	27.0	-
Ours-GRU-DO1	63.12	44.27	29.82	19.34
Ours-GRU-DO2	61.89	43.86	29.99	19.85
Ours-GRU-DO3	62.63	44.16	30.03	19.83
Ours-GRU-DO4	63.14	45.14	31.09	20.94

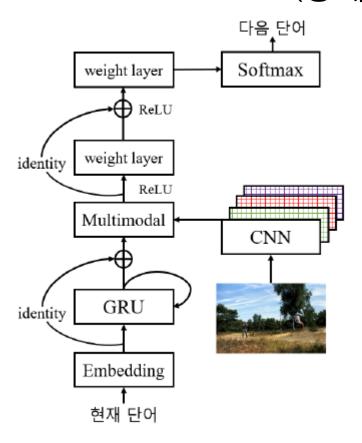
Flickr 30K	B-1	B-2	B-3	B-4
m-RNN (Baidu)[2]	60.0	41.2	27.8	18.7
DeepVS (Stanford)[5]	57.3	36.9	24.0	15.7
NIC (Google)[4]	66.3	42.3	27.7	18.3
Ours-GRU-DO1	63.01	43.60	29.74	20.14
Ours-GRU-DO2	63.24	44.25	30.45	20.58
Ours-GRU-DO3	62.19	43.23	29.50	19.91
Ours-GRU-DO4	63.03	43.94	30.13	20.21







Residual Net + 한국어 이미지 캡션 생성 (동계학술대회 16)



Residual Net: F(x) weight layer identity x H(x) = F(x) + x F(x) = F(x) + x

표 3. 한국어 이미지 캡션 생성 실험 결과

모델	B-1	B-2	B-3	B-4
GRU-D01[4]	68.64	54.42	44.07	35.50
GRU-D02[4]	68.56	54.74	44.52	36.02
GRU-D03[4]	67.89	53.68	43.25	34.77
GRU-D04[4]	68.56	54.65	44.18	35.51
GRU-ResNet	68.96	54.93	44.61	36.14



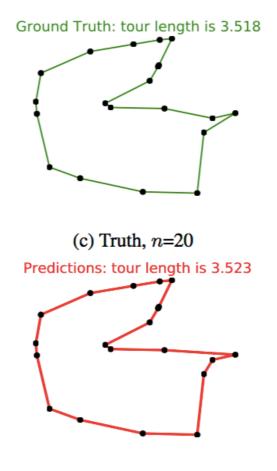
- (a) 한 남자가 카메라를 향해 미소를 짓고 있다.
- (b) 한 남자가 낚시를 하고 있다
- (c) 한 무리의 사람들이 분주한 거리에 서 있다.
- (d) 두 개들이 잔디를 관통하여 달리는 중이다.

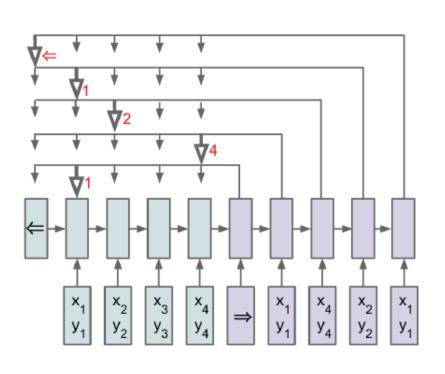
차례

- 자연어처리 소개
- 딥러닝 소개
- 딥러닝 기반의 자연어처리
 - Classification Problem
 - Sequence Labeling Problem
 - Sequence-to-Sequence Learning
 - Pointer Network
 - Machine Reading Comprehension

Pointer Network (Google NIPS 15)

- Travelling Salesman Problem: NP-hard
- Pointer Network can learn approximate solutions: O(n^2)



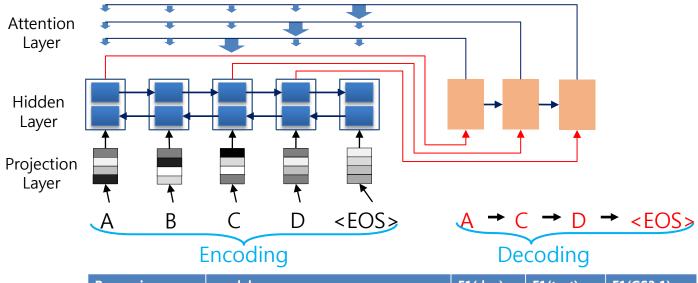


(f) Ptr-Net, m=5-20, n=20

포인터 네트워크 기반 상호참조해결

(KCC 16, Pattern Recognition Letters 17)

