Ch4.
Intent Determination and Spoken Utterance Classification

지식표현기법 | SLU | 2017. 10. 10.

2016010662 윤주성
2017010615 하종수

Index

발표는 다음과 같은 순서로 진행됩니다

Preview

SUC는 SLU의 special form, Chapter 4에서는 SUC techniques에 대해서 리뷰

본 챕터의 목적

- ✓ SUC techniques에 대해서 전체적으로 리뷰함
- ✓ Intent determination system에서 성공적이었던 applications에 대해서 설명함

Spoken Utterance Classification(SUC) 이란?

- ✓ SUC는 SLU의 Special form 임
- ✓ Call routing system (Call type classification) 에서 주로 많이 사용됨
- ✓ Pre-defined call-type이나 type of intent로 분류하는 task임

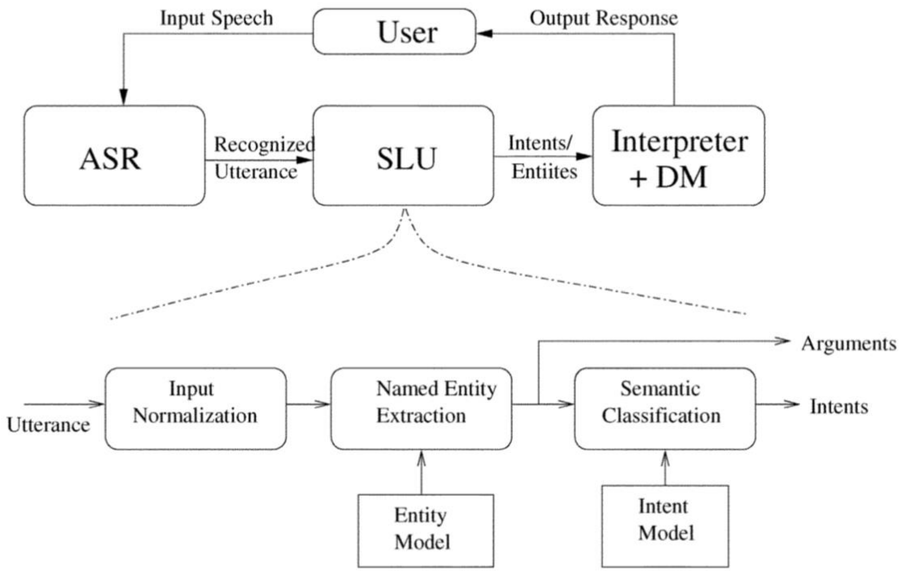


Fig. 1. AT&T SLU run-time system. (Gupta et al., 2006)

4.1 Background

Keyword detection based approach

AT&T's Voice Recognition Call Processing (VRCP) system (Wilpon et al., 1990)

고객 서비스 상담원

- ✓ Call center에서 customer service representatives (CSRs) 를 위해 사용 됨
- ✓ 큰 call center의 경우 CSRs의 수가 천 명 이상 됨에 따라 재정적 부담 발생
- ✓ AT&T VRCP system은 유저에게 3가지 키워드만 말하게 함 (please say collect, calling card, or third party)
- ✓ SLU에선 primitive approach였지만 시스템은 매우 성공함 (1년에 10억 건 처리함)
- ✓ 이러한 종류의 machine-initiative 시스템을 Interactive Voice Response system(IVR)이라 함

Interactive Voice Response system (IVR)

- ✓ 모든 Interaction은 machine에 의해 제어됨
- ✓ 유저에게 특정 질문을 던짐
- ✓ 유저가 machine내에서 미리 정의된 키워드 중 하나를 말하기를 기대함
  ex) mail delivery system, pizza delivery system
- ✓ VoiceXML 형태로 많이 구현됨

```
<vxml version="2.0">
<form id= "get_pizza_topping">
<field name= "pizza_topping">
<prompt> What kind of topping would you like to have? </prompt>
<grammar> olive | pepperoni | mushroom </grammar>
<noinput> Please say the topping you wish to have. </noinput>
<nomatch> I didn't understand that. </nomatch>
<filled> Thank you, I added <value expr= "pizza_topping" />
<submit next="next_document.vxml" /> </filled>
</field>
</form>
</vxml>
```

Figure 4.1 A VXML example for getting pizza topping information (from the VXML lecture by François Mairesse)

4.1 Background

Keyword detection based approach

앞으로 계속 언급됨

AT&T How May I Help You?(HMIHY) (Gorin et al., 1997)

- ✓ Call routing system
- ✓ Open-ended prompt가 "How May I Help You"
- ✓ AT&T's call center에 배포 된 후 전국적으로 크게 성공함 (Gorin et al., 2002)
- ✓ 6+1가지 call type으로 분류 (Billing, Rates and Plans, …, Other)
- ✓ Not specific("I have a question")한 경우나 Not Informative("Good afternoon")한 경우 reprompt 수행

✓ 용어 정리,
Call type == predefined categories

HMIHY: How may I help you?
User: Hi, I have a question about my bill *(Billing)*
HMIHY: OK, what is your question?
User: May I talk to a human please? *(CSR)* Utterance level intents for better understanding
HMIHY: In order to route your call to the most appropriate department can you tell me the specific reason you are calling about?
User: There is an international call I could not recognize *(Unrecognized_Number)*
HMIHY: OK, I am forwarding you to the human agent. Please stay on the line.

Figure 4.2 A conceptual example dialogue between the user and the AT&T HMIHY system

4.1 Background

call classification & semantic frame filling의 관계

Note

Call classification & Semantic frame filling 의 차이

- ✓ Semantic frame filling과 달리 Call classification 에서는 유저가 주는 argument 를 크게 신경 안 씀
- ✓ 왜냐하면 Call classification 은 단지 적합한 콜센터부서로 routing 하는 것이 목표
- ✓ Argument는 단지 더 옳은 분류를 하기 위한 용도로 사용
  ex) 특정 약의 이름(argument)은 prescription renewal or refill department 부서로 연결하는데 도움을 줌

둘은 완전히 '다른' Task지만 template filling 할 땐 complimentary함

- ✓ 각각의 call-type을 template으로도 볼 수 있음
  1. classification: 먼저 유저의 발화를 goal에 맞게 분류
     ex) *flight reservation* or *ground transportation* in ATIS
  2. frame filling: 분류한 goal에 맞는 template을 채움
- ✓ Finer grain categorization에서는 template filling task과 semantic representation이 비슷함
  ex) AT&T VoiceTone system에서 predicate/argument-type sematic representation 사용

```
<frame name="ShowFlight" type="Void">
    <slot name="topic" type="Topic">
    <slot name="flight" type="Flight">
</frame>
<frame name="GroundTrans" type="Void">
    <slot name="city" type="City">
    <slot name="type" type="TransType">
</frame>
<frame name="Flight" type="Flight">
    <slot name="DCity" type="City">
    <slot name="ACity" type="City">
    <slot name="DDate" type="Date">
</frame>
```

Figure 3.2

둘 다 payments department로 Routing 되는 건 같지만 전보다 표현력이 좋아짐

I am calling to check whether you received my payment. → *Verify(Payment)*

I am calling to make a payment. → *Make(Payment)*

Goals In ATIS:
flight reservation, *ground transportation*, *airplane specifications*, etc

4.1 Background

Dialogue act tagging

Dialogue act tagging in Chapter 9

- ✓ Dialogue acts에선 discourse analysis & understanding를 위해 users' utterance 를 predefined classes로 분류하고자 함
- ✓ 예를 들면, 아래와 같이 intent에 따라 다양하게 쓰일 수 있음
  - Question -> rhetorical question or yes/no question
  - Statement -> command or suggestion
- ✓ 이런 관점에서 dialogue act tagging 라는 task가 call-type classification 또는 intent determination system 과 human/human conversation understanding system 사이의 연결고리라고 할 수 있음

4.2 Task Description

SUC? 발화로부터 시멘틱 클래스를 찾는 Task

Spoken Utterance Classification (SUC)

SLU의 special form이며, 발화로 부터 시멘틱 클래스를 알아내는 분류 문제

SUC task aims at classifying a given speech utterance $\boldsymbol{X_r}$ into one of $\boldsymbol{M}$ semantic classes

$$\hat{C}_r = \arg\max_{C_r} P(C_r|X_r). \tag{4.1}$$

$$\hat{C}_r \in \mathcal{C} = \{C_1, \ldots, C_M\} \text{ (where } r \text{ is the utterance index)}$$

- ✓ Utterance variations 에 대처할 수 있어야함
  ex) "I want to fly from Boston to New York next week" == "I am looking to fly from JFK to Boston in the coming week"
- ✓ 표현의 자유도가 높아도 utterance는 특정 정보와 바인딩 되는 clear structure를 가짐
  ex) Flight class: "Show all flights" == "Give me flights"
  Fare class: "Show me fares"
- ✓ 그러므로, Class C와 word sequence W 의 관계를 잡아내는 feature function $f_i(C, W)$ 을 선택해야함
- ✓ Text로부터 feature가 추출되면 text classification 문제가 됨
  ex) class-posterior probability $P(C_r|W_r)$

Binary bigram feature function

$$f^{BG}_{c,w_x w_y}(C_r, W_r) = \begin{cases} 1, & \text{if } c = C_r \wedge w_x w_y \in W_r, \\ 0, & \text{otherwise.} \end{cases}$$

주로 word n-gram 사용

4.3 Technical Challenges

Spoken utterance의 특성상 통계 기반, Data-driven approach 가 적절함

Disfluencies in utterance

- ✓ Repetitions
- ✓ False starts
- ✓ Filler words (such as *uh*)

Noise in speech recognition

- ✓ Background noise
- ✓ Mismatched domain
- ✓ Proper names (ex) city or person, etc)

Out-of-domain utterances

- ✓ Irrelevant question

-> Typical call routing system에서는
20-30% word error rate 를 보임 (Gupta et al., 2006)

4.3 Technical Challenges

Technical challenges를 극복할 때 도움이 되는 것들

Note: Redundancy 활용

- ✓ Redundancy는 앞에서 소개한 challenge에 대해서 Robust 하게 만드는데 있어서 매우 중요함
- ✓ Redundancy란 key concepts외에 secondary surrounding words and entities를 의미함
- ✓ Call-type을 위한 key concepts가 없어도, 주변의 secondary words, entities가 correct classification에 도움을 줌
- ✓ Natural language call classification은 template filling에 비해 Redundancy라는 큰 이점이 있음

In ATIS, 원문) *I wond- I'd like to make uh air- I mean flight re-reservation from Boston to uh New York tomorrow, uh no the day after*

인식 결과) *I wond- I'd like to make uh airy mean flights serve nation from boston to uh new yolk tomorrow, uh no the day after*

-> 스펠링 오류가 있지만 ***Flight Reservation*** class로 분류 가능

Note: Stopwords 활용

- ✓ Stopwords 매우 중요! Discriminative words 될 수 있음
- ✓ Discriminative words (ex) ***from, to*** in flight reservation)
- ✓ Call classification & template filling task에서 모두 중요

4.3 Technical Challenges

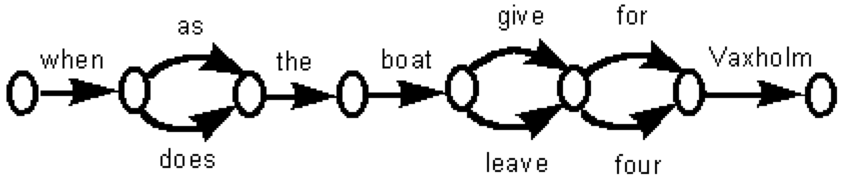
Challenges에 robust하기 위해 제안된 방법들

Methods to be robust to all the challenges

- ✓ 아직 해결 안된 robustness에 대한 이슈들이 많음 (Intent determination error의 1/3은 speech recognition noise 때문)
- ✓ 이를 해결 하기 위해, n-best list, word confusion networks, prior knowledge 를 사용하는 방법들이 제안됨

when as the boat give for Vaxholm
when does the boat give for Vaxholm
when as the boat leave for Vaxholm
when does the boat leave for Vaxholm
when as the boat give four Vaxholm
when does the boat give four Vaxholm
when as the boat leave four Vaxholm
when does the boat leave four Vaxholm

N-best list



Word confusion networks

TABLE I
KEYWORDS USED FOR EACH CLASS ON THE *AP-TITLES* DATASET

Class	*Keywords*
japan	japan, tokyo, yen
bush	bush, george, president, election
israel	israel, jerusalem, peres, sharon, palestinian, israeli, arafat
britx	britain, british, england, english, london, thatcher
gulf	gulf, iraq, saudi, arab, iraqi, saddam,hussein, kuwait
german	german, germany, bonn, berlin, mark
weather	weather, rain, snow, cold, ice, sun, sunny, cloudy
dollargold	dollar, gold, price
hostages	hostages, ransom, holding, hostage
budget	budget, deficit, taxes
arts	art, painting, artist, music, entertainment, museum, theater
dukakis	dukakis, boston, taxes, governor
yugoslavia	yugoslavia
quayle	quayle, dan
ireland	ireland, ira, dublin
burma	burma
bonds	bond, bonds, yield, interest
nielsens	nielsens, rating, t v, tv
boxoffice	box office, movie
tickertalk	stock, bond, bonds, stocks, price, earnings

