Spoken Language Understanding
(focusing on Speech-act Classification)

Ko, Youngjoong

Dept. of Computer Engineering,
Dong-A University

Contents

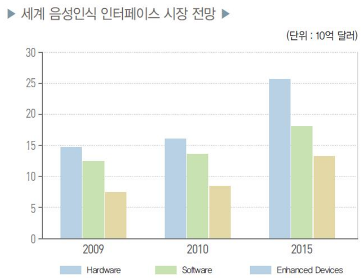
Basic Concepts of SLU

❖ 사용자의 음성을 이용한 자연어 기반의 인터페이스 사용의 증가

- Smart phone, Robot etc

Basic Concepts of SLU

❖ 음성 인식 인터페이스의 전망 및 규모



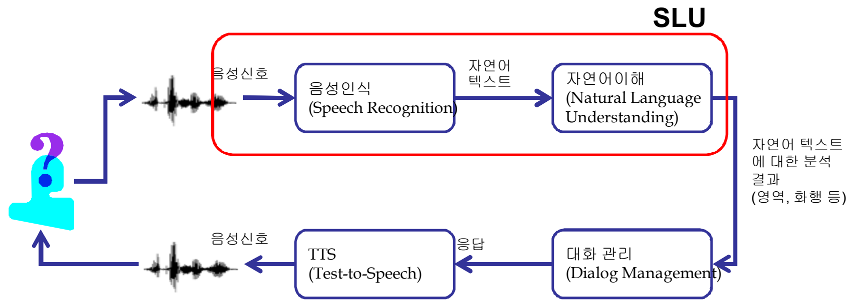
▶ 음성인식 인터페이스 국내 시장 규모 ▶

(단위 : 억 원)

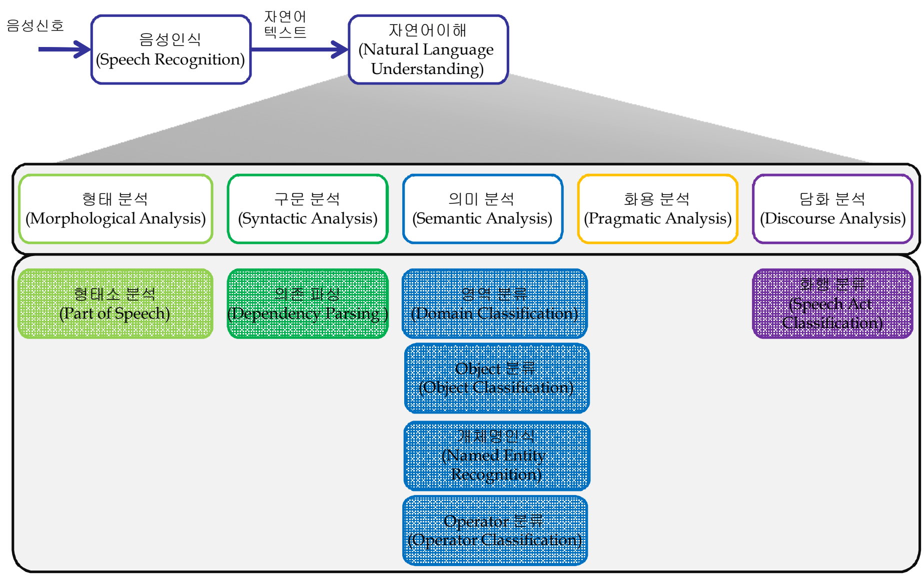
구분	2008년	2009년	2010년	2011년	2012년	2013년	204년	2015년	성장률
국내 시장	3,012	3,485	3,797	4,234	4,736	5,201	5,712	6,274	10.3%

Basic Concepts of SLU

- **Spoken Language Understanding**
  - 음성인식 + Natural Language Understanding(NLU)



Basic Concepts of SLU



Basic Concepts of SLU

- **Domain**
  - 자연어 문장이 포함되는 범주(Category) 또는 주제(Topic)
- **Domain Classification**
  - 자연어 문장을 분석하여 적절한 범주로 분류하는 것

이번 주말 강원도 날씨를 알려줘

영역 분류
(Domain Classification)

날씨(Weather)

Basic Concepts of SLU

- **Named Entity**
  - 인명(Person), 지명(Location), 기관명(Organization) 등과 같은 고유명사
- **Named Entity Recognition (개체명 인식)**
  - 발화 문장에 개체명을 인식하여 해당 태그를 달아주는 것

이번 주말 강원도 날씨는 어떻니?

이번 주말/date 강원도/loc
날씨/NNG 는/JX 어떻/VA 니/EC

Basic Concepts of SLU

- **Object**
  - 사용자가 얻고자 하는 정보
  - 날씨정보(weather_info), 습도정보(humidity_info), 버스번호(bus_number) 등
- **Object Classification**
  - 발화에서 사용자가 원하는 결과를 판단할 수 있는 구간을 인식하여 사용자가 얻고자 하는 정보를 분류

이번 주말 강원도 날씨는 어떻니?

weather_info
이번 주말 강원도 날씨

Basic Concepts of SLU

- **Operator**
  - 사용자가 요구하는 행동
  - 설정(set), 수정(mod), 삭제(del), 찾기(lookup) 등
- **Operator Classification**
  - 발화 속에 포함된 사용자가 요구하는 행동을 분류하는 것

강원도의 이번 주말 날씨를 알려줘

Operator 분류
(Operator Classification)

Lookup

Basic Concepts of SLU

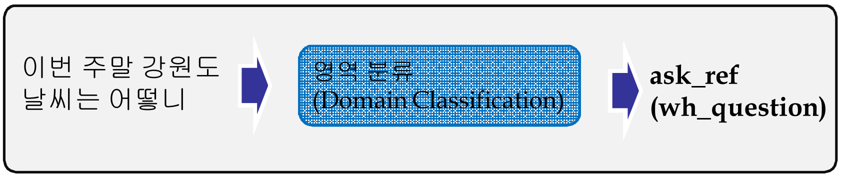
❖ Speech-act

➢ 발화 속에 포함된 대화 목적을 수행하기 위한 화자의 의도된 행위

ask_ref	정보 요구	화자가 청자에게 어떤 변수의 값을 요구(WH-question)
ask_if	정보 요구	화자가 청자에게Yes/No의 답을 요구(Y/N-question)
inform	정보 제공	화자가 청자에게 정보를 제공
response	응답	ask_ref, ask_if에 대한 대답
request	행위 요구	화자가 청자에게 어떠한 행위를 요구
accept	호응	화자가 청자의 발화에 호응
confirm	확인	확인을 요구하는 발화에 대한 응답
reject	거절	대화를 계속 진행할 수 없는 상황