- 상호참조해결: "A씨는... B씨는 ... 그는 ..." → 그: A or B?
- **입력**: 단어(형태소) 열, 출발점(대명사, 한정사구(이 별자리 등))
 - X = {A:0, B:1, C:2, D:3, <EOS>:4}, Start_Point=A:0
- **출력**: 입력 단어 열의 위치(Pointer) 열 → Entity
 - $Y = \{A:0, C:2, D:3, <EOS>:4\}$
- 특징: End-to-end 방식의 대명사 상호참조 해결 (mention detection 과정 X)



Beam size	model	F1(dev)	F1(test)	F1(GS3.1)
1	BI_GRU_concat h=[400,400]	84.37		
3	BI_GRU_concat h=[400,400]	85.50		
5	BI_GRU_concat h=[400,400]	85.83	82.38	83.95
10	BI_GRU_concat h=[400,400]	85.83		
Rule-bas	sed system (Multi-pass sieve)	-	55.49	62.77

결과 예제

- 입력
 - 입력문장: 우리:0 나라:1 국회:2 에서:3 의결:4 되:5 ㄴ:6 법률:7 안:8 은:9 정부:10 로:11 이송:12 후:13 이:14 기한:15 내:16 에:17 대통령:18 이:19 공포:20 하:21 ㅁ:22 으로 써:23 확정:24 되:25 ㄴ다:26 ::27 헌법:28 에:29 명시:30 되:31 ㄴ:32 이:33 기한:34 은:35 며칠:36 이:37 ㄹ까:38 ?:39 < EOS >:40
 - 출력열 출발점: 이 기한:15
- 출력 정답 (Coref0 순서)
 - 이_기한:15 (출발점) → 이_기한:34 → 며칠:36 → <EOS>:40
- Attention score (100점)
 - 이_기한:15 **→ 이_기한:34**
 - 이 기한 내:16 (3), 헌법:28 (1), 이 기한:34 (80), 며칠:36 (10), <EOS>:40 (2)
 - 이_기한:34 **→ 며칠:36**
 - **며칠:36 (89)**, <EOS>:40 (9)
 - 며칠:36 → **<EOS>:40**
 - <EOS>:40 (99)
- 참고: 규칙기반 결과
 - {이_기한:15, 이_기한:34} 며칠:36 생략 됨
 - {법률_안:8, 며칠:36} (X)

포인터 네트워크를 이용한 한국어 의존구문분석 (동계학술대회 16)

 SBJ
 MOD
 OBJ

 CJ그룹이
 대한통운
 인수계약을
 체결했다

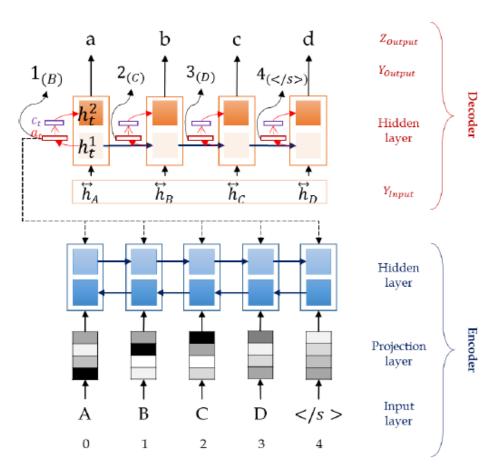


표 1. 의존 구문 분석 디코더 입력 기준

Input sentence

나/NP:201[0] <sp>:2[1] 당신/NP:2834[2] 의/JKG:9[3] <sp>:2[4] 그림자 /NNG:6803[5] 를/JKO:12[6] <sp>:2[7] 밟/VV:2241[8] 으며/EC:163[9] <sp>:2[10] 가/VV:262[11] ㅂ니다/EF:213[12] ./SF:3[13] </s>:1[14]

