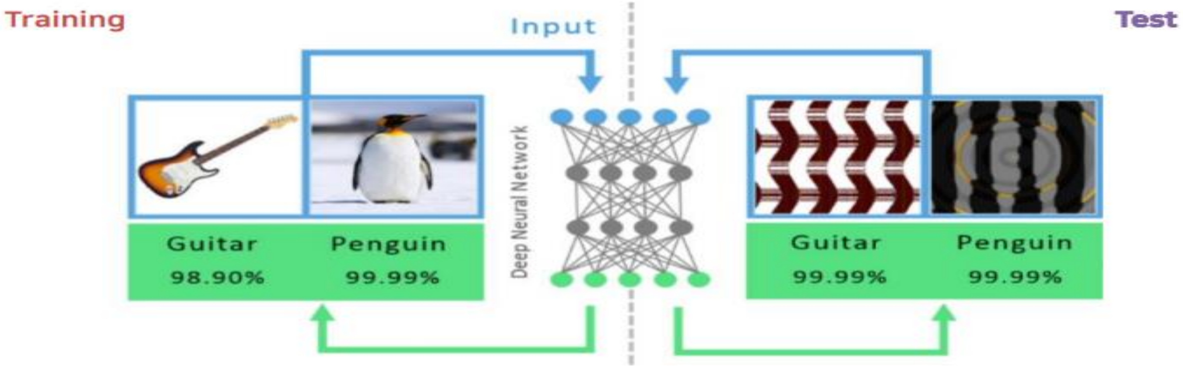
Towards Open Intent Detection

Hanlei Zhang, Hua Xu, Shaojie Zhao, Qianrui Zhou

전북대학교 IT정보공학과
202018392
박나현

Open intent detection

- 대화시스템에서 아직 난제
  - 기존 심층 신경망에서는 학습한 class에 대한 확률만 출력(closed world 가정)
  - Training 동안 user의 모든 intent를 고려하는 것은 불가능



k개의 class로 분류

User utterances	Intent Label
Book a flight from LA to Madrid.	Booking flight
Can you get me a table at Steve's?	Restaurant reservation
Book Delta ticket Madison to Atlanta.	Booking flight
Schedule me a table at Red Lobster.	Restaurant reservation
……	……
What time is it now?	Asking time (**Open**)
Where is the nearest school?	Asking place (**Open**)

k+1번째 새로운 class = open intent

Open intent detection

- **Goal**
  - Open intent 를 찾는 동시에 known intent 를 정확하게 분류하는 것
  - K-class known intent (prior knowledge)만을 이용하여 (k+1)번째 class인 open intent를 식별
- **Two main difficulties**
  - Known intent로 학습한 representation이 unseen open class를 찾기에 rubust하지 않다.
  - Representation등 결정조건들이 여전히 불완전하다