Prior knowledge

4.4 Benchmark Data Sets

콜센터 데이터셋은 많지만 연구용으로 쓸 수 없기 때문에 다른 데이터 셋을 구축 및 사용함

Call center data

- ✓ 장점: No data scarcity -> millions of calls, transcribed or annotated using semi-automated techniques
- ✓ 단점: sensitive data로 인해 공유할 수 없음

사용 가능

DARPA ATIS corpus

- ✓ 16 different intents (flight reservation, or aircraft capacity)
- ✓ 단점: heavily skewed, flight reservation (70%), 이미 error rate가 5%이하 임

Three notable examples

- ✓ The CMU's Let's Go corpus (from ICSI, AT&T, and Univ of Edinburgh)
- ✓ The DisCoH corpus (from ICSI, AT&T, and Univ of Edinburgh)
- ✓ The European MEDIA corpus (used in the LUNA project)

4.5 Evaluation Metrics - Direct

Direct(연구용) & indirect(비지니스용) evaluation metrics로 나뉨

Accuracy

$$Acc = \frac{\text{\# utterances where the top scoring call-type is amongst the true call-types}}{\boxed{\text{\# of utterances}}}$$

Utterance level에서 평가

- ✓ 하나의 utterance가 하나 이상의 intent를 가질 수 있음 (ex) Can you tell me my balance? I need to make a transfer)
- ✓ 대부분의 경우 두번째 intent가 generic one (greeting, talk to a human agent) or vague one 이면 무시 됨

F-measure(c)

$$F - Measure(c) = \frac{2 \times R(c) \times P(c)}{R(c) + P(c)}$$

$$R(c) = \frac{TP(c)}{TP(c) + \boxed{FN(c)}} \quad \text{missed}$$

$$P(c) = \frac{TP(c)}{TP(c) + \boxed{FP(c)}} \quad \text{False alarm}$$

Threshold

- ✓ Confidence를 고려하기 때문에 accuracy보다 informative 함
- ✓ Confidence level, c 를 넘으면 call-type에 accept됨
- ✓ When c = 0, low precision (many false alarms) + 100% recall (no misses)
- ✓ Confidence value c -> acceptance threshold 로도 사용 될 수 있음
- ✓ 모든 call-type에 대해서 confidence level c 못 넘으면, *Other* class로 맵핑

AT&T HMIHY system에서 사용

		True condition	
	Total population	Condition positive	Condition negative
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error
	Predicted condition negative	**False negative**, Type II error	**True negative**

4.5 Evaluation Metrics - Indirect

Direct(연구용) & indirect(비지니스용) evaluation metrics로 나뉨

The repeat call rate

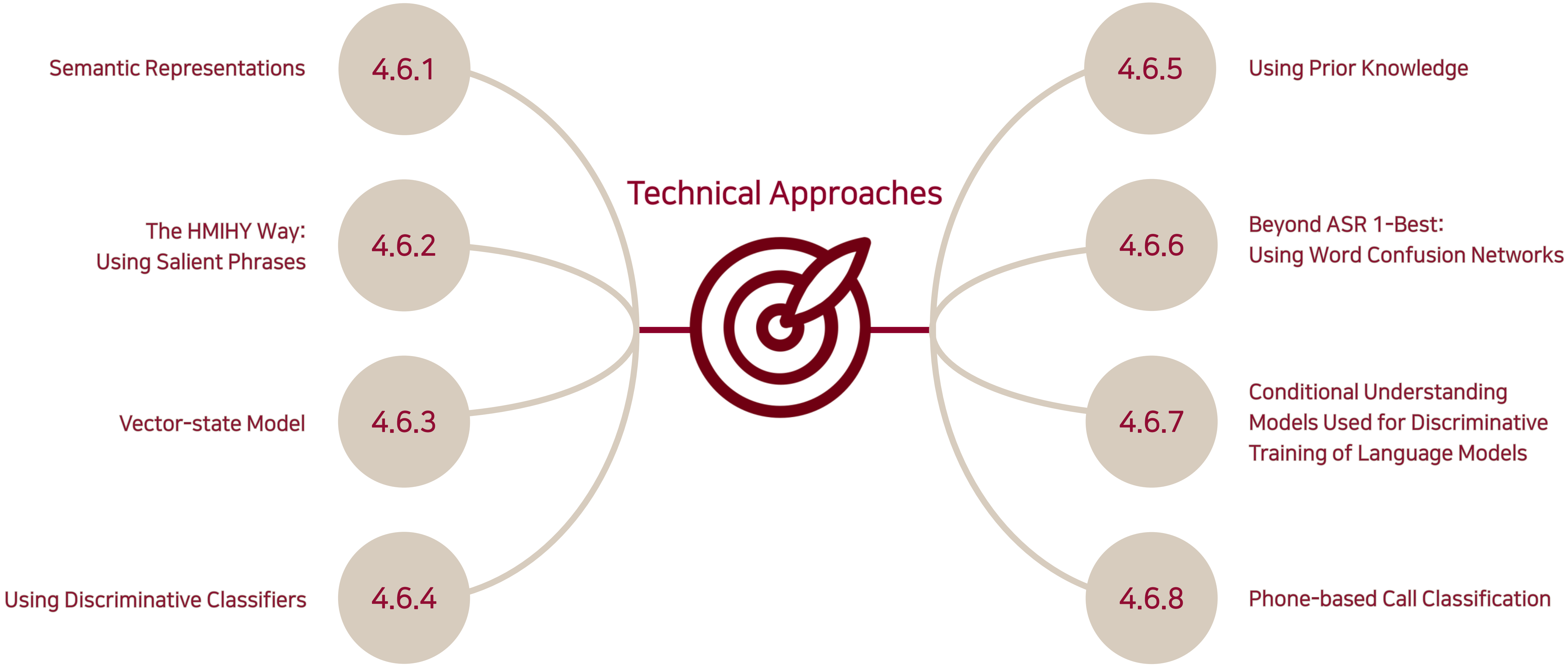
- ✓ 가정: 만족하지 못한 유저는 다시 연락할 것 (Task를 완료하지 못함)
- ✓ 24시간 안에 같은 유저로부터 다시 연락이 오는 경우에 대해서 rate 측정

The hang-up rate

- ✓ Routing 되기 전, 유저에게 직접 서비스에 대한 rating을 받음

4.6 Technical Approaches

Call classification algorithm 에 대한 소개



4.6 Technical Approaches

Call classification algorithm 에 대한 소개

어떠한 semantic representation이 intent를 구성하는지 부터 알아봄

Technical Approaches

- 4.6.1 Semantic Representations
- 4.6.2 The HMIHY Way: Using Salient Phrases
- 4.6.3 Vector-state Model
- 4.6.4 Using Discriminative Classifiers
- 4.6.5 Using Prior Knowledge
- 4.6.6 Beyond ASR 1-Best: Using Word Confusion Networks
- 4.6.7 Conditional Understanding Models Used for Discriminative Training of Language Models
- 4.6.8 Phone-based Call Classification

Semantic Representations (Action oriented vs Intent oriented)

input-action(ex)routes) pair 형태의 semantic representation 도 있지만 이는 generalization이 잘 안된 형태임
-> 왜냐하면 Application마다 call-type이 다르므로, 시간이 지나고, Application이 바뀌면 representation 도 바뀜

Representation language 은 아래 항목에 대해서 Robustness를 달성할 수 있어야함

- ✓ Time-varying nature of the spoken language
- ✓ Time-varying nature of external world (Introduction of a new service)
- ✓ Annotator disagreement

Action oriented semantic representation

"Charge_on_bill"

I see charges on my bill that I do not recognize.
I want credit for some charges on my bill.

predicate-argument $v_i(o_j)$ representation (Intent oriented, one possible way)

- ✓ Linguistics & AI 에서 많이 연구된 Representation임
- ✓ Utterance를 system이 하는 action 이 아닌 speaker 의 intent 로 labeling!
- ✓ Predicate v_i 는 Domain-independent verb (Action)
- ✓ Argument o_j는 Domain-related object
- ✓ 재사용에 용이함

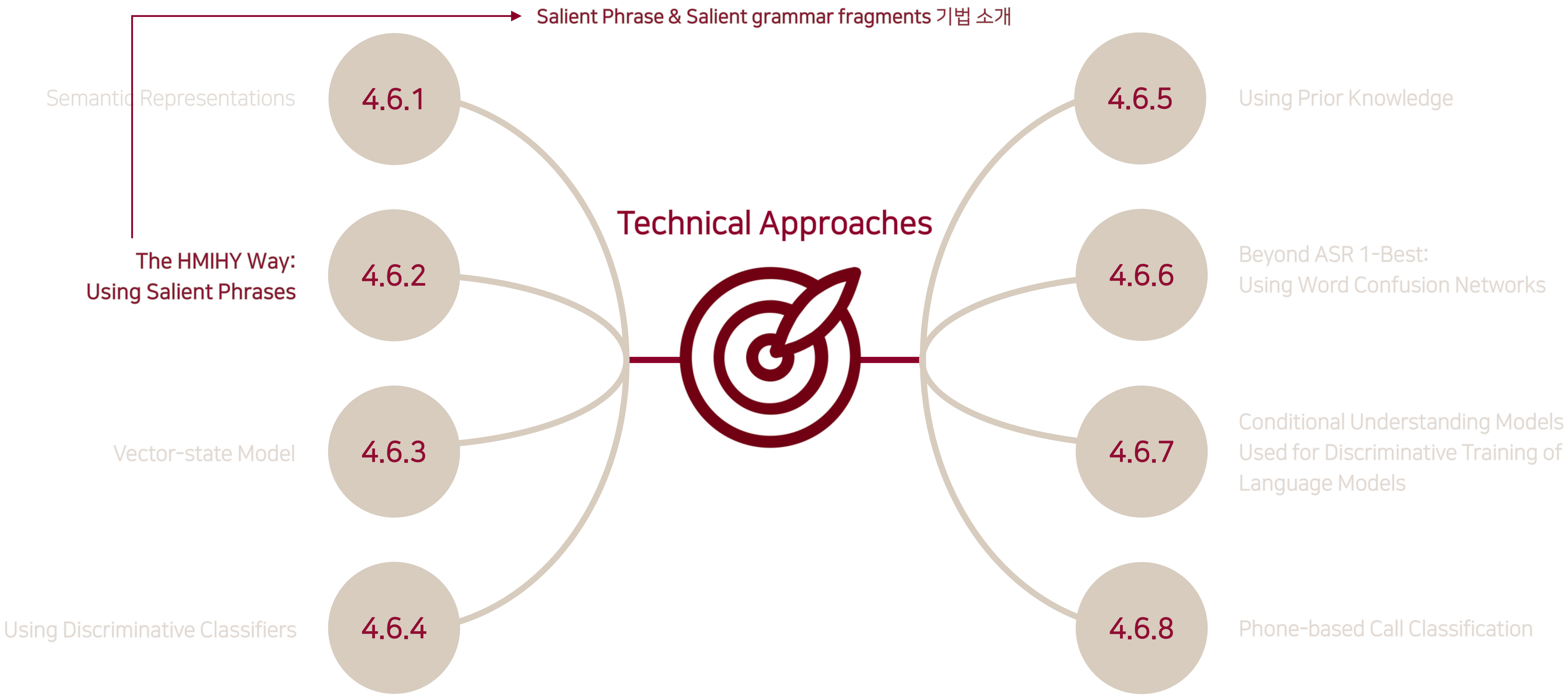
Intent oriented semantic representation

I see charges on my bill that I do not understand.	Explain(Bill_Charge)
I want credit for some charges on my bill.	Request(Credit)
I am just wanting to tell you that I have made the payment.	Report(Payment)
I am calling to check if you received my payment.	Verify(Payment)
I dialed a wrong number.	Report(WrongNumber)

(Gupta et al., 2006)

4.6 Technical Approaches

Call classification algorithm 에 대한 소개



4.6 Technical Approaches

Call classification algorithm 에 대한 소개

THE HMIHY way: Using Salient Phrases

Salient phrase for *Calling Plans* call-type

- ✓ Salient Phrases에 매우 의존함
- ✓ Salient Phrase란? Input utterance내에서 특정 call-type에 대해 핵심적인 구문을 의미함
  ex) "I would like to change oh about long distance service two one charge nine cents a minute"

$$S(f) = \sum_{c_k} P(c_k|f)I(f, c_k)$$ Salience score

$$I(x, y) = log\frac{P(x, y)}{P(x)}$$

$$I(v_1, v_2) = \log_2[P(v_2 | v_1)/P(v_2)]$$

x, y의 mutual information(MI)

f Fragment(word N-gram)

c_k Call-type

Phrase fragments 라고도 함

0.5 < $maxP(c_k|f)$ <0.9 이면 괄호

Stopwords
-> more discriminative

Phrase	$I(f_n^n, f_1^{n-1})$	$maxP(c_k\|f)$	$\text{argmax}_{c_k} P(c_k\|f)$
I would like	7.1	0.24	
made a long distance	7.4	0.93	Billing_Credit
long distance	7.3	0.55	(Billing_Credit)
the area code for	5.6	0.92	Area_Code
area code	6.9	0.65	(Area_Code)

Figure 4.3 Example salient phrases with associated call-types

Salient phrase 선택을 위한 3가지 기준들

- ✓ Salience score가 기준보다 낮은 경우 제외 $S(f) < \theta_1$
- ✓ 특정 call-type에 속하지 않는다고 판단 될 경우 제외 $maxP(c_k|f) < \theta_2$
- ✓ 마지막단어와 전단어의 mutual information 값이 기준보다 낮은 경우 제외 $I(f_n^n, f_1^{n-1}) < \theta_3$