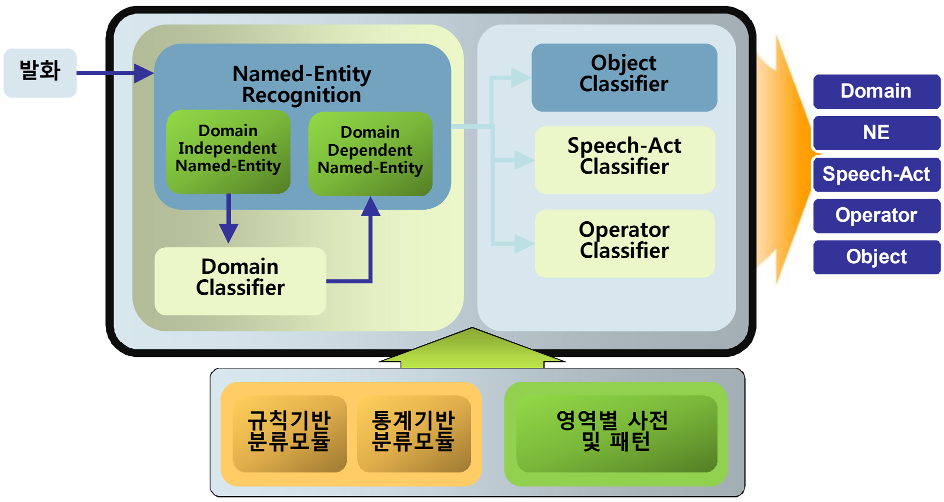
❖ Speech-act Classification

➢ 발화 속에 포함된 사용자가 요구하는 행동을 분류하는 것



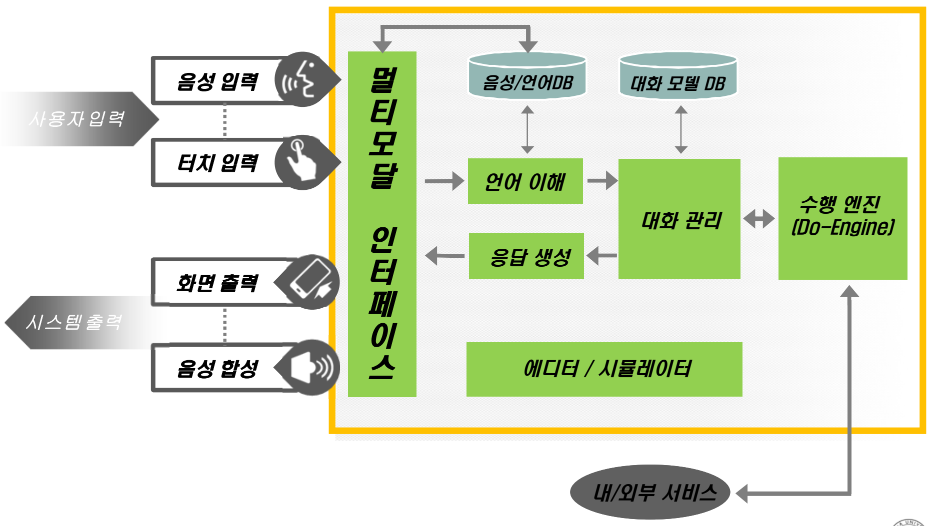
Basic Concepts of SLU

❖ Overview of SLU systems



Basic Concepts of SLU

❖ **Overview of SLU systems**



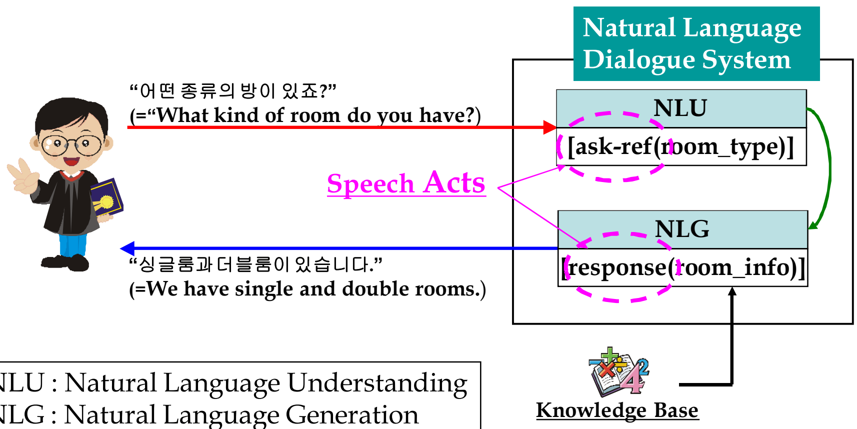
13

Speech-act Classification

❖ **Natural Language Dialogue System**

➢ Speech-act Classification

- 발화에 나타난 화자에 의해 의도된 언어적 행동
- 대화시스템에서 대화를 이해하고 발화를 생성하는데 필수적인 요소



NLU : Natural Language Understanding
NLG : Natural Language Generation

14

Speech-act Classification

- **Two Approaches**
  - **Rule-based** Speech-act Classification
    - 언어적 자질과 화행간의 관계를 정의한 규칙을 사용 (Lee's work, 1996)

      Rule 1. If the utterance have the unknown type pronoun such as *"누구(=who)"*, *"무엇(=what)"*, the utterance is classified as *ASK-REF*
      Rule 2. If the utterance have the unknown type pre-noun such as *"어떤(=What)"*, *"어느(=which)"*, the utterance is classified as *ASK-REF*
      Rule 3. If the utterance have the unknown type adverb such as *"어떻게(=How)"*, *"언제(=when)"*, the utterance is classified as *ASK-REF*
      …
      Rule n. If the utterance start the *"아니오 (=No)"*, the utterance is classified as *CORRECT*

    - 영역 지식에 의존적이기 때문에 확정성과 이식성이 떨어짐
  - **Machine Learning based** Speech-act Classification
    - SVM, CNN etc.
    - 영역내 확장성 및 영역간 이식성이 좋음

Speech-act Classification

- **What are main issues in speech-act classification using machine learning techniques?**
  - **What features** are effective for speech-act classification?
    - Previous Speech-act: Sub-Dialogue Problem
    - Syntactic features?, Lexical Features, Bigram or Unigram?
  - How do we estimate **the weight of features**?
    - Data Sparseness Problem in small size of Categories
    - Utilization of the category distributions of features

Speech-act Classification

- ❖ **What Features?**
  - ➢ **Previous Speech-act**
    - ▪ Sub-Dialogue Problem