Problem formulation

- Intent label set

$$I = \{I^{\text{Known}}, \text{Open}\}$$

- Known intent label set

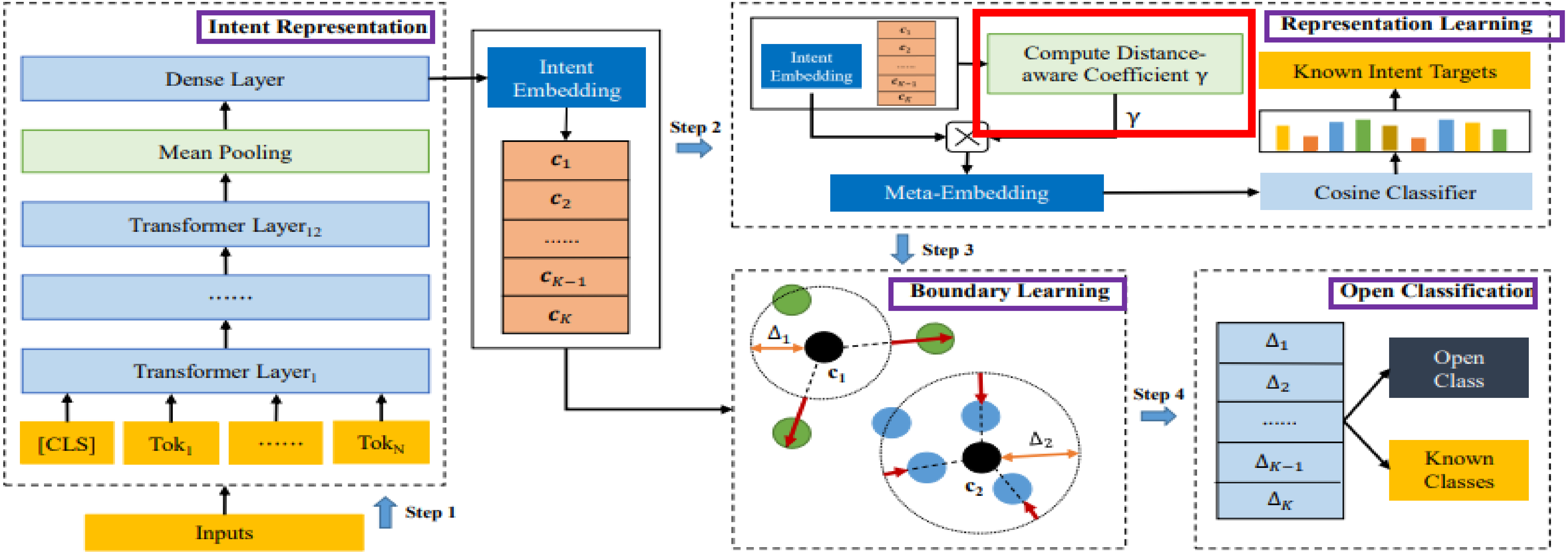
$$I^{\text{Known}} = \{I_1, \ldots, I_K\}$$

- Data set

$$D = \{\underbrace{D^{\text{Train}}, D^{\text{Valid}}}_{I^{\text{Known}}}, \underbrace{D^{\text{Test}}}_{I}\}$$

Model architecture

- Four main step



1. Intent Representation

Hidden layer size

- Pre-trained BERT의 마지막 hidden layer에서 intent representation 추출
- **Mean pooling** $[CLS, T_1, \cdots, T_N] \in \mathbb{R}^{(N+1)\times H}$
  - to synthesize the high-level semantic features in one sentence
  - Get the averaged representation $\boldsymbol{x}_i \in \mathbb{R}^H$

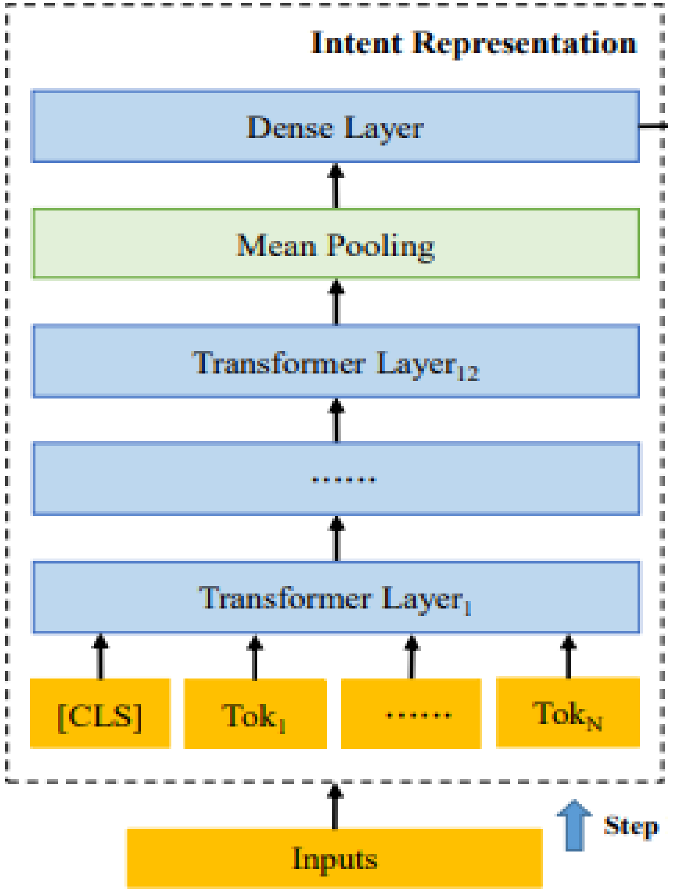
$$\boldsymbol{x}_i = \text{mean-pooling}([CLS, Tok_1, \cdots, Tok_M]),$$