Input criteria	Input explanation
	입력되는 각 어절의 시작 형태소의 인덱스
Dp0	Y _{input} ={나/NP[0], 당신/NP[2],그림자/NNG[5], 밟
	/VV[8],
	Y _{output} ={빏/VV[8], 그림자/NNG[5], 빏/VV[8], 가
	/VV[11], [14]}
	$Z_{output} = \{NP_{SBJ}, NP_{MOD}, NP_{OBJ}, VP, VP\}$

표 3. 한국어 의존 구문 분석 성능 비교(자동 분석 형태소 이용)

Dependency parsing	UAS	LAS
이창기[4] with MI	90.37	88.17
이건일[5]	90.33	-
나승훈[6]: String-to-dependency SMT 모델	88.16	-
나승훈[7]: Stack LSTM 기반	90.11	87.70
나승훈[8]: Stack LSTM + 음절과 형태소 결합	90.44	87.93
Our: Dp0 concat2 [800, 400] dropout[0.1, 0.1]	91.36	88.67
Our: Dp1 concat2 [400, 200] dropout[0.1, 0.1]	91.46	89.05
Our: Dp1 concat2 [400, 200] dropout[0.2, 0.2]	91.55	89.16

차례

- 자연어처리 소개
- 딥러닝 소개
- 딥러닝 기반의 자연어처리
 - Classification Problem
 - Sequence Labeling Problem
 - Sequence-to-Sequence Learning
 - Pointer Network
 - Machine Reading Comprehension

SQuAD

- Stanford Question
 Answering Dataset (SQuAD)
 - Reading comprehension dataset
 - 100,000+ questions
 - 536 Wikipedia articles
 - 23,215 passages
 - 107,785 questions

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall? gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail? graupel

Where do water droplets collide with ice crystals to form precipitation?