발화자	발화	현재 화행	이전발화를 이용한 발화간 자질	담화스택을 이용한 발화간 자질
User	방을 하나 예약하고 싶은데요	Inform	Dialog-start	Dialog-start, NULL
Agent	어떤 방을 원하시죠?	Ask-ref	Inform	Inform, NULL
User	어떤 종류의 방이 있습니까?	Ask-ref	Ask-ref	Ask-ref, NULL
Agent	더블룸과 싱글룸이 있습니다.	Response	Ask-ref	Ask-ref, SS
User	방값이 얼마죠?	Ask-ref	Response	Response,SE
Agent	싱글은 삼만원이고 더블은 사만원 입니다.	Response	Ask-ref	Ask-ref,SS
User	싱글룸으로 해주세요.	Response	Response	Ask-ref,SE

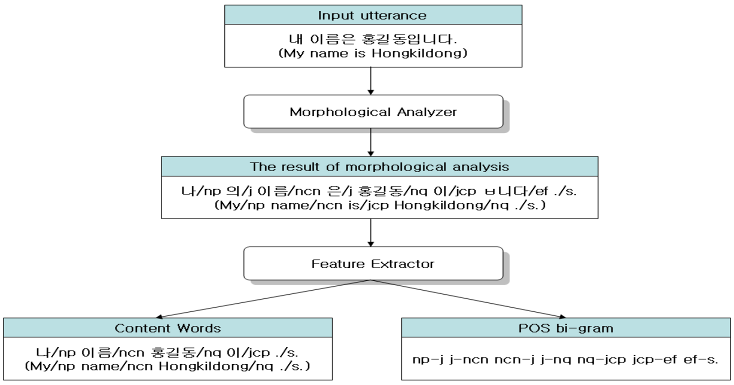
Speech-act Classification

- ❖ **What Features? (Ko et al, 2012)**
  - ➢ **Proposed Solution of Sub-Dialogue Problem: Discourse Stack**
  - ➢ **Adjacency Pairs**: Request Type - Response Type
    - ▪ *Request Type*: ask-ref, ask-if, ask-confirm, offer, suggest, request
    - ▪ *Response Type*: accept, reject, response, acknowledge
    - ▪ *Others*: opening, introducing-oneself, correct, inform, expressive, promise, closing

```
For each utterance
Begin
Reference :
   If (Stack is Empty)
      Use Speech acts and DSI of previous utterance
   Else
      Use Speech acts of discourse stack's top and DSI of previous utterance
Operation :
   If (Utterance is Request Type)
      If (Stack is not Empty)
         Give SS to DSI of current utterance
      Push speech acts of current utterance in discourse stack.
   Else if (Utterance is Response Type)
      Pop speech acts in discourse stack
      If (Stack is not Empty)
         Give SE to DSI of current utterance
End
```

Speech-act Classification

- **What Features? (Ko et al, 2012)**
  - Syntactic features?, Lexical features, Bigram or Unigram?
  - **The Problem of Syntactic Features?**
    - 구문 유형의 구성: [문장유형, 주동사, 시제, 부정여부, 양상, 단서단어]
    - **Low accuracy of Syntactic Analysis**



19

Speech-act Classification

- **Experimental Data**
  - **Korean Dialogue Corpus**
    - 한국어 화행 분석을 위한 많은 논문들에서 사용
    - 호텔, 비행, 관광 예약 내용
    - **528대화, 10,285 발화 (19.48 대화당 발화)**

Tag	Values
SP	Customer
KS	미국 조지아대 어학연수에 참가 신청을 한 학생인데요.
EN	I'm a student and registered for a language course at University of Georgia in U.S.
SA	Introducing-oneself
DS	[2]

SP : 화자, **KS** : 한국어, **EN** : 영어, **SA** : Speech Acts, **DS** : Discourse Structure

20

Speech-act Classification

❖ Experimental Data

➢ Speech-act Distributions

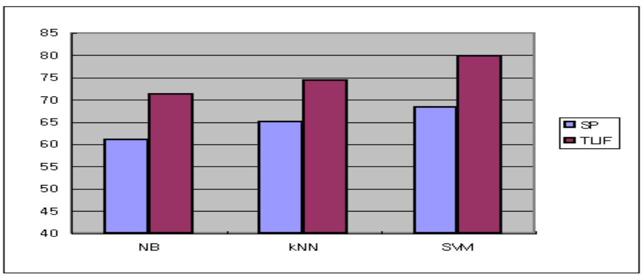
Speech act type	Ratio (%)	Speech act type	Ratio (%)
Accept	2.49	Introducing-oneself	6.75
Acknowledge	5.75	Offer	0.4
Ask-confirm	3.16	Opening	6.58
Ask-if	5.36	Promise	2.42
Ask-ref	13.39	Reject	1.07
Closing	3.39	Request	4.96
Correct	0.03	Response	24.73
Expressive	5.64	Suggest	1.98
Inform	11.9	Total	100

Speech-act Classification

❖ Experimental Results

- ➢ 구문 유형 (**SP**)
- ➢ 두 단계 발화 내 자질 (**TLIF**)

	NB	kNN	SVM
SP	61.05	65.1	68.33
TLIF	71.22	74.38	79.95

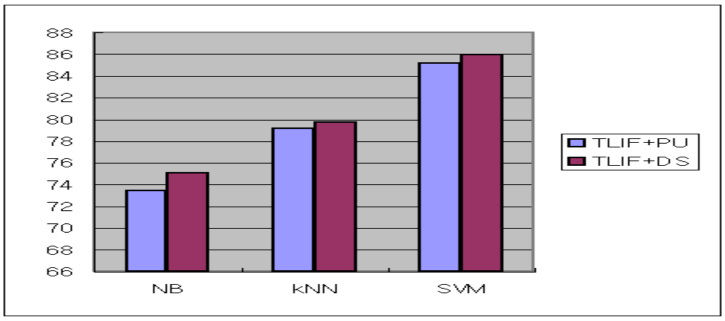


Speech-act Classification

❖ Experimental Results

- Evaluation of Discourse Stack (담화스택)
  - TLIF + PU (이전발화의 화행)
  - TLIF + DS (담화스택을 이용한 이전발화 화행)

	NB	KNN	SVM
TLIF+PU	73.48	79.18	85.18
TLIF+DS	75.1	79.75	85.95



23

Speech-act Classification

❖ The Weight of Features (Ko et al, 2012)

- The problems of previous feature weighting scheme
  - **Data Sparseness Problem** in small size of Categories
- The new feature weighting scheme
  - **Step 1**: Construction of two-level Speech-act Hierarchy
  - **Step 2**: Shrinkage-based estimation of feature probabilities
- **Two-level Speech-act Hierarchy**