- **Dense layer**
  - To further strengthen feature extraction capability
  - Get the intent representation $\boldsymbol{z}_i \in \mathbb{R}^D$ Dimension of the intent representation

$$\boldsymbol{z}_i = h(\boldsymbol{x}_i) = \sigma(W_h \boldsymbol{x}_i + b_h)$$

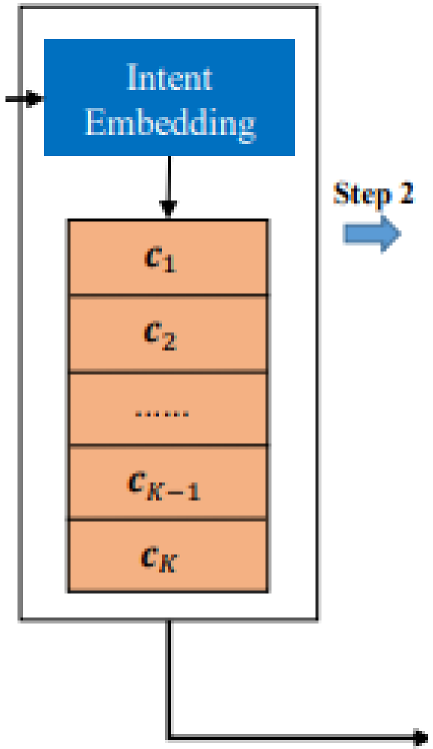
ReLU

$W_h \in \mathbb{R}^{H\times D}$ and $b_h \in \mathbb{R}^D$



2. Distance-aware Representation Learning

- **차별적인 intent feature 학습**
  - 중심 계산
  - Meta embedding을 얻기 위해 Distance-aware 계수를 계산
  - 코사인 분류기를 사용하여 각 샘플이 거리 정보를 인식 가능



- **Centroids Calculation**
  - 평균 벡터

$S = \{(z_i, y_i), \ldots, (z_N, y_N)\}$

$$\boldsymbol{c}_k = \frac{1}{|S_k|} \sum_{(\boldsymbol{z}_i, y_i) \in S_k} \boldsymbol{z}_i,$$