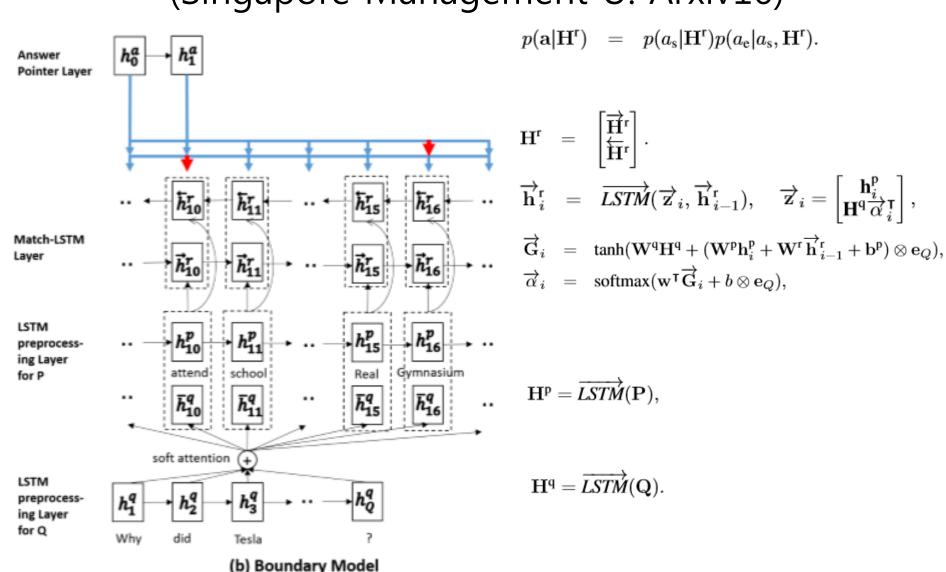
within a cloud

SQuAD Leaderboard

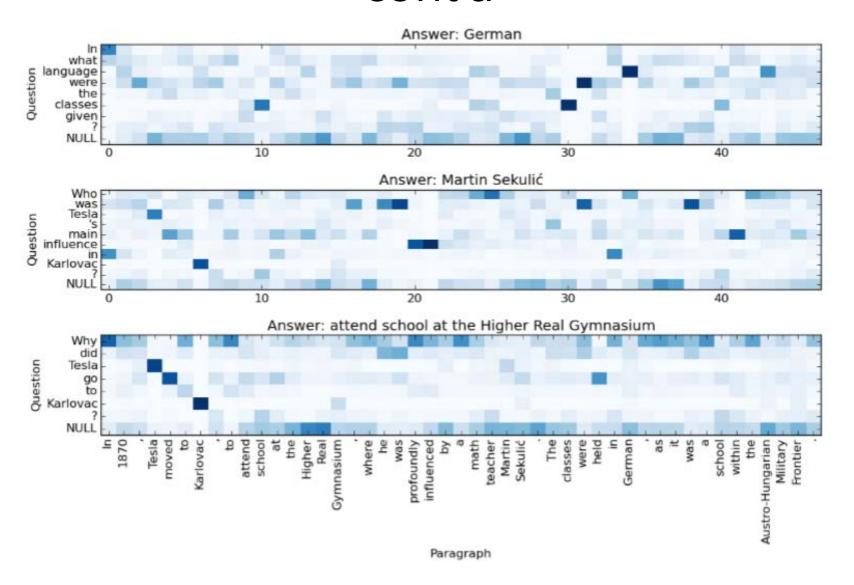
Rank	Model	EM	F1
1 Mar 2017	r-net (ensemble) Microsoft Research Asia	76.922	84.006
2 Mar 2017	ReasoNet (ensemble) MSR Redmond	75.034	82.552
3 Feb 2017	BiDAF (ensemble) Allen Institute for AI & University of Washington https://arxiv.org/abs/1611.01603	73.744	81.525
4 Apr 2017	SEDT+BiDAF (ensemble) CMU https://arxiv.org/abs/1703.00572	73.586	81.393
4 Jan 2017	Multi-Perspective Matching (ensemble) IBM Research https://arxiv.org/abs/1612.04211	73.765	81.257
5 Apr 2017	T-gating (ensemble) Peking University	72.758	81.001
6 Apr 2017	r-net (single model) Microsoft Research Asia	72.338	80.717
7 Nov 2016	Dynamic Coattention Networks (ensemble) Salesforce Research https://arxiv.org/abs/1611.01604	71.625	80.383
7 Apr 2017	QFASE NUS	71.898	79.989
8 Apr 2017	Interactive AoA Reader (single model) Joint Laboratory of HIT and iFLYTEK Research	71.153	79.937

9 Mar 2017	jNet (single model) USTC & National Research Council Canada & York University https://arxiv.org/abs/1703.04617	70.607	79.821
9 Apr 2017	Ruminate Reader (single model) New York University	70.639	79.456
10 Mar 2017	ReasoNet (single model) MSR Redmond	70.555	79.364
10 Mar 2017	Document Reader (single model) Facebook AI Research	70.733	79.353
10 Dec 2016	FastQAExt German Research Center for Artificial Intelligence https://arxiv.org/abs/1703.04816	70.849	78.857
11 Apr 2017	Multi-Perspective Matching (single model) IBM Research https://arxiv.org/abs/1612.04211	70.387	78.784
12 Apr 2017	SEDT+BiDAF (single model) CMU https://arxiv.org/abs/1703.00572	68.478	77.971
12 Mar 2017	RaSoR (single model) Google NY, Tel-Aviv University https://arxiv.org/abs/1611.01436	69.642	77.696
13 Apr 2017	T-gating (single model) Peking University	68.132	77.569
14 Nov 2016	BiDAF (single model) Allen Institute for AI & University of Washington	67.974	77.323
Helifa Established Association (1995)	Allen Institute for AI & University of Washington	68.436	77.323
Nov 2016	Allen Institute for AI & University of Washington		
14 Dec 2016	FastQA German Research Center for Artificial Intelligence https://arxiv.org/abs/1703.04816 Match-LSTM with Ans-Ptr (Boundary) (ensemble)		
14 Dec 2016	FastQA German Research Center for Artificial Intelligence https://arxiv.org/abs/1703.04816	68.436	77.07
14 Dec 2016	FastQA German Research Center for Artificial Intelligence https://arxiv.org/abs/1703.04816 Match-LSTM with Ans-Ptr (Boundary) (ensemble) Singapore Management University	68.436	77.07