	Parent	Child
Root	**Type1**: Utterances of request type	Ask-if
		Ask-ref
		Ask-confirm
		Offer
		Suggest
		Request
	Type2: Utterances of response type	Accept
		Response
		Reject
		Acknowledge
	Type3: Utterances with a speaker emotion	Expressive
		Promise
		Closing
	Type4: Utterances of usually life	Opening
		Introducing-oneself
		Correct
		Inform

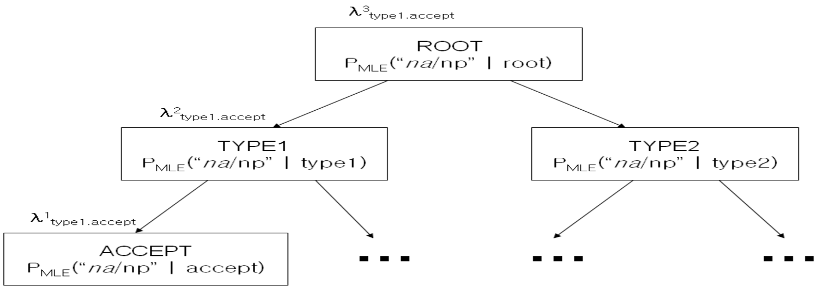
24

Speech-act Classification

❖ The Weight of Features

➢ Shrinkage-based Estimation

$$P(f_t | s_j ; \breve{\theta}_j) = \lambda_j^1 \hat{\theta}_{jt}^1 + \lambda_j^2 \hat{\theta}_{jt}^2 + \ldots + \lambda_j^k \hat{\theta}_{jt}^k$$



$P_{SHIRINKAGE}(\text{"}na\text{/np"} \mid \text{accept}) = \lambda^1_{type1.accept} P_{MLE}(\text{"}na\text{/np"} \mid \text{accept}) + \lambda^2_{type1.accept} P_{MLE}(\text{"}na\text{/np"} \mid \text{type1}) + \lambda^3_{type1.accept} P_{MLE}(\text{"}na\text{/np"} \mid \text{root})$

Speech-act Classification

❖ The Weight of Features

➢ Parameter Estimation on EM

Initialize:

Set the λ_j's to some initial values, say $\lambda_j^i = \frac{1}{k}$

Iterate:

1. Calculate the degree to which each estimate predicts the features f_t in the held-out feature set, H_j, from speech acts s_j :

$$\beta_j^i = \sum_{w_t \in H_j} P(\hat{\theta}_j^i \text{ was used to generate } f_t) = \sum_{w_t \in H_j} \frac{\lambda_j^i \hat{\theta}_{jt}^i}{\sum_m \lambda_j^m \hat{\theta}_{jt}^m} \quad (5)$$

2. Compensate the degree for loss that is caused by large variation of each degree :

$$\beta_i^j = \beta_i^j + \frac{\sum_m \beta_j^m}{m} \quad (6)$$

3. Derive new weights by normalizing the β's :

$$\lambda_j^i = \frac{\beta_j^i}{\sum_m \beta_j^m} \quad (7)$$

Terminate: Upon convergence of the likelihood function

Speech-act Classification

The Weight of Features

- **Result of Parameter Estimation on EM**

# training documents	Speech Acts			Mixture Weights		
	Root	Parent	Child	Root	Parent	Child
250	Root	Type1	Ask-ref	0.289	0.32	0.39
			Suggest	0.257	0.275	0.467
		Type2	Expressive	0.263	0.335	0.4
		Type3	Reject	0.259	0.269	0.47
		Type4	Inform	0.297	0.336	0.366
8349	Root	Type1	Ask-ref	0.282	0.295	0.422
			Suggest	0.217	0.22	0.562
		Type2	Expressive	0.229	0.279	0.49
		Type3	Reject	0.212	0.215	0.571
		Type4	Inform	0.26	0.332	0.406

- **Feature weighting scheme for SVM**

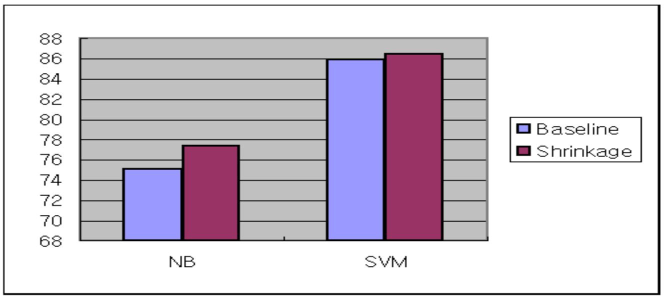
$$w_{ik} = 1.0 + P(f_{ik} \mid s_j; \breve{\theta}_j)$$

Speech-act Classification

Experimental Results

- **Shrinkage-based feature weighting scheme**

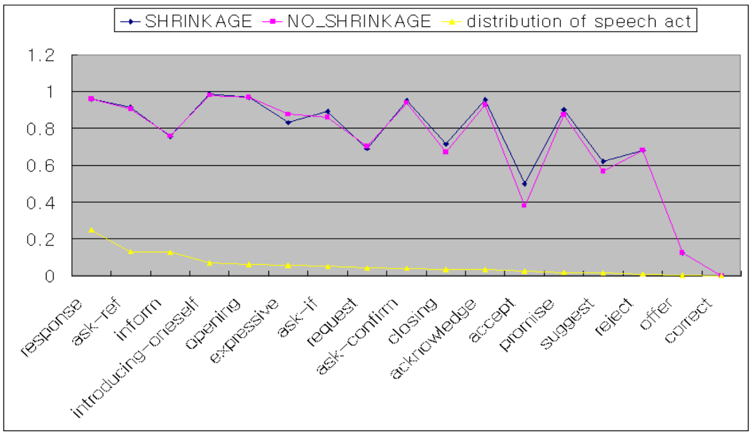
	NB	SVM
Baseline	75.1	85.95
Shrinkage	77.4	86.5