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Consider the two salient phrases “a wrong number” and “the wrong number”.
Clearly, these should not be treated independently, but rather combined into a single unit

Table 2
Salient and background phrase fragments

$0.5 < maxP(c_k|f) < 0.9$ 이면 괄호

MI	Phrase fragments	P_{max}	Call-Type
7.4	made a long distance	0.93	Billing credit
7.3	long distance	0.55	(Billing credit)
7.1	I would like	0.24	
6.9	area code	0.65	(Area code)
6.3	could you tell me	0.37	
5.6	the area code for	0.92	Area code
5.3	I'm trying	0.33	
5.0	a wrong number	0.98	Billing credit
4.9	a long distance call	0.62	(Billing credit)
4.8	the wrong number	0.98	Billing credit
4.4	I'm trying to	0.33	
4.3	long distance call	0.62	(Billing credit)
4.3	I just made a	0.93	Billing credit
4.1	I'd like to	0.18	

Phrase fragments 라고도 함

$0.5 < maxP(c_k|f) < 0.9$ 이면 괄호

Stopwords
-> more discriminative

Phrase	$I(f_n^n, f_1^{n-1})$	$maxP(c_k\|f)$	$\text{argmax}_{c_k} P(c_k\|f)$
I would like	7.1	0.24	
made a long distance	7.4	0.93	Billing_Credit
long distance	7.3	0.55	(Billing_Credit)
the area code for	5.6	0.92	Area_Code
area code	6.9	0.65	(Area_Code)

Figure 4.3 Example salient phrases with associated call-types

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Salient grammar fragments (SGFs)

용어 정리

- ✓ F(wrong) = (a | the | was) wrong (number | call | eos)
- ✓ | == 'or'
- ✓ — == 'previous row'

- ✓ Salient phrase에 단어를 더하면 discriminative power는 좋아지지만 coverage는 떨어짐
- ✓ Salient phrase들은 salient grammar fragments (SGFs)로 클러스터링 됨
- ✓ 클러스터링 사용하는 이유는 더 generalization이 잘 되기 때문임
- ✓ Discriminative power ($maxP(c_k|f)$) 도 보존하면서 Coverage ($P(f|c_k)$) 도 높이는 방향으로 구축
- ✓ 각 Call-type에 해당하는 구축된 grammar fragments는 $SGF(c_k)$로 표기함

Fragment	$max P(c_k \mid f)$	$P(f \mid c_k)$
wrong	0.92	0.48
wrong number	0.98	0.41
(a|the|was)wrong(number|call|eos)	0.97	0.42
F(wrong)|F(dialed)	0.95	0.50
SGF(Billing_Credit)	0.95	0.64

Figure 4.4 An example for building a salient grammar fragment (SGF)

Table 3
Growth of a salient grammar fragment for distinguishing billing credit queries

Prob correct P(Cr | *G*)	Coverage P(*G* | Cr)	Fragment *G*
0.92	0.48	wrong
0.98	0.41	wrong number
0.95	0.45	wrong (number |eos |call)
0.97	0.42	(a |the |was) wrong (number |eos |call)
0.95	0.50	*F*(wrong) | *F*(dialed)
0.95	0.57	*F*(wrong) | *F*(dialed) | *F*(credit)
0.95	0.59	— | *F*(disconnected)
0.95	0.64	— | *F*(misdialed) | *F*(cut off)

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Salient grammar fragments (SGFs)

- ✓ SGFs 구축 후, finite state machines(FSMs) 형태로 사용됨
- ✓ ASR output 또는 input text s 에 대해서 FSM operation을 통해 SGFs 의 발생여부를 탐색함
- ✓ Posteriori probability에서 최대값을 갖는 class를 선택함
- ✓ 최대값이 threshold를 넘지 못하는 경우, utterance는 ***other*** class로 분류됨 $P(c_k|f_{i(s)}) < \theta$

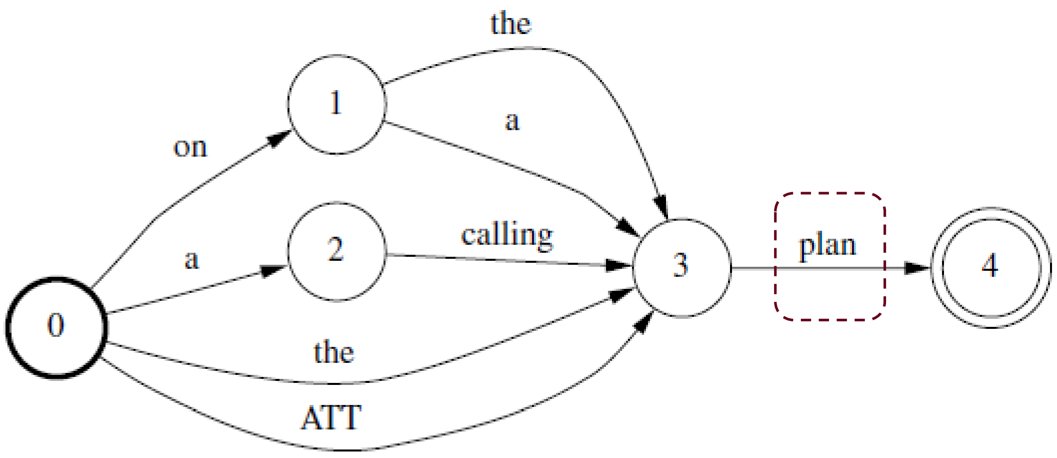


Figure 4.5 An example salient grammar fragment

$$i(s) = \operatorname*{argmax}_i (\max_{c_k} P(c_k|f_i))$$ Feature function 선정 후

$$K(s) = \operatorname*{argmax}_{c_k} P(c_k|f_{i(s)})$$ Call-type Posterior prob 계산

Note:
확률은 후에 confidence score로
사용 가능하며 이를 위해 정규화 필요함

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Salient grammar fragments (SGFs)

- ✓ Call-type에 대한 raw score를 actual probability로 만들기 위해서 normalize 필요
- ✓ Logistic regression을 통해서 0~1 사이 확률 값으로 변환
- ✓ β는 Newton-Raphson Method (Agresti, 1990) 를 통해 학습함
- ✓ Normalization를 통해 reliable confidence score를 얻을 수 있음

Call-type에 대한 raw score

$$\boxed{P(c_k | f_{i(s)})}$$

$$p' = \frac{1}{1 + e^{-(\beta_0+\beta_1\times p+\beta_2\times l+\beta_3\times c)}}$$

length

Coverage
(% of the words occurring inside
a salient phrase in an utterance)

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

IR에서 사용되었던 vector 기반의 모델을 분류에 적용

Technical Approaches

- 4.6.1 Semantic Representations
- 4.6.2 The HMIHY Way: Using Salient Phrases
- 4.6.3 **Vector-state Model**
- 4.6.4 Using Discriminative Classifiers
- 4.6.5 Using Prior Knowledge
- 4.6.6 Beyond ASR 1-Best: Using Word Confusion Networks
- 4.6.7 Conditional Understanding Models Used for Discriminative Training of Language Models
- 4.6.8 Phone-based Call Classification

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Vector-state Model (Chu-Carroll and Carpenter, 1999)

- ✓ IR 기반의 모델, the Bell Labs에서 시도 Stop-words 제거, stemming 등 IR에서 사용되는 기법 적용
- ✓ $F \times K$ term-document matrix 생성 (term은 word 또는 word phrase, document는 클래스, call-type)
- ✓ 각 term은 tf-idf 로 weighted vector 로 표현됨
- ✓ Baseline algorithm은 user utterance 와 route의 minimum cosine distance (전통적인 IR 방법대로)

Baseline algorithm에 비해 개선된 점

- ✓ 1st improvement: 0 ~ 1의 값을 갖는 similarity score (sigmoid function 사용)
- ✓ 2nd improvement: training corpus 가 클 때, SVD로 효율을 높임 (Routing matrix $W = USV^T$로 분해), 후에 LSA 방법으로 발전 (Later Cox and Shahshahani, 2001)

term-document matrix의 transpose 된 형태

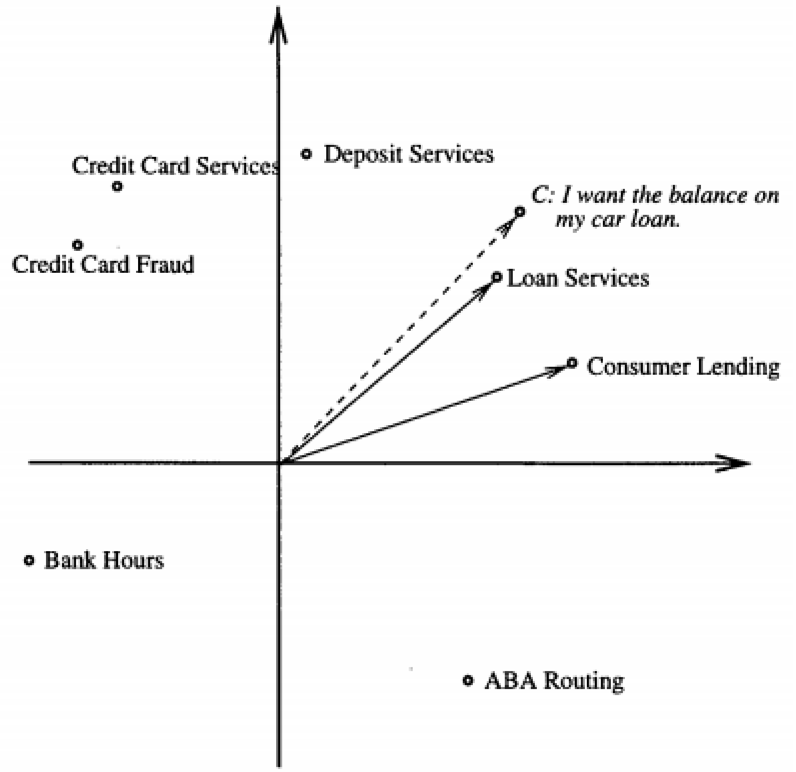


Figure 4
Two-dimensional vector representation for the routing module.

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Vector-state Model (Chu-Carroll and Carpenter, 1999)

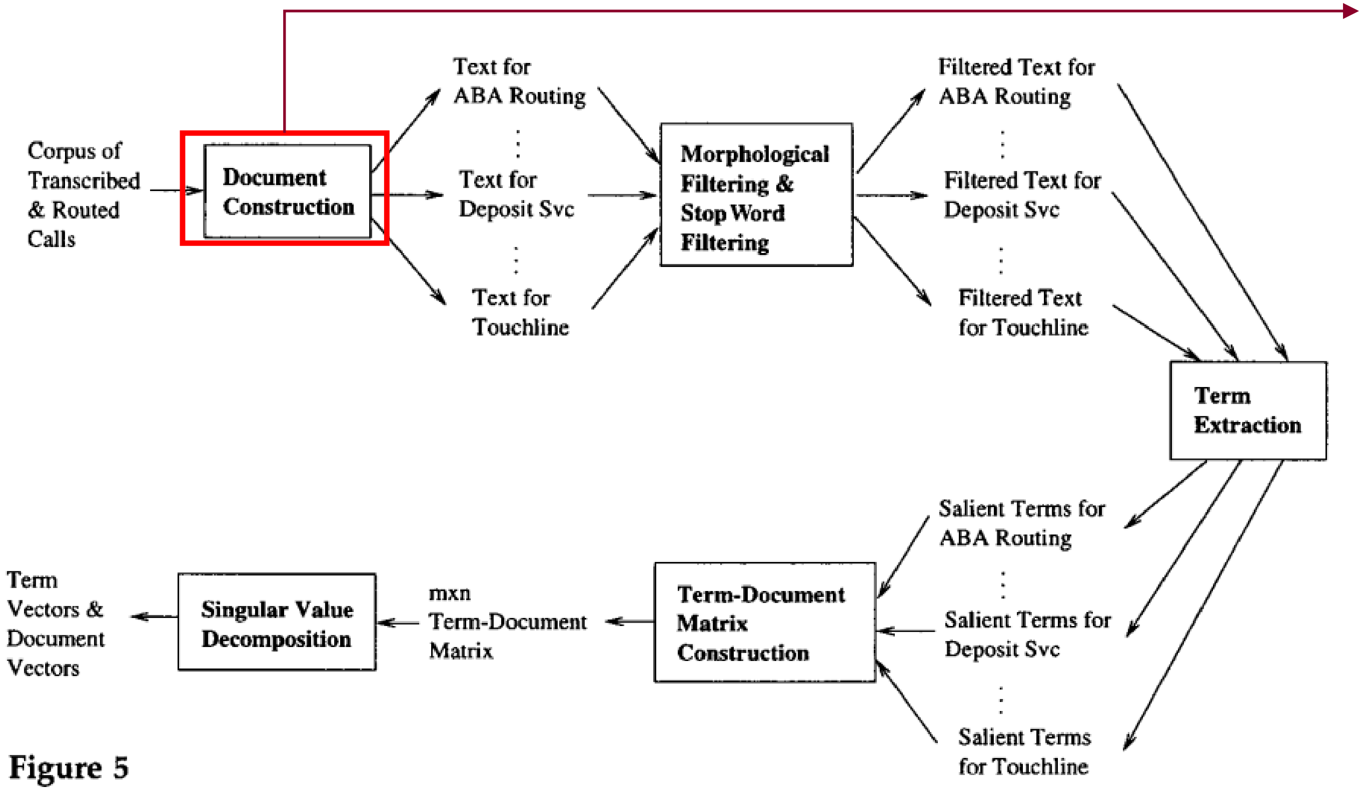


Figure 5
Training process for the routing module.