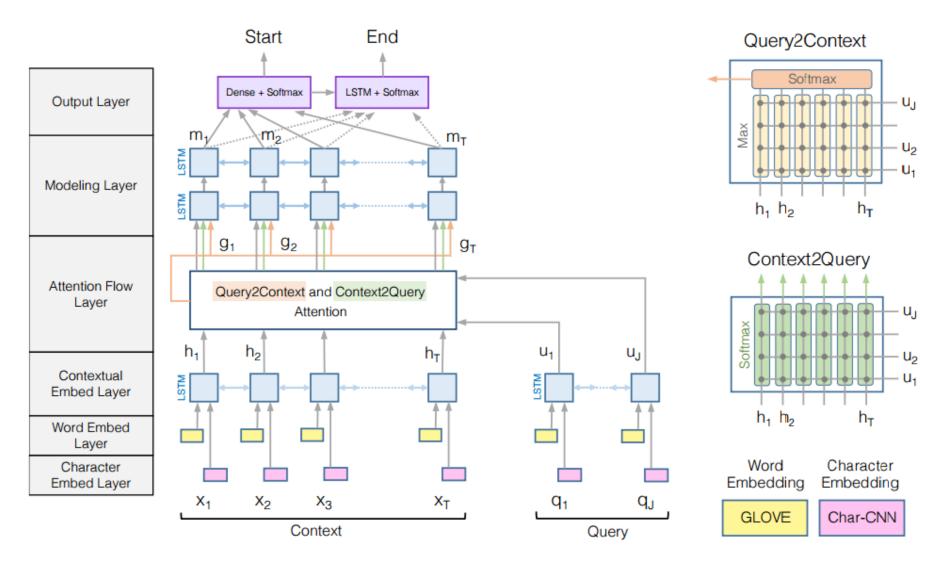
Match LSTM with Answer Pointer (Singapore Management U. Arxiv16)



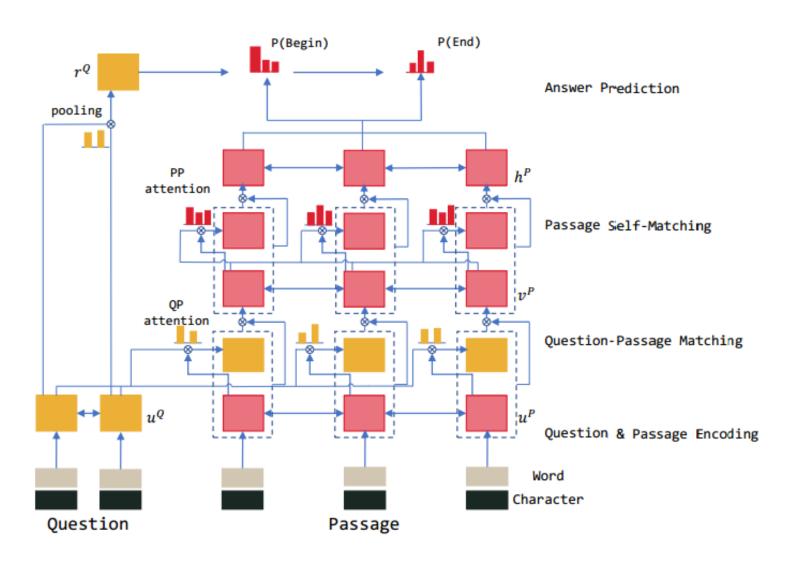
Match LSTM with Answer Pointer – cont'd



Bi-directional Attention Flow Model (Allen Institute Arxiv16)



R-Net: SQuAD 1st (Microsoft Research Asia)



개발 현황

- Match LSTM, BIDAF, R-Net 모델 자체 구현 완료
 - Theano를 이용하여 자체 구현 → baseline model
 - 다양한 변형 모델 개발 중
- 한국어 데이터 실험 중
 - 기업체에서 한국어 데이터 구축 중
 - 다양한 모델을 한국어 데이터에 적용 중