Speech-act Classification

❖ **Experimental Results**

➢ **Shrinkage-based feature weighting scheme**

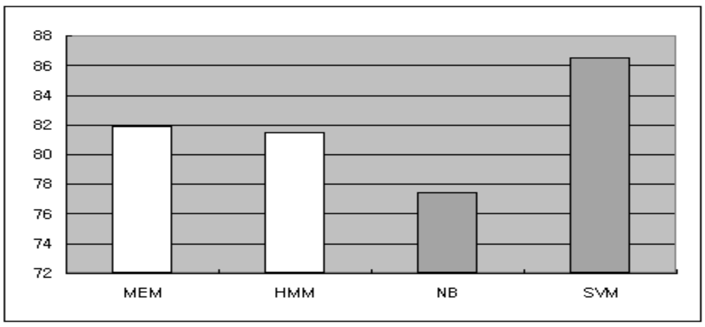


Speech-act Classification

❖ **Experimental Results**

➢ **Comparing Other Methods**

MEM	HMM	NB	SVM
81.9	81.5	77.4	86.5



Speech-act Classification

- **The Weight of Features (Ko, 2012 and 2015)**
  - **Category Distributions of Features**
  - Binary feature in Speech-act classification
    - Simpler but more effective than other schemes such *tf*, *idf*, and *tf.idf*
    - Why? An utterance is much shorter than a document
  - **Two weighting schemes** for the classification cases with the small number of features
    - 1) Apply to the **entropy** concept to estimate the feature importance with all category distributions of each feature
    - 2) The **ratio of positive and negative category distributions**

Speech-act Classification

- **Category Distributions of Features**
  - **Estimation of feature probabilistic distribution for categories**
    - ELE (Expected Likelihood Estimator)

$$P(f_i|c_j) = \frac{N(f_i, c_j) + 0.5}{\sum_{t=1}^{|V|} N(f_t, c_j) + 0.5 \times |V|},$$

$$P(f_i|\bar{c}_j) = \frac{N(f_i, \bar{c}_j) + 0.5}{\sum_{t=1}^{|V|} N(f_t, \bar{c}_j) + 0.5 \times |V|},$$

  - **Entropy value of Category Probabilities (ECP)**

$$\mathrm{ECP}(f_i) = \frac{\mathrm{MaxEntropy}}{-\sum_{j=1}^{|C|} P(c_j|f_i)\log P(c_j|f_i)},$$

$$\mathrm{MaxEntropy} = -\sum_{j=1}^{|C|} P_{\mathrm{uniform}}(c_j)\log P_{\mathrm{uniform}}(c_j)$$
$$= -\sum_{j=1}^{|C|} \frac{1}{|C|}\log\frac{1}{|C|} = \log|C|,$$

Speech-act Classification

- **Category Distributions of Features**
  - **Log-Odds Ratio (LOR) of positive and negative categories**

$$\mathrm{LOR}(f_i, c_j) = \log\left(\frac{P(f_i|c_j)}{P(f_i|\bar{c}_j)} + \alpha\right),$$

  - For test utterances

$$\mathrm{TW}(f_i) = \max_{c_j} \mathrm{LOR}(f_i, c_j),$$

- **Traditional Feature Weighting Scheme**

$$\mathrm{binary}_{ij} = \begin{cases} 1 & \text{if } tf_{ij} \geq 1 \\ 0 & \text{if } tf_{ij} = 0 \end{cases},$$

$$tf_{ij} = \begin{cases} tf_{ij} & \text{if } tf_{ij} \geq 1 \\ 0 & \text{if } tf_{ij} = 0 \end{cases},$$

$$idf_{ij} = \begin{cases} \log\left(\frac{N}{df_{ij}}\right) & \text{if } tf_{ij} \geq 1 \\ 0 & \text{if } tf_{ij} = 0 \end{cases},$$

$$tf.idf_{ij} = \begin{cases} tf_{ij} \cdot \log\left(\frac{N}{df_{ij}}\right) & \text{if } tf_{ij} \geq 1 \\ 0 & \text{if } tf_{ij} = 0 \end{cases},$$

33

Speech-act Classification

- **Experimental Results**
  - **Two Data Sets**
    - **1) RES:** Hotel, Travel, Airplane Reservation
    - **2) SM:** Schedule Management **(New)**
      - ✓ **SM-11:** 954 dialogues (22.3 utterances per a dialogue and 11 speech-acts, 23,310 utterances-17,054 for training and 4,256 for test data)
      - ✓ **SM-8:** 8 categories after removing 3 rare categories
  - **Performance Comparison of Conventional Schemes**

		SVM		k-NN	
		Micro-avg F1	Macro-avg F1	Micro-avg F1	Macro-avg F1
RES	*Binary*	***85.72***	***75.84***	***79.82***	***71.41***
	tf	85.3	75.6	78.01	69.15
	idf	85.41	75.77	77.34	69.46
	tf.idf	84.63	74.88	77.13	70.59
SM-8	*Binary*	***94.31***	***88.94***	***90.8***	***85.17***
	tf	94.17	88.48	90.52	84.77
	idf	94.09	88.34	89.42	84.46
	tf.idf	94.09	88.54	89.42	84.33
SM-11	*Binary*	***94.1***	***73.27***	***90.34***	***61.42***
	tf	94.00	73.00	90.01	60.9
	idf	94.08	73.24	89.38	61.29
	tf.idf	93.96	73.16	89.26	61.16

34

Speech-act Classification

❖ **Experimental Results**

➢ **Performance Comparison of the Proposed Schemes (*ECP* and *LOR*)**

		SVM		k-NN	
		Micro-avg F1	Macro-avg F1	Micro-avg F1	Macro-avg F1
RES	*Binary*	85.72	75.84	79.82	71.41
	ECP	***86.39***	***78.06***	80.18	71.65
	LOR	85.15	76.71	***82.3***	***74.49***
SM-8	*Binary*	94.31	88.94	90.8	85.17
	ECP	***94.78***	***89.43***	91.63	86.14
	LOR	93.18	88.31	***91.93***	***87.12***
SM-11	*Binary*	94.1	73.27	90.34	61.42
	ECP	***94.67***	***73.79***	91.33	71.16
	LOR	93.04	72.78	***91.7***	***71.99***

➢ **Performance Comparison of the Proposed model and other Models**

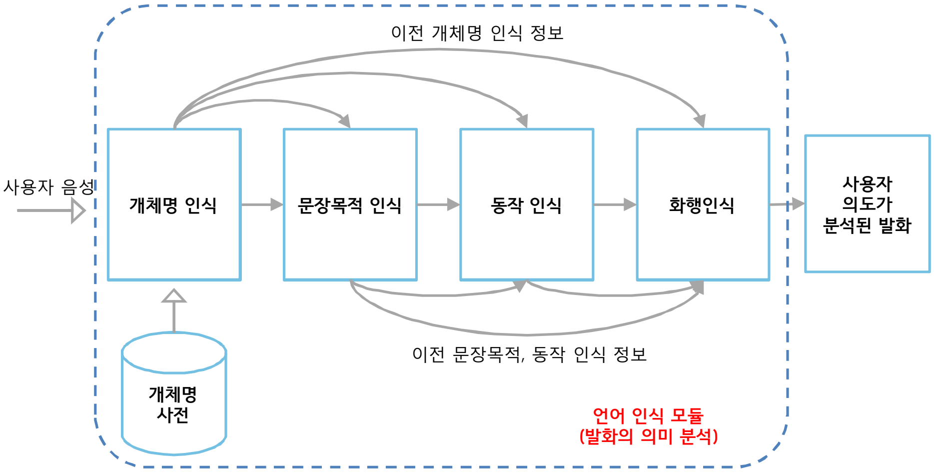
Classification model	Classifier	Micro-avg F_1
Choi's model	MEM	83.57
Lee's model	HMM with decision tree	81.5
Kang's model	SVM	84.52
Proposed model	SVM	86.39