- ✓ 각 destination(call-type) 마다 One (virtual) Document 생성
- ✓ Document는 caller의 utterance(raw text)를 담고 있음
  ex) call-type이 deposit service 인 경우,
  "I want to check the balance in my checking account"

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Vector-state Model (Chu-Carroll and Carpenter, 1999)

- ✓ Service, services, servicing -> service 라는 root form으로 변경
- ✓ Ignore list (um, uh, ah, aah, ahh) 제거
- ✓ Stop list (the, be, for) 제거

논문에서 25개로 정함

논문에서 1,200개로 정함

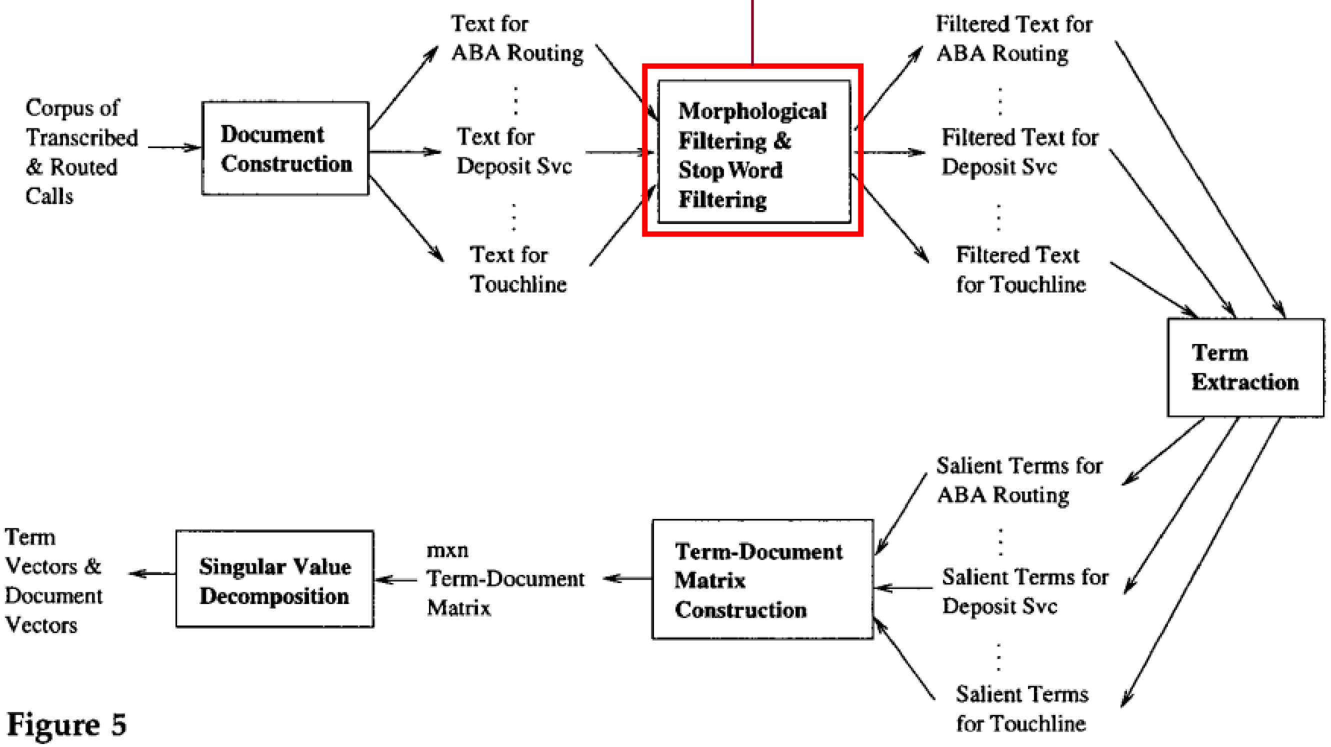


Figure 5
Training process for the routing module.

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Vector-state Model (Chu-Carroll and Carpenter, 1999)

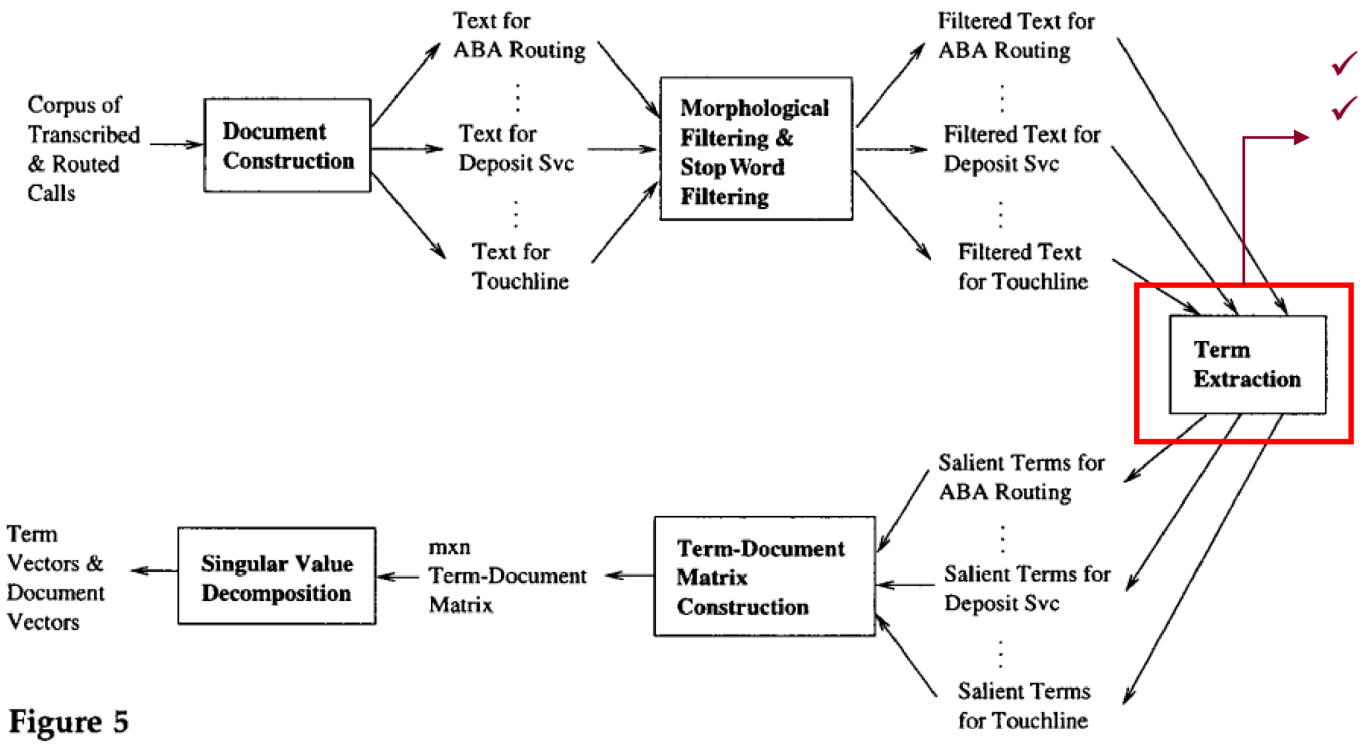


Figure 5
Training process for the routing module.

논문에서
unigram은 2, bigram, trigram은 3으로 정함

- ✓ Co-occurrence, n-gram terms 추출
- ✓ 모든 n-gram term의 숫자를 카운트 후 threshold 정하고, salient terms이 될 만한 n-gram을 추출
- ✓ Salient terms으로 각 document에 대해서 bag of salient term 적용
- ✓ N-gram 내에 있는 단어들에 대해서도 k-gram (1<k<n) 적용해서 추출
  ex) word sequence "Checking account balance"
  -> trigram: "check+account+balance"
  -> bigrams: "check+account" and "account+balance"
  -> unigrams: "check", "account" and "balance"

논문에서 trigrams 62개, bigrams 275개, unigrams 420개 만큼 추출됨

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Vector-state Model (Chu-Carroll and Carpenter, 1999)

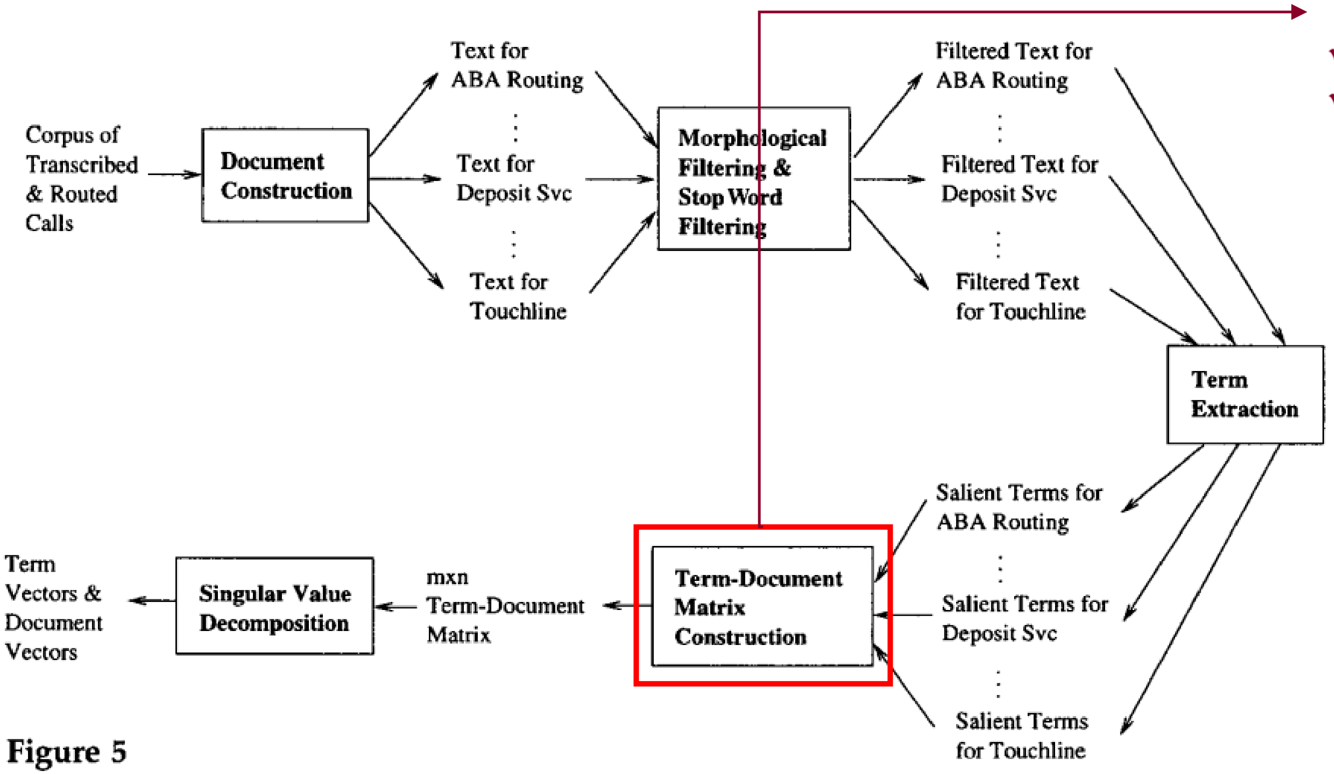


Figure 5
Training process for the routing module.