한국어 실험 - 1/2

• 한국어 데이터셋

- 학습: 16,421 질문셋

- 평가: 1,825 질문셋

성능

Match-LSTM: F1=48.30%, EM=31.62%

- R-Net: F1=45.41%, EM=29.42%

- BIDAF: F1=? %, EM=? %

한국어 결과 데이터 예제 - 1/3

• 뉴스 기사

- 이희솔/nnp[0] 은/jx[1] 1/sn[2] 차/nnb[3] 시기/nng[4] 에서/jkb[5] 155/sn[6] kg/nnb[7] 을 /jko[8] 가뿐/xr[9] 하/xsa[10] 게/ec[11] 들/vv[12] 어/ec[13] 올리/vv[14] 었/ep[15] 다 /ef[16] ./sf[17] ...
- 리우/nnp[53] 올림픽/nng[54] 역도/nng[55] 여자/nng[56] 75/sn[57] kg/nnb[58] 이상 /nng[59] 급/nng[60] 그룹/nng[61] A/sl[62] 경기/nng[63] 에서/jkb[64] ...
- 이/mm[106] 경기/nng[107] 에서/jkb[108] 는/jx[109] 중국/nnp[110] 의/jkg[111] <mark>명수핑 /nng[112]</mark> 이/jks[113] **금메달/nng[114]** 을/jko[115] 차지/nng[116] 하/xsv[117] 았/ep[118] 다/ef[119] ./sf[120] **명수핑/nng[121]** 은/jx[122] 인상/nng[123] 130/sn[124] kg /nnb[125] ,/sp[126] 용상/nng[127] 177/sn[128] kg/nnb[129] ,/sp[130] 합계/nng[131] 307/sn[132] kg/nnb[133] 을/jko[134] 기록/nng[135] 하/xsv[136] 았/ep[137] 다 /ef[138] ./sf[139]
- 은메달/nng [140] 을/jko[141] ...
- 4/sn[228] 번/nnb[229] 째/xsn[230] 메달/nng[231] 을/jko[232] 수확/nng[233] 하/xsv[234] 있/ep[235] 다/ef[236] ./sf[237] </s>[238]

• 질문

- 리우 올림픽 여자 역도 75kg 우승자는 누구야?
- Key: 멍수핑/nng Sys: 멍수핑/nng

한국어 결과 데이터 예제 - 2/3

• 잘된 것

- 누구
 - Q: 리우 올림픽 여자 역도 75kg 우승자는 누구야?
 - Key: 멍수핑/nng

Sys: 멍수핑/nng

- 무엇
 - Q: 뉴로피드백은 무슨 시스템이야?
 - Key: 뇌파를 조절하여 인간의 능력을 극대화 하는 시스템
 - Sys: 뇌파를 조절하여 인간의 능력을 극대화 하는 시스템
- 얼마?
 - Q: 화도진스케이트장 이용료는 얼마야?
 - Key: 1천원

Sys: 1천원

- 이유
 - Q; 리우 올림픽에서 박태환의 성적이 부진했던 이유가 뭐야?
 - Key: 그동안 많은 경기를 뛰지 못했다.
 - Sys: 많은 경기를 뛰지 못했다.
- 방법
 - Q: 리우올림픽 측은 초록색 수영장 물을 어떻게 처리했어?
 - Key: 하수관으로 배출돼 조정, 카누, 트라이애슬론 경기장의 물로 활용
 - Sys: 하수관으로 배출돼 조정, 카누, 트라이애슬론 경기장의 물로 활용

한국어 결과 데이터 예제 - 3/3

• 잘 안된 것

- 왜
 - Q: 롯데는 이번 시즌에 왜 성적이 부진했어?
 - Key: 주축 선수들의 줄부상
 - Sys: 좌측 정강이 뼈 분쇄 골절 진단
- 무엇
 - Q: 두산베어스의 팀 특징은 뭐야?
 - Key: 공격적이고 적극적인 타격
 - Sys: 강한 팀으로 거듭나는 이유 중 하나는 견고한 수비
- 언제
 - Q: 류현진 메이저리그 데뷔전이 언제였어?
 - Key: 2013년 4월 3일
 - Sys: 2013년
- 무엇
 - Q: 전국기능대회 입상자는 어떤 혜택이 있어?
 - Key: 금메달 1200만원, 은메달 800만원, 동메달 400만원의 상금과 함께 해당직종 산업 기사 실기시험 면제, 세계대회 출전 기회 및 기능경기대회 후원업체 취업기회 등
 - Sys: 기술 전문가로 자리 잡을 수 있도록