35

Simultaneous Recognition Model

❖ **Traditional SLU Model**

➢ Pipeline Process



36

Simultaneous Recognition Model

- **Problems of Traditional SLU Model**
  - Error Propagation
  - Duplicated Features, Not share useful features
  - Required 4 different classifiers
- **Solution of these Problems**
  - **Simultaneous Recognition model (Ko et al, 2015)**
- **Example of SLU Recognition**

사용자 발화 : 지금 뉴욕은 얼마나 더워?

개체명 : 지금 / Time , 뉴욕 / Location
문장목적 : 기온정보 / Temperature_Info
동작 : 조회 / Lookup
화행 : 요구 / Request

37

Simultaneous Recognition Model

- **Pseudo Tag Addition Method for Sequential Analysis**
  - Different Recognition Units
    - POS Unit for NE
    - Sentence Unit for Object, Operator and Speech-act

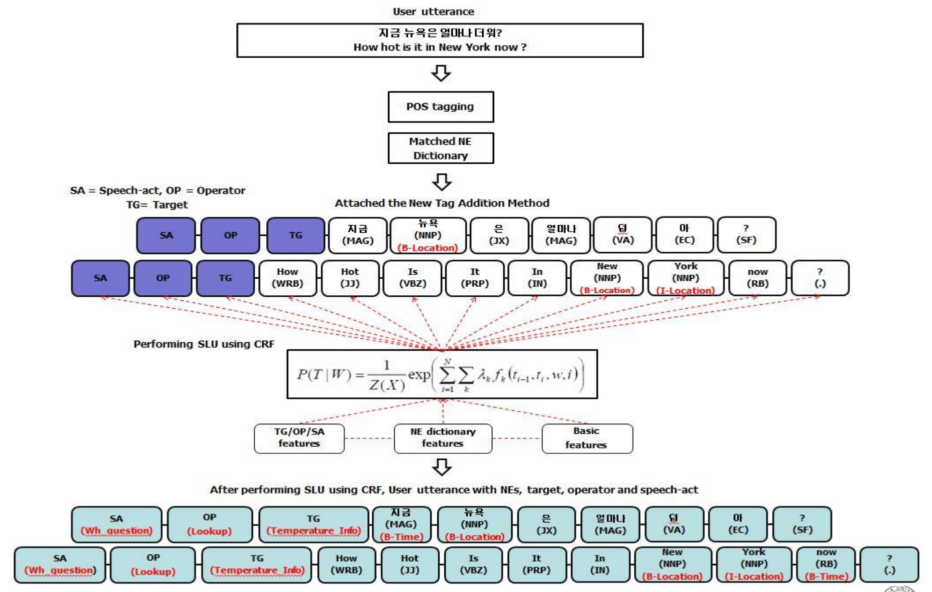
가상 태그 부여 방법

사용자 발화 : 지금 뉴욕은 얼마나 더워?

가상 태그 부여 방법을 적용한 사용자 발화:
[SA] [OP] [OB] 지금 뉴욕은 얼마나 더워?

38

Simultaneous Recognition Model



39

Simultaneous Recognition Model

❖ **Example of Simultaneous Recognition Result for SLU**

사용자 발화

지금 뉴욕은 얼마나 더워?

⇩

형태소 분석

지금/MAG 뉴욕/NNP+은/JX 얼마나/MAG 덥/VA+아/EC+?/SF

⇩

개체명 사전 매치 후 BIO 태그 부착

지금/MAG [뉴욕/NNP/B_Location]+은/JX 얼마나/MAG 덥/VA+아/EC ?/SF

⇩

가상 태그 부여 방법 적용

[SA] [OP] [OB] 지금/MAG 뉴욕/NNP/B_Location+은/JX 얼마나/MAG 덥/VA+아/EC ?/SF

⇩

학습된 CRF 모델을 토대로 개체명, 문장목적, 동작, 화행 분류

[SA/Request] [OP/Lookup] [OB/Temperture_Info] [지금/MAG/B_TIme] [뉴욕/NNP/B_Location]+은/JX 얼마나/MAG/ 덥/VA/+아/EC ?/SF

40

Simultaneous Recognition Model

❖ Three Feature Sets for Simultaneous Recognition Model

- ➢ Example of Utterance
  - "지금 뉴욕은 얼마나 더워" - 개체명 사전 매치 : 뉴욕
  - 지금/MAG 뉴욕/NNP/**B_Location**+은/JX 얼마나/MAG 덥/VA+아/EC+?/SF

➢ Basic Feature

형태소 어휘/태그 자질	현재 위치의 형태소 어휘/태그 정보 (Ex. 현재 위치의 형태소가 "뉴욕" 일 경우 : 뉴욕/NNG)
형태소 어휘 시퀀스 자질	현재 형태소 어휘 위치를 기준으로 최대 윈도우 사이즈 2를 가진 형태소 어휘 정보 - (-1,0) (0,1) (-2,-1,0) (0,1,2) (-1,0,1) (-2,-1,0,1) (-1,0,1,2) (-2,-1,0,1,2) 단, 현재 형태소 어휘 위치는 숫자로 표현한다. (Ex. 현재 형태소 형태소가 "뉴욕/NNP" 일 경우 : (지금/-1+뉴욕/0) (뉴욕/0+얼마나/1) (null/-2+지금/-1,뉴욕/0) ...