- ✓ call-type(destination)에 대한 Bag of salient term로 $m \times n$ term-document frequency matrix A 를 만듦
- ✓ m 은 salient terms 개수
- ✓ n 은 call-type(destination) 개수

$$A \xrightarrow{\text{Normalized to unit length}} B \xrightarrow{\text{IDF 적용}} C$$

$$B_{t,d} = \frac{A_{t,d}}{(\sum_{1 \le e \le n} A_{t,e}^2)^{1/2}} \qquad C_{t,d} = IDF(t) \cdot B_{t,d}$$

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Vector-state Model (Chu-Carroll and Carpenter, 1999)

- ✓ Weighted term-document frequency matrix C에 대해서 SVD적용
- ✓ Document vector의 dimension의 크기 줄여줌
- ✓ Documents는 $V_r \cdot S_r$로 나타낼 수 있음
- ✓ Caller의 Utterance도 비슷한 과정을 거쳐서 $V_q \cdot S$로 나타냄(Pseudo doc)

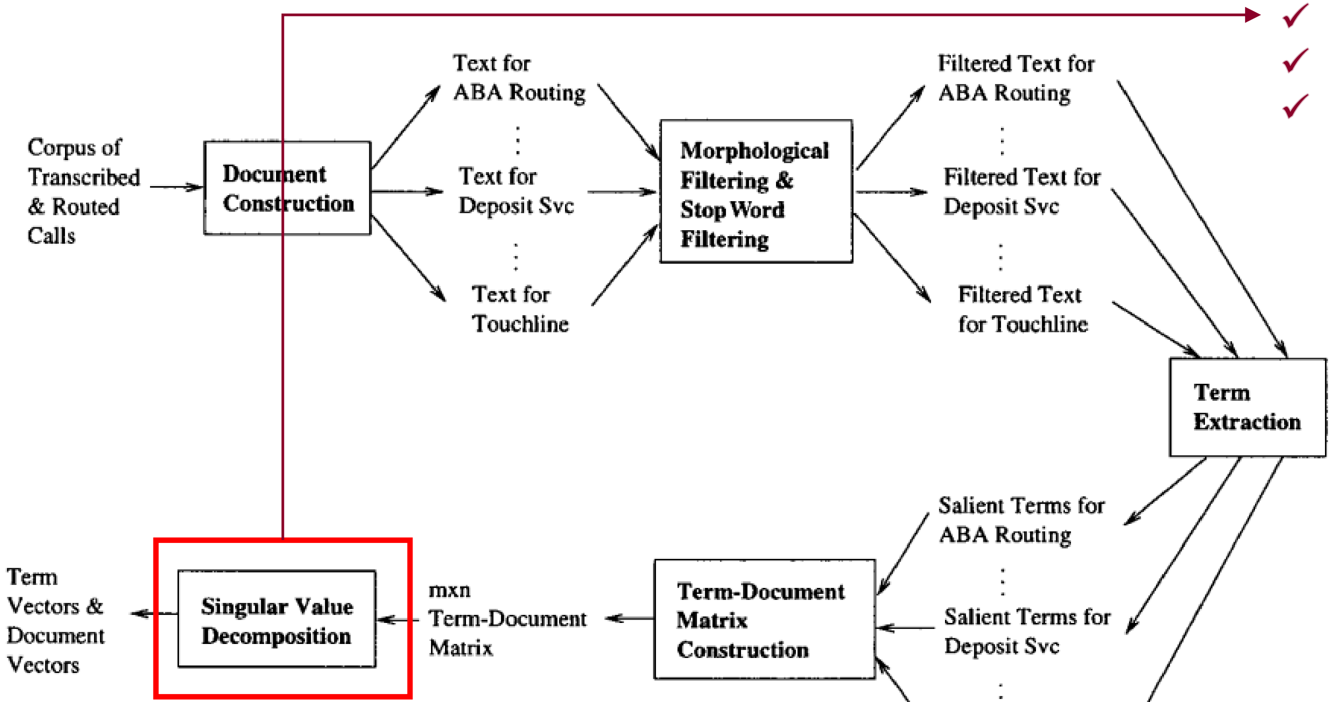


Figure 5
Training process for the routing module.

$$C = U \cdot S \cdot V^{\mathrm{T}},$$

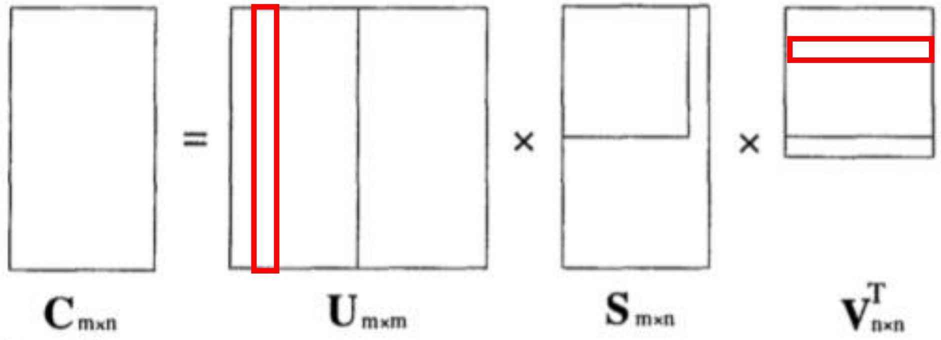


Figure 6
Singular value decomposition.

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Vector-state Model (Chu-Carroll and Carpenter, 1999)

- ✓ 기존 cosine similarity 방식은 비슷한 정도를 나타내긴 하지만,
  correct routing을 위한 closeness의 likelihood를 directly 하게 나타내진 않음
- ✓ 개선 방법으로 Logistic regression 사용, d_a, d_b는 destination(call-type)에 따른 coefficient (least-squares method 로 구함)

$$\boxed{cos(\mathbf{x}, \mathbf{y})} = \frac{\mathbf{x} \cdot \mathbf{y}^{\mathrm{T}}}{\sqrt{\sum_{1 \le i \le n} x_i^2 \cdot \sum_{1 \le i \le n} y_i^2}}$$

$$Conf(d_a, d_b, x) = 1/(1 + e^{-(d_a \boxed{x} + d_b)})$$

(a) Cosine Scores

1.	Consumer Lending	0.979
2.	Loan Services	0.260
3.	Home Loans	0.077
4.	Collateral Control	0.069
5.	Operator	0.038

(b) Confidence Scores

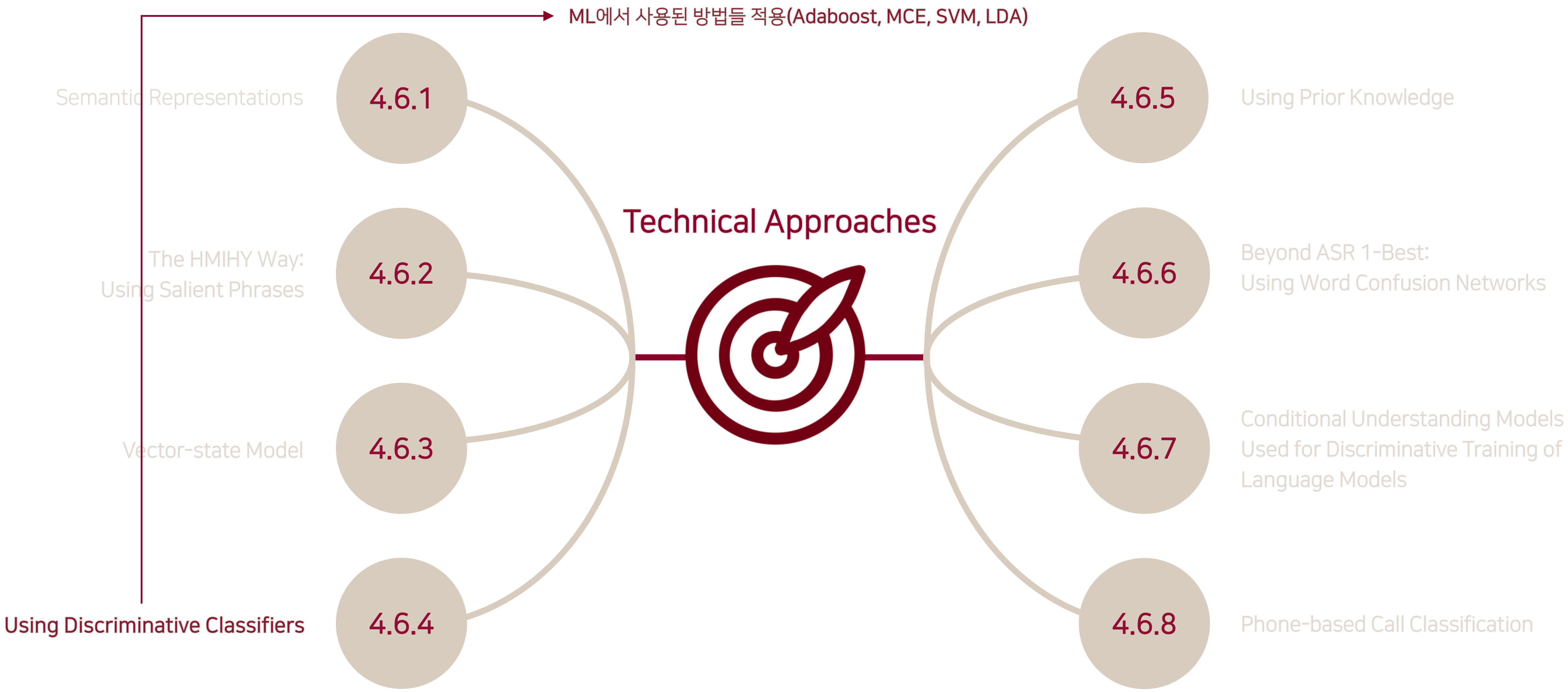
1.	Consumer Lending	0.913
2.	Deposit Services	0.070
3.	PC Banking	0.049
4.	Loan Services	0.035
5.	Auto Leasing	0.032

Figure 9
Ranking of candidate destinations.

- ✓ 적용 결과 Ranking 바뀜
- ✓ 정확도도 92.2% -> 93.5%로 올랐다고 함

4.6 Technical Approaches

Call classification algorithm 에 대한 소개



4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Using Discriminative Classifiers

- ✓ Intent determination은 classification 문제로 볼 수 있으므로 ML의 discriminative classification techniques 적용
- ✓ Discriminative classifier가 generalization performance 가 좋아서 spoken dialogue application 에서 성공적이였음 (Gupta et al., 2006)
- ✓ Input space의 차원이 매우 크기 때문에, **SVMs** (Vapnik, 1998)이나 **Adaboost** (Schapire and Singer, 2000) 방식이 사용됨
- ✓ Context를 고려하기 위해서, n-grams 방식 사용

AdaBoost.MH

- ✓ Iterative algorithm 임, 각 iteration 마다 weak classifier 가 훈련되고 weak classifiers 를 합친 combined classifier 사용
- ✓ AT&T HMIHY system 에 적용됨 (Schapire and Singer, 2000)
- ✓ 성능은 Salience-based classification과 비슷함. 특히 high false alarm rate 부분에서는 더 뛰어남

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Using Discriminative Classifiers

term-document matrix의 transpose 된 형태

Minimum classification error (MCE) based approach

- ✓ Bell Labs에서 연구됨 (Kuo and Lee, 2000)
- ✓ Routing matrix R의 elements 를 discriminative training을 통해 수정해서 Vector-based routing 방법을 개선함
  특히 low rejection rates 부분을 개선함
- ✓ MCE의 nonlinear optimization problem 을 해결하기 위해 generalized probabilistic descent (GPD) algorithm 사용

나중에 Boosting기법과 함께 응용됨 (ARF, Zitouni et al., 2003)

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Using Discriminative Classifiers

용어 정리

- ✓ R_t: t번째 iteration에서 parameter set
- ✓ x_t: training data with class k

Minimum classification error (MCE) based approach

$$d_k(\vec{x}, R) = -g_k(\vec{x}, R) + G_k(\vec{x}, R)$$

Misclassification 함수, discriminant 함수와 anti-discriminant 함수로 나뉨

$$g_k(\vec{x}, R) = \vec{r}_j.\vec{x}$$

Discriminant 함수 (normalized query $\vec{x}$ 와 destination vector $\vec{r}_j$ 의 내적),
cosine score와 같음! query, destination vector 모두 normalized 된 상태이기 때문 $\frac{\vec{r}_j \cdot \vec{x}}{\|\vec{r}_j\|\|\vec{x}\|}$

$$G_k(\vec{x}, R) = \left(\frac{1}{K-1} \sum_{j \neq k, 1 \leq j \leq K} g_j(\vec{x}, R)^{\eta}\right)^{\frac{1}{\rho}}$$

Anti-discriminant 함수,
정답 클래스를 제외한 나머지($K-1$)에 대한 score 계산

$$l_k(\vec{x}, R) = \frac{1}{1 + e^{-\gamma d_k(\vec{x}, R) + \theta}}$$

Loss 함수,
misclassification 함수를 인자로 갖는 sigmoid 함수로 정의

$$\delta R_t = -\epsilon_t \nabla l_k(\vec{x}_t, R_t)$$

갱신할 때 사용하는 텀은 training data $\vec{x}_t$(label k)에 대한 loss 함수 l_k 의 Gradient 사용

$$R_{t+1} = R_t + \delta R_t$$

Routing matrix 가 iteratively하게 갱신 됨

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Using Discriminative Classifiers

Large margin classifier (SVM)

- ✓ Binary SVM classifier를 generalization 해서 multiclass call classification problem에 적용하긴 어려움
- ✓ 여러 개의 binary classifier를 조합 할 수 있는 모델로 global optimization을 통해 문제 해결 (Haffner et al., 2003)
- ✓ AT&T's HMIHY 의 call-type classification error rate를 낮춤 (특히 false rejection rate를 낮춤)

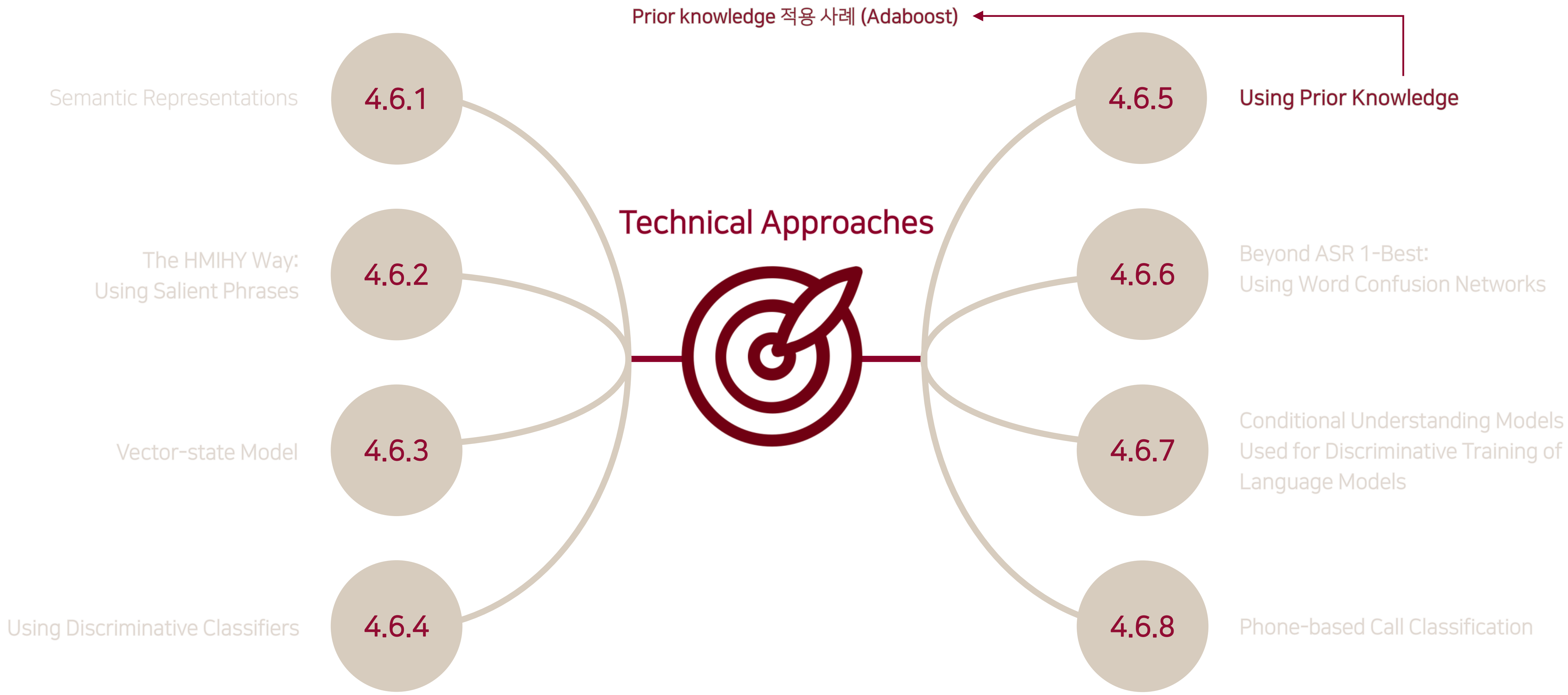
Vector를 분류가 잘못 될 때마다 adaptive하게 변경하는 기법

Linear discriminant analysis (LDA) based approach

- ✓ 비슷한 시기에 연구됨 (Cox, 2003)
- ✓ Generalized probabilistic descent (GPD), corrective training (CT), linear discriminant analysis (LDA) 기법 사용
- ✓ LDA는 training vector, test vector의 linear transformation 에 의존함
- ✓ LDA가 maximization of distance 기반이기 때문에 cosine distance가 아닌 Euclidean distance를 사용
- ✓ Cox 논문에서는 test set에서 큰 성능 개선을 보이지는 못함 (training data의 부족)

4.6 Technical Approaches

Call classification algorithm 에 대한 소개



4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Using Prior Knowledge

ex) ATIS CMU(Ward and Issar, 1994), MIT (Seneff, 1992) system

- ✓ Slot filling task에서는 manually written semantic grammars 사용이 보편적이었음
- ✓ 반면, Intent determination task에서는 키워드 기반의 시스템 외에는 rule에만 의존하는 시스템은 없었음
- ✓ Intent determination 에 prior knowledge 사용 시도
  ex) Boosting 기법과 함께 call-type에 대한 manually written grammars 적용 (Schapire et al., 2005)

중요도 계수, 실험적으로 결정됨

Loss function :

x_i 가 true class면 y_{ic} 는 1, 아니면 -1

$$sum_i \sum_c [\ln(1 + e^{-y_{ic} h(x_i,c)}) + \eta RE(\pi(c|x_i) \| \sigma(h(x_i, c)))]$$

Relative entropy measure
(KL divergence)
-> prior model에 맞게 학습되었는지
체크하기 위한 목적

Adaboost negative log conditional likelihood

prior knowledge로 예측한 확률
π는 rule을 의미함

데이터기반의 모델((logistic Adaboost)이 예측한 확률

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Using Prior Knowledge

Loss function :

$$sum_i \sum_c [\ln(1 + e^{-y_{ic} h(x_i, c)}) + \eta RE(\pi(c|x_i) || \sigma(h(x_i, c)))]$$

prior knowledge로 예측한 확률
π는 rule을 의미함

용어정리
- ✓ l : 클래스
- ✓ n_w: 키워드가 존재하는 class들의 개수(적을 수록 효과적인 키워드)
- ✓ k: 전체 클래스 개수
- ✓ w: 키워드가 존재함
- ✓ $\overline{w}$: 키워드가 존재하지 않음

$$\pi(\ell|w) = \begin{cases} \frac{0.9}{n_w}, & \text{if } w \text{ is a keyword for } \ell \\ \frac{0.1}{(k-n_w)}, & \text{otherwise} \end{cases}$$

$$\pi(\ell|\overline{w}) = 1/k$$

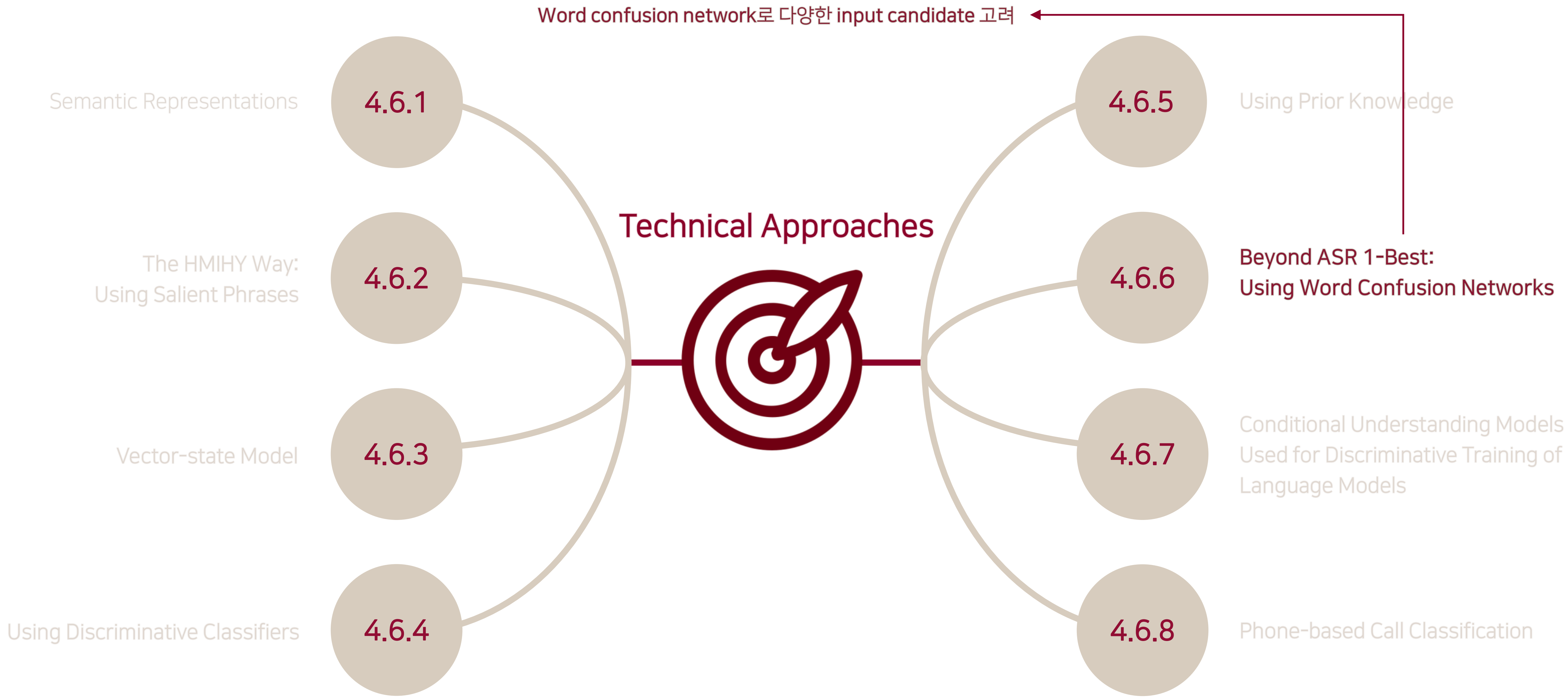
TABLE I
KEYWORDS USED FOR EACH CLASS ON THE *AP-TITLES* DATASET

Class	*Keywords*
japan	japan, tokyo, yen
bush	bush, george, president, election
israel	israel, jerusalem, peres, sharon, palestinian, israeli, arafat
britx	britain, british, england, english, london, thatcher
gulf	gulf, iraq, saudi, arab, iraqi, saddam,hussein, kuwait
german	german, germany, bonn, berlin, mark
weather	weather, rain, snow, cold, ice, sun, sunny, cloudy
dollargold	dollar, gold, price
hostages	hostages, ransom, holding, hostage
budget	budget, deficit, taxes
arts	art, painting, artist, music, entertainment, museum, theater
dukakis	dukakis, boston, taxes, governor
yugoslavia	yugoslavia
quayle	quayle, dan
ireland	ireland, ira, dublin
burma	burma
bonds	bond, bonds, yield, interest
nielsens	nielsens, rating, t v, tv
boxoffice	box office, movie
tickertalk	stock, bond, bonds, stocks, price, earnings

키워드 기반의 Prior knowledge

4.6 Technical Approaches

Call classification algorithm 에 대한 소개



4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Beyond ASR 1-Best: Using Word Confusion Networks

- ✓ SLU에서는 ASR error에 Robust 한 것이 매우 중요함
- ✓ Spontaneous conversational speech에서 word error rate(WER)이 20-30% 정도를 기록함

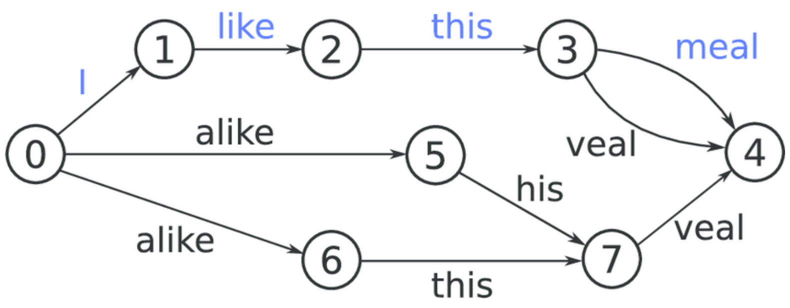
- ✓ 기존 방법의 문제, 1-best 위주의 선택
  1-best ASR output: "I have a few questions about my maybe ill."
  original utterance : "I have a few questions about my May bill"

- ✓ 1-best 만 의존하지 말자, intermediate ASR output 도 사용하자! (Rescoring 전략)
  ex) n-best lists / word lattices / word confusion networks

- ✓ Lattice 란?
  Directed graph of words (가능한 문장들을 인코딩 할 수 있는 형태)

when as the boat give for Vaxholm
when does the boat give for Vaxholm
when as the boat leave for Vaxholm
when does the boat leave for Vaxholm
when as the boat give four Vaxholm
when does the boat give four Vaxholm
when as the boat leave four Vaxholm
when does the boat leave four Vaxholm

N-best list



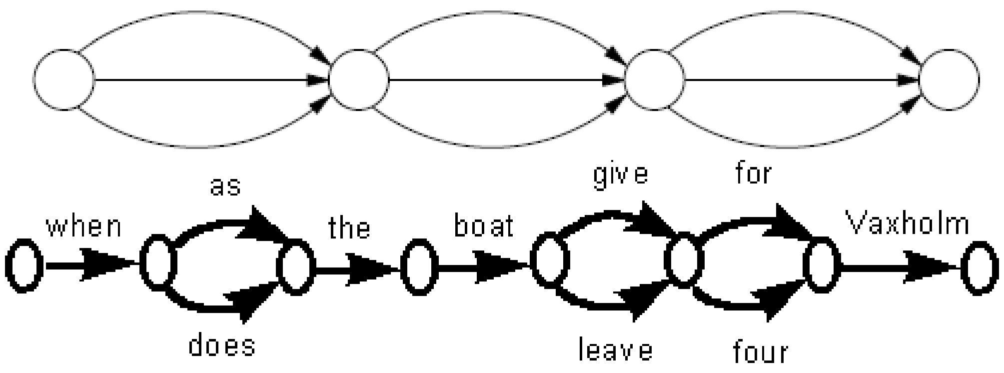
Lattice

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Beyond ASR 1-Best: Using Word Confusion Networks

Word Confusion Network:



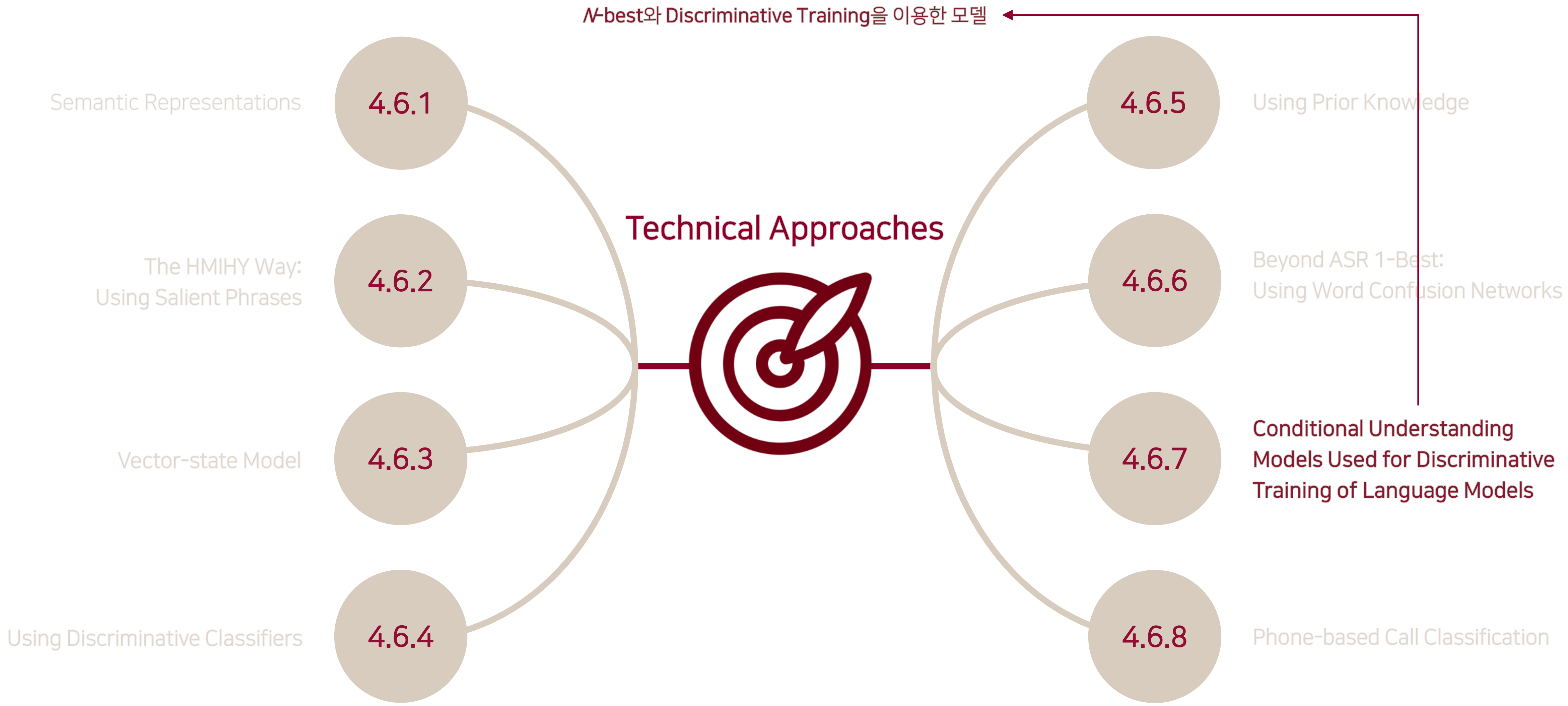
- ✓ Word confusion networks (WCNs) 란?
  Lattice의 특수한 형태, word set의 concatenation 으로 구성 됨
- ✓ AT&T HMIHY 에서 salient phrase를 더 잘 찾기 위해 사용됨(Tur et al., 2002)
- ✓ 1-best로 ASR에서 salient phrases 매칭 시킬 때보다 WCNs 방법(다른 매칭들 고려)이 더 효과적인 임
- ✓ Salient phrases를 더 잘 detect하기 위해 AT&T HMIHY system에 적용됨 (Tur et al., 2002)

WCNs의 장점들

- ✓ 구조상 word alignment가 되어있음 (language processing에 유용할 수 있음)
- ✓ WCNs안의 단어들에 대해 Posterior probability를 구할 수 있고 이를 통해 confidence score를 구할 수 있음
- ✓ 기존 Lattice보다 훨씬 작은 형태임 (1/100 of those ASR lattices), 성능도 높음 (Mangu et al., 2000)

4.6 Technical Approaches

Call classification algorithm 에 대한 소개



4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Conditional Understanding Models Used for Discriminative Training of Language Models

Automatic Speech Recognition

목표
- ✓ Acoustic signal로부터 word sequence 생성

$$\hat{W}_r = \arg\max_{W} P(W|X_r).$$

AM (HMMs)

LM (n-gram model with Markovian assumption)

$$\hat{W}_r = \arg\max_{W} \frac{\boxed{P(X_r|W)}\boxed{P(W)}}{P(X_r)},$$
$$= \arg\max_{W} P(X_r|W)P(W).$$

$$\hat{W}_r = \arg\max_{W} \left[P^{\boxed{\frac{1}{L}}}(X_r|W)P(W) \right]$$

Knowledge sources (KSs) 적용 (Baker et al., 2009)
- ✓ Lattice or N-best로 생성한 word-sequence hypotheses 를 re-rank 할 때 사용

WER의 감소가 CER의 감소로 이어진다고 볼 순 없음

Word error rate(WER) 발생

Class error rate (CER) 발생

word sequence (Best hypothesis)

label

utterance

1st step

2nd step

$P(C_r|W_r)$

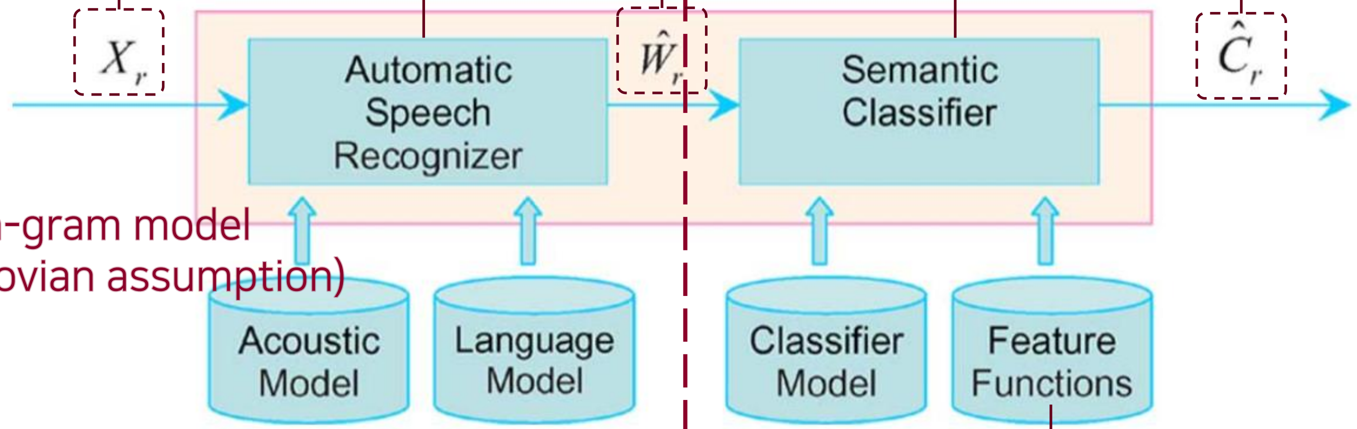


Fig. 1. Traditional spoken utterance classification system is composed of two isolated stages: an automatic speech recognition system is followed by a semantic classification system.

$$f^{BG}_{c,w_x w_y b}(C_r, W_r) = \begin{cases} 1, & \text{if } c = C_r \wedge w_x w_y b \in W_r \\ 0, & \text{otherwise.} \end{cases}$$

AM의 텀이 normalize 안돼서 확률이 아님
LM scaling factor L 도입

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Conditional Understanding Models
Used for Discriminative Training of Language Models

- ✓ 만약 Best hypothesis $\widehat{W}_r$가 틀리면 Semantic classifier가 틀릴 확률이 높음
- ✓ 하나만 의존하지 말고, 여러 경우의 수에 대해서 가능성을 열어 놓고 rescoring 하자 -> *N*-best

TABLE I
ASSIGNMENT OF SEMANTIC CLASSES TO WORD SEQUENCES

SEMANTIC CLASS	WORD SEQUENCE	$D(.)$
C_r^0: GROUND SERVICE	W_r^0: WHAT IS THE TRANSPORTATION IN ATLANTA	-20.04
C_r^1: FARE	W_r^1: WHAT IS THE ROUND TRIP FARE FROM ATLANTA	$\boxed{-17.56}$
C_r^2: CITY	W_r^2: WHAT IS THE TRANSPORTATION ATLANTA	-25.46
C_r^3: FLIGHT	W_r^3: WHAT IS THE TRANSPORTATION AND ATLANTA	-28.49
C_r^4: FARE BASIS	W_r^4: WHAT IS THE ROUND TRIP FARE FROM THE ATLANTA	-27.98
C_r^5: AIRPORT SERVICE	W_r^5: WHAT IS THE TRANSPORTATION THE ATLANTA	-29.09
GROUND SERVICE	WHAT IS THE GROUND TRANSPORTATION IN ATLANTA	

N-best list

현재로서는 잘못 예측한 샘플의 스코어가 가장 높음
word sequence의 rescoring 및
LM, classifier의 parameter 학습 필요

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Conditional Understanding Models Used for Discriminative Training of Language Models

Discriminative Learning Algorithm

목표

✓ Discriminative training은 인식/분류 결과 높이는 것

$$\hat{C}_r = \arg\max_{C_r}\left[\sum_{W_r} P(C_r, W_r|X_r)\right]$$

AM (HMMs)

LM (n-gram model with Markovian assumption)

$$= \arg\max_{C_r}\left[\sum_{W_r} P(C_r|W_r, X_r)\boxed{P^{\frac{1}{L}}(X_r|W_r)}\boxed{P(W_r)}\right]$$

$$\cong \arg\max_{C_r}\left[\sum_{W_r} P(C_r|W_r)P^{\frac{1}{L}}(X_r|W_r)P(W_r)\right]$$

Class-discriminant function에 사용

$$\cong \arg\max_{C_r}\left[\max_{W_r\in\aleph}\boxed{P(C_r|W_r)P^{\frac{1}{L}}(X_r|W_r)P(W_r)}\right]. \quad (4.7)$$

N-best list (denoted by $\aleph$)

Word error rate(WER) 발생

Class error rate (CER) 발생

label

utterance

Integrated approach (Yaman et al., 2008)

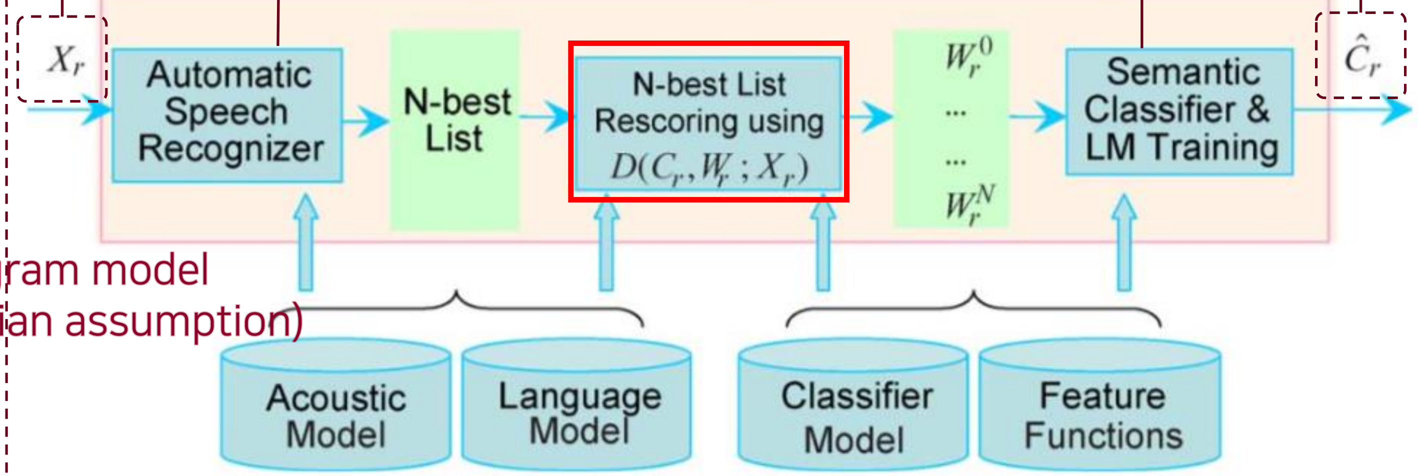


Fig. 2. ASR and semantic classification phases are integrated using the N-best lists.