형태소 태그 시퀀스 자질	현재 형태소 태그 위치를 기준으로 최대 윈도우 사이즈 2를 가진 형태소 어휘 정보 - (-1,0) (0,1) (-2,-1,0) (0,1,2) (-1,0,1) (-2,-1,0,1) (-1,0,1,2) (-2,-1,0,1,2) 단, 현재 형태소 태그 위치는 숫자로 표현한다. (Ex. 현재 형태소 형태소가 "뉴욕/NNP" 일 경우 : (MAG/-1+NNP/0) (NNP/0+JX/1) (null/-2+MAG/-1,NNP/0) ...
어절 내 자질	현재 형태소의 어절 내 위치 (Ex. 현재 형태소가 "뉴욕/NNP"일 경우 : 어절은 뉴욕/NNP+은/JX 어절의 처음 위치이므로 Start이며, Start, Continue, End로 구성 형태소 태그 / 어절 길이 현재 형태소 태그와 형태소를 포함하는 어절의 길이

Simultaneous Recognition Model

❖ Three Feature Sets for Simultaneous Recognition Model

- ➢ Example of Utterance
  - "지금 뉴욕은 얼마나 더워" - 개체명 사전 매치 : 뉴욕
  - 지금/MAG 뉴욕/NNP/**B_Location**+은/JX 얼마나/MAG 덥/VA+아/EC+?/SF

➢ NE Feature

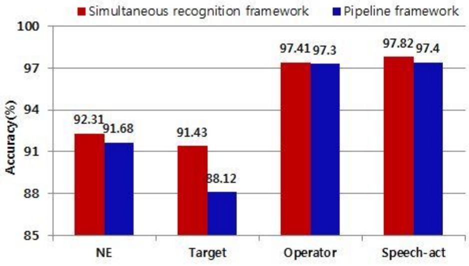
개체명 사전 자질	현재 위치의 형태소가 개체명 사전으로부터 매치되는 경우, 형태소 어휘/태그/개체명 정보를 자질로 사용 (Ex. 뉴욕/NNG/B_Location)
이전 개체명 자질	현재 위치의 형태소 기준으로 이전 위치의 개체명 정보 (Ex. 현재 형태소가 "얼마나/MAG"일 경우, 이전 형태소의 개체명 "B_Location"을 자질로 사용
개체명 출현 자질	현재 발화에 출현한 모든 개체명 종류 (Ex. 예제 발화에서는 "SN_B_Location") 이 자질은 특정 개체명들은 특정 도메인에서만 나오는 특징으로 인해 사용

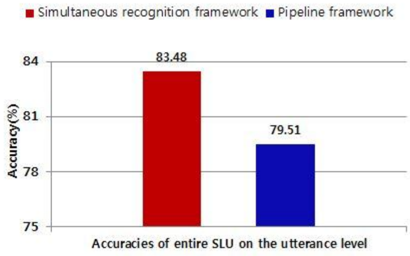
Sentence Feature

동사 자질	현재 발화의 동사 자질 (Ex. 예제 발화에서는 "덥/VA+아/EC")
형태소 어휘 Unigram 자질	현재 발화의 형태소 어휘 Unigram 자질 (Ex. 지금, 뉴욕, 은, 얼마나, 덥, 아, ?)
형태소 어휘/태그 Bigram 자질	현재 발화의 형태소 어휘/태그 Bigram 자질 (Ex. 형태소 어휘 : 지금_뉴욕, 뉴욕_은, 은_얼마나 ...) (Ex. 형태소 태그 : MAG_NNP, NNP_JX, JX_MAG)
형태소 어휘/태그 Trigram 자질	현재 발화의 형태소 어휘/태그 Trigram 자질 (Ex. 형태소 어휘 : 지금_뉴욕_은, 뉴욕_은_얼마나 ...) (Ex. 형태소 태그 : MAG_NNP_JX, NNP_JX_MAG.)
육하원칙 자질	현재 발화의 육하원칙 [왜, 어떻게, 무엇, ...] 자질 이진 자질로써 존재하면 True, 존재하지 않으면 False (Ex. 예제 발화에서는 "얼마나"가 존재하므로 "True"

Simultaneous Recognition Model

- **Experimental Settings**
  - **MADS (Multi-Applications Dialogues for Smart phones)** data set
    - 6 domains: weather, clock, alarm, schedule, exchange and traffic
    - 1,925 utterances, 8 NE, 28 objects, 5 operators and 6 speech-act tags
  - Five-fold cross validation and CRF (Mallet toolkit)
  - Accuracy in utterance level
  - Paired *t*-test and Wilcoxon signed rank test
- **Performance Comparisons**





43

Simultaneous Recognition Model

- **Processing Time and Significant Test**

	Test Data
Proposed Framework	15 sec.
Pipeline Framework	19 sec.

	Proposed *vs.* pipeline
Paired *t*-test	0.00001
Wilcox signed rank test	0.021

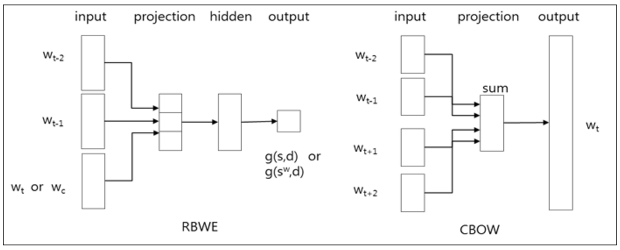
- **Comparison of Other Method (Jeong and Lee, 2008)**

	NE+Speech-act	All (four components)
Proposed Framework	90.61	83.48
Triangular-chain CRF	87.07	16.4

44

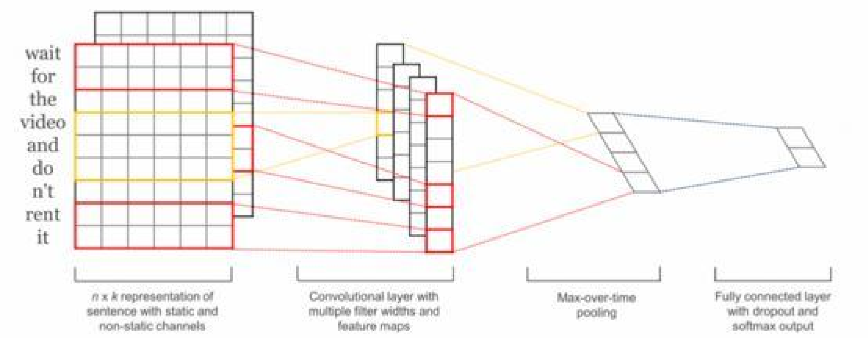
Deep Learning for Speech-act Classification

- **Word Embedding (Ko et al., under review)**
  - Train with 31 billion lexical from several corpora in Korean
  - Vocabulary Size: 100,000
  - Dimension: 64
  - Two Approaches
    - **Ranking-Based Word Embedding (RBWE)**
    - **Continuous Bag-of-Words (CBOW)**



Deep Learning for Speech-act Classification

- **Convolution Neural Network (CNN)**



Deep Learning for Speech-act Classification

❖ **Distribution of the Training Data and Three Types of Test Data**

		RES	SM-8
Train		8,349	17,054
Test	ALL	1,932	4,254
	No-OOV	1,494	4,036
		77%	95%
	In-OOV	438	218
		23%	5%