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Conditional Understanding Models
Used for Discriminative Training of Language Models

Discriminative Learning Algorithm

Class-discriminant function (높을 수록 좋음)

$$D(C_r, W_r; X_r) = \log[P(C_r|W_r)P^{\frac{1}{L}}(X_r|W_r)P(W_r)]. \quad (4.8)$$

Decision rule

$$\hat{C}_r = \arg\max_{C_r}\left[\max_{W_r \in \aleph} D(C_r, W_r; X_r)\right]. \quad (4.9)$$

Minimum classification error(MCE) 기반으로 학습함

$$d_r(X_r) = -D(C_r^0, W_r^0; X_r) + \log\left[\frac{1}{T-1}\sum_{n=1}^{T}\exp[\eta D(C_r^n, W_r^n; X_r)]\right]^{\frac{1}{\eta}} \quad (4.12)$$

Misclassification function

$$\ell_r(d_r(X_r)) = \frac{1}{1+\exp(-\alpha d_r(X_r)+\beta)}. \quad (4.13)$$

Loss function

Word error rate(WER) 발생

Class error rate (CER) 발생

label

utterance

Integrated approach (Yaman et al., 2008)

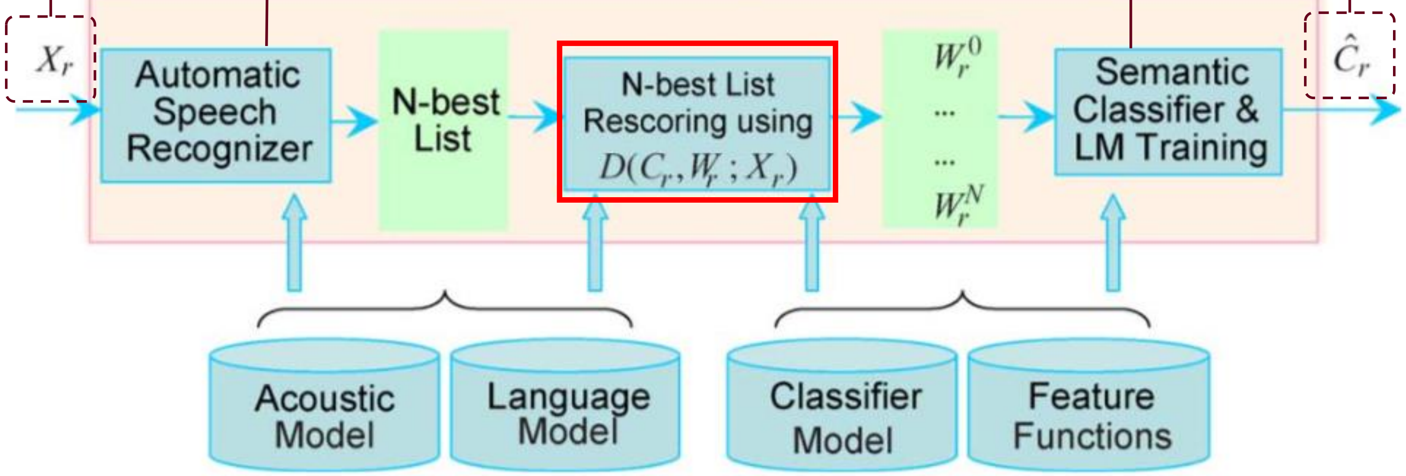


Fig. 2. ASR and semantic classification phases are integrated using the N-best lists.

두가지 목표
- ✓ Rescoring the sentence
- ✓ Adjusting LM and classifier parameters

4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Conditional Understanding Models Used for Discriminative Training of Language Models

Discriminative Learning Algorithm

Misclassification function

$$d_r(X_r) = -D(C_r^0, W_r^0; X_r) + \log\left[\frac{1}{T-1}\sum_{n=1}^{T}\exp[\eta D(C_r^n, W_r^n; X_r)]\right]^{\frac{1}{\eta}} \quad (4.12)$$

Loss function

$$\ell_r(d_r(X_r)) = \frac{1}{1+\exp(-\alpha d_r(X_r)+\beta)}. \quad (4.13)$$

Total loss function

$$L(\Lambda_W, \Lambda_\lambda) = \sum_r \ell_r(d_r(X_r))$$

LM update

$$\boxed{p_{w_x w_y}^{(t+1)}} = p_{w_x w_y}^{(t)} - \varepsilon_{LM} \sum_r \frac{\partial \ell_r(d_r(X_r))}{\partial p_{w_x w_y}}$$

$$= p_{w_x w_y}^{(t)} - \varepsilon_{LM}\alpha \sum_r \ell_r(d_r)[1-\ell_r(d_r)]\frac{\partial d_r(X_r)}{\partial p_{w_x w_y}}$$

classifiers update

$$\boxed{\lambda_k^{(t+1)}} = \lambda_k^{(t)} - \varepsilon_\lambda \sum_r \frac{\partial \ell_r(d_r(X_r))}{\lambda_k}$$

$$= \lambda_k^{(t)} - \varepsilon_\lambda \alpha \sum_r \ell_r(d_r(X_r))[1-\ell_r(d_r(X_r))]\frac{\partial d_r(X_r)}{\partial \lambda_k}$$

Word error rate(WER) 발생

Class error rate (CER) 발생

Integrated approach (Yaman et al., 2008)

utterance

label

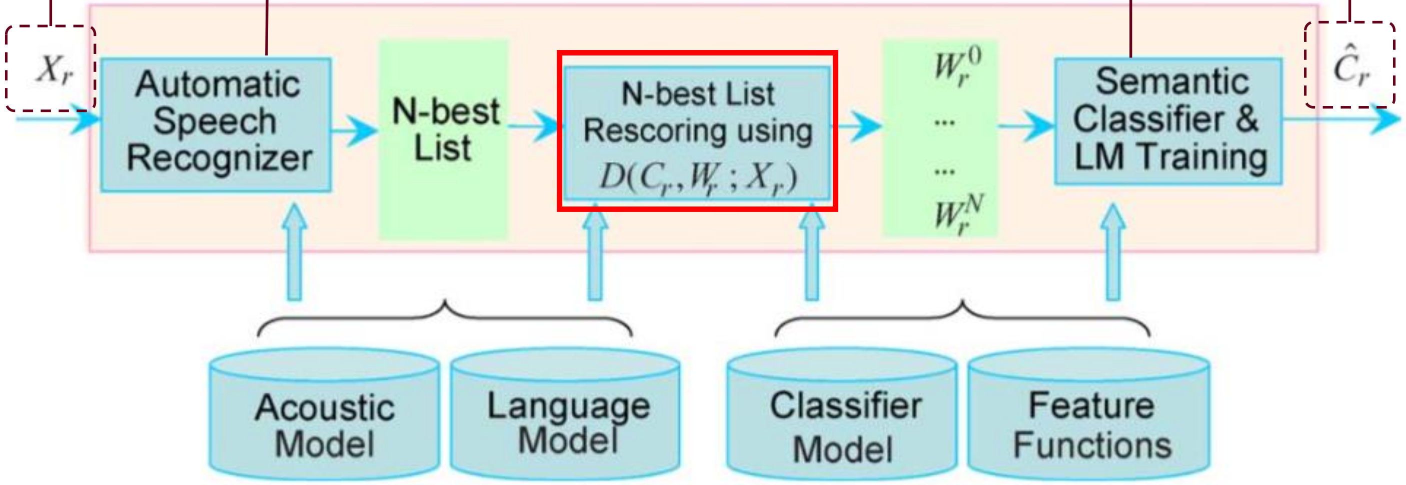


Fig. 2. ASR and semantic classification phases are integrated using the N-best lists.

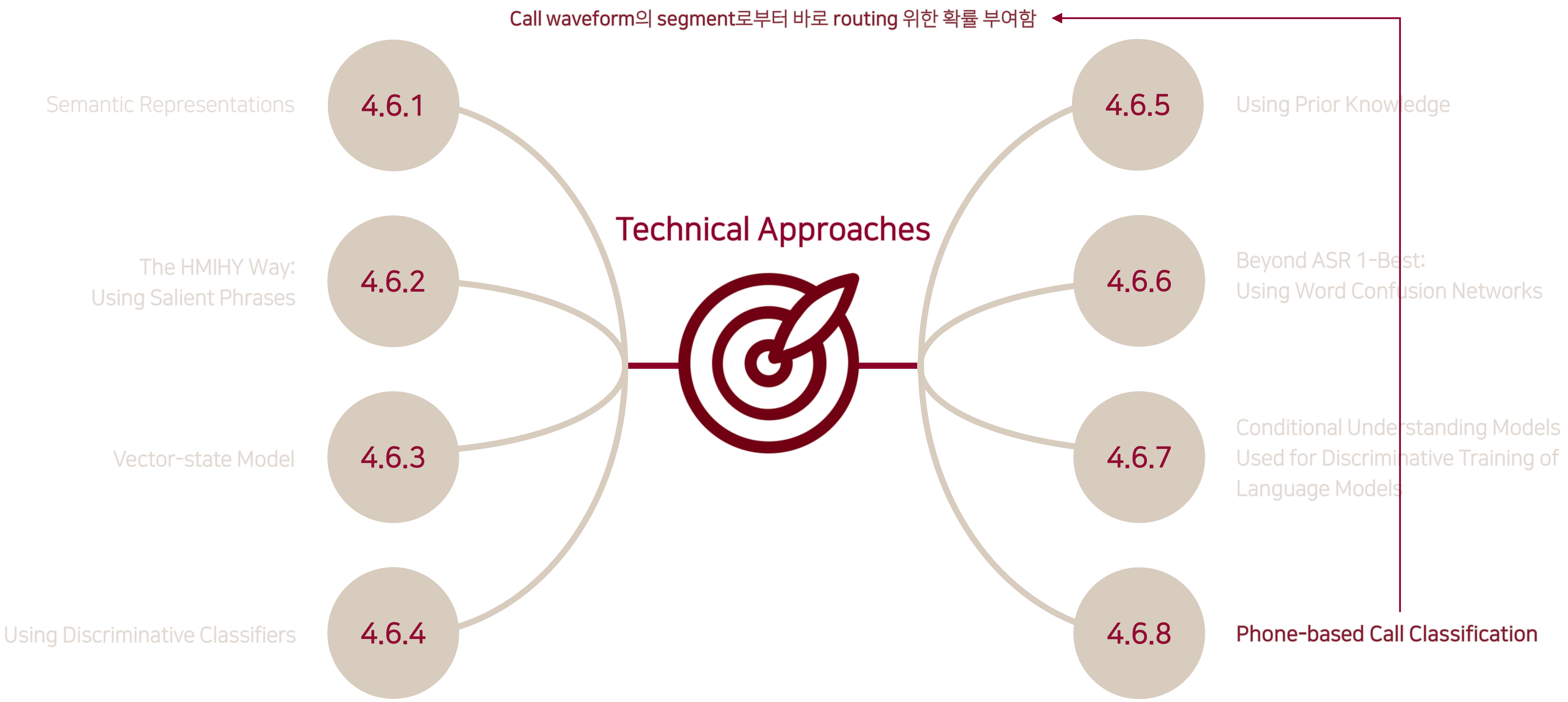
$p_{w_x w_y} = \log P(w_y|w_x)$ 초기 확률값은 MLE로 구함

the log-probabilities $p_{w_x w_y}^{(0)}$ estimated by maximum likelihood

λ_k denote the weight of the kth feature function $f_k(C_r, W_r)$

4.6 Technical Approaches

Call classification algorithm 에 대한 소개



4.6 Technical Approaches

Call classification algorithm 에 대한 소개

Phone-based Call Classification

- ✓ 대부분의 call classification algorithm은 word, word n-gram level feature 사용
- ✓ 그러나 Phone을 사용한 연구도 있음 (noise에 robust함)

Huang and Cox, 2003

- ✓ Call waveform의 segment에 바로 routing을 위한 확률 부여함
- ✓ Salient phrase대신 Salient acoustic morphemes을 사용하고 Linear discriminant analysis(LDA)을 적용해서 성능 개선함

Alshawi, 2003 (비슷한 아이디어를 제안)

- ✓ 차이점은 phone recognizer를 위한 self-training algorithm (unsupervised) 제안 한 것

4.7 Discussion and Conclusion

해결하지 못한 에러에 대한 분석 및 intent determination의 중요성

Discussion

✓ word *n*-gram-based statistical classification model을 많이 사용하는 편임
✓ 하지만 해결하지 못한 문제들이 아직 많이 있음 특히 "복잡한" 일들을 위한 "진짜 같은, 자연스러운" 발화는 아직 해결하지 못함
✓ ATIS corpus를 사용해서 intent determination의 detailed errors에 대해 발표함 (Tur et al, 2010)
✓ 간단한 Task에 대해서도 5%정도의 error가 있음 (word n-gram based Boosting 모델 사용)
✓ 5%의 errors를 6가지로 군집화 함
1) Preposition phrases embedded in noun phrases (전치사로 생기는 이슈, syntactic parser 필요)
2) Wrong functional arguments of utterances (syntactic parser 필요)
3) Annotation errors (human mistake)
4) Utterances with multiple sentences (intent in last sentence)
5) Other (ambiguous utterance, ill-formulated query, preprocessing)
6) Difficult cases (unseen data)

Error Type	All Train	25% Train	10-Fold
1	42.5%	33.8%	24.5%
2	22.5%	13.8%	30.0%
3	2.5%	6.1%	18.4%
4	0%	0%	8.0%
5	17.5%	12.5%	7.2%
6	15.0%	33.8%	11.7%

Figure 4.7 The distribution of error categories for intent determination using all and 25% data, and using all the training and the test set with 10-fold cross validation

Conclusion

✓ Intent determination task는 음성처리커뮤니티와 NLU 분야가 처음으로 만나게 된 곳
✓ Users' utterances에 대해 얕은 이해를 제공하지만 실제적인 application이 이미 많이 등장함
✓ 위와 같은 이유 때문에 Intent determination task는 중요하다고 할 수 있음

감사합니다

Any Questions?