Deep Learning for Speech-act Classification

❖ **Experimental Results**

		SVM	*DBN*		*RBWE*		*CBOW*	
		Micro	*Micro*	*ER*	*Micro*	*ER*	*Micro*	*ER*
RES	*ALL*	84.9	86.2	8.6	**87.0**	**13.9**	86.7	11.9
	No-OOV	85.3	86.9	10.9	**87.4**	**14.3**	87.3	13.6
	In-OOV	83.8	83.8	0.0	**87.4**	**22.2**	87.0	19.8
SM	*ALL*	93.1	92.1	-14.5	**94.8**	**24.6**	94.4	18.8
	No-OOV	93.7	92.4	-20.6	**95.3**	**25.4**	94.8	17.5
	In-OOV	81.2	87.6	34.0	**92.2**	**58.5**	86.2	26.6

SVM	*DBN*		*RBWE*		*CBOW*	
Macro	*Macro*	*ER*	*Macro*	*ER*	*Macro*	*ER*
75.7	78.0	9.5	**85.4**	**39.9**	85.0	38.3
77.3	79.8	11.0	**87.0**	**42.7**	85.6	36.6
70.0	65.7	-14.3	**86.0**	**53.3**	85.6	52.0
85.0	81.5	-23.3	**92.6**	**50.7**	91.5	43.3
85.7	82.3	-23.8	**93.1**	**51.7**	92.3	46.2
74.4	76.7	9.0	**87.6**	**51.6**	80.7	24.6

Deep Learning for Speech-act Classification

- **Experimental Results**

		RES	*SM*
SVM	*unigram*	84.9	93.1
	unigram + bigram	85.3 (+0.5)	93.1 (+0.0)
RBWE	*unigram*	**87.0 (+2.4)**	**94.8 (+1.9)**
CBOW	*unigram*	86.7 (+2.1)	94.4 (+1.4)

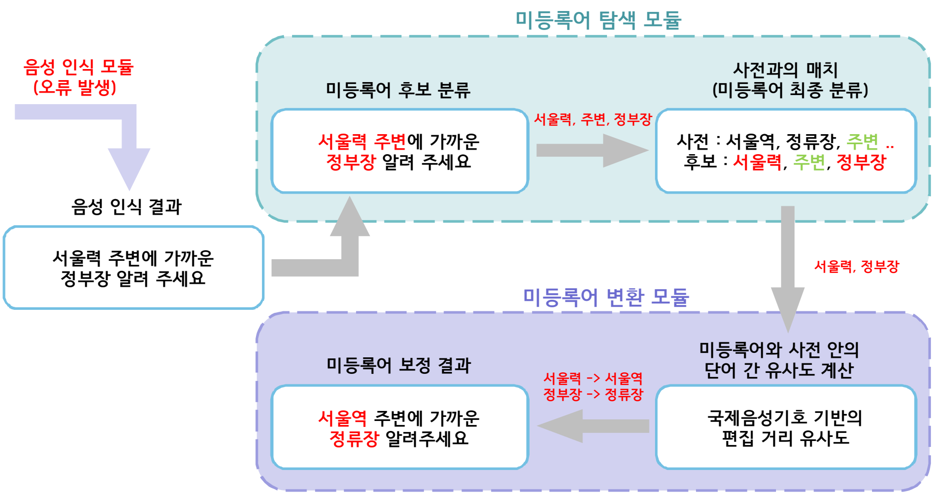
OOV Handling Module

- **How to handle OOVs from Speech Recognition (논문 작성 중)**
  - OOV detection module
  - OOV correction module

- **Necessity of OOV Handing for SLU**
  - Output of speech recognition is the input of SLU
  - Very important for obtaining the accurate results of SLU

Simultaneous Recognition Model

❖ **Example of OOV Handling Process**



Simultaneous Recognition Model

❖ **Performance of OOV Candidates Detection**

미등록어 후보 분류 성능 (CRF) (%)				
교통	날씨	알람	환율	시계
90.24	93.21	96.82	93.58	96.67

❖ **Error Reduction Rate of OOV**

전체 미등록어 개선율 (%)				
교통	날씨	알람	환율	시계
58.07	34.04	34.40	44.61	52.57

Simultaneous Recognition Model

❖ **Reconstruction Rate of Speech Recognition Results to Original Sentences**

음성 인식 오류가 없는 원본 문장과의 일치율 (%)					
	교통	날씨	알람	환율	시계
음성 인식 결과	67.6	71.2	72.93	57.76	66.5
미등록어 처리 모듈을 수행하여 만든 새로운 음성 인식 결과	85	78.27	79.19	72.02	81.4

References

- **Youngjoong Ko** “A New Term Weighting Scheme for Text Classification Using the Odds of Positive and Negative Class Probabilities.” *Journal of the Association for Information Science and Technology, Wiley-Blackwell, Vol. 66, No. 12, pp. 2553-2565*, December 2015.
- Changsu Lee, **Youngjoong Ko** and Jungyun Seo. ”A Simultaneous Recognition Framework for the Spoken Language Understanding Module of Intelligent Personal Assistant Software on Smart Phones.” *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015)*, pp. 818-822, in Beijing, China, July 2015.
- **Youngjoong Ko**. “New Feature Weighting Approaches for Speech-act Classification.” *Pattern Recognition Letters, Elsevier Science, Vol. 51, pp. 107-111*, January 2015.
- Sangwoo Kang, **Youngjoong Ko** and Jungyun Seo. ”Hierarchical Speech-act Classification for Discourse Analysis.” *Pattern Recognition Letters, Elsevier Science, Vol. 34, No. 10, pp. 1119-1124*, July 2013.
- **Youngjoong Ko.** ”A Study of Term Weighting Schemes Using Class Information for Text Classification.” *Proceedings of the 35th Annual International ACM SIGIR Conference (SIGIR 2012), pp. 1029-1030*, in Portland Oregon, USA, August 2012.
- Kyungsun Kim, **Youngjoong Ko** and Jungyun Seo. ”An Effective Application of Contextual Information using Adjacency Pairs and a Discourse Stack for Speech-Act Classification.” *International Journal of Innovative Computing Information and Control (IJICIC), Vol. 8, No. 11, pp. 7713-7728*, November 2012.
- Kyoungman Bae and **Youngjoong Ko**. “Speech-act Classification Using Convolution Neural Network and Word Embedding.” *Pattern Recognition Letters,* Under Review.

Thank you for your attention!

http://web.donga.ac.kr/yjko/

고 영 중