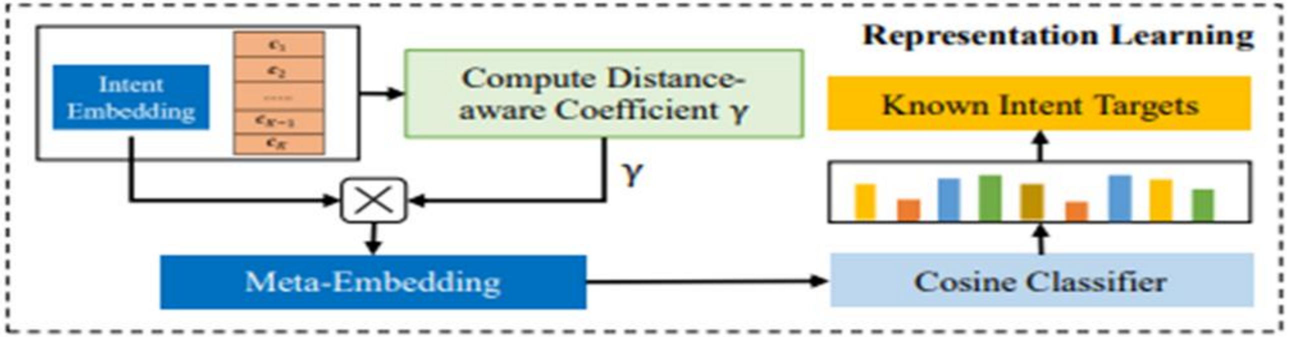
K class의 feature vector 집합

2. Distance-aware Representation Learning

- **Meta embedding with distance-aware concept**
  - **Initial intent representation의 문제**
    - “easy”인지 “hard”인지 확인하기 어려워 차별적인 representation 학습에 불리
  - **해결방법: Distance-aware 개념 활용**
    - 구별능력 강화위해 meta-embedding 얻음
    - 가장 정보력 있는 중심 index(k_a,k_b)로 구별 능력 평가하기 위해 계산

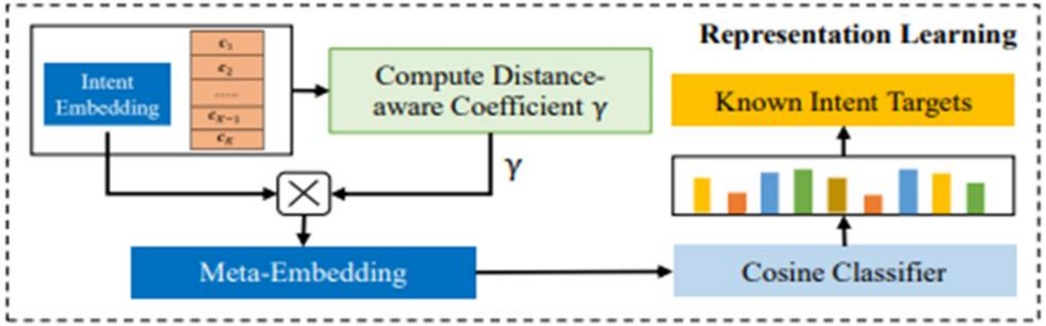
$$k_a = \underset{k}{\operatorname{argmin}} \{\|\boldsymbol{z}_i - \boldsymbol{c}_k\|_2\}_{k \in I^{\text{Known}}},$$

$$k_b = \underset{k}{\operatorname{argmin}} \{\|\boldsymbol{z}_i - \boldsymbol{c}_k\|_2\}_{k \in I^{\text{Known}} \setminus \{k_a\}},$$



2. Distance-aware Representation Learning

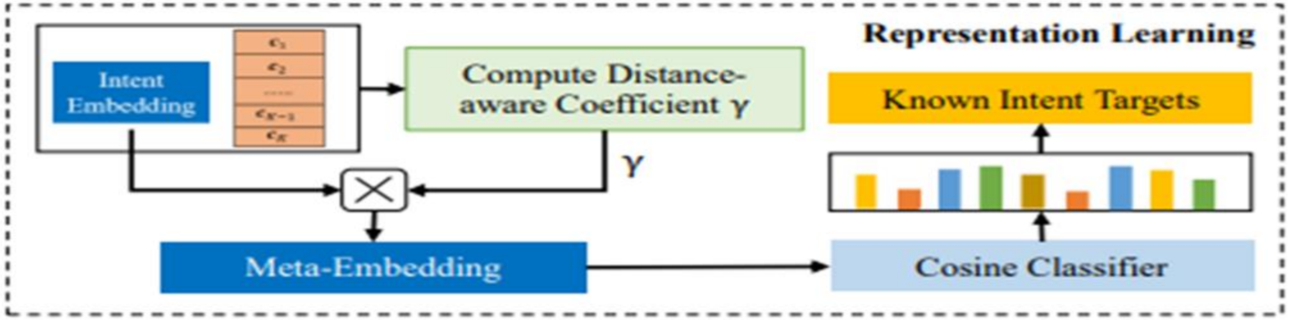
- **Meta embedding with distance-aware concept**
  - 구별 쉽게 하기 위한 조건
    - Discriminative 예시는 가장 가까운 중심과 가까워야하고 그 다음 가까운 중심으로부터 멀어야 한다.
  - **두 길이의 차이는 분리 능력 반영(intent의 차별화)**
  - **계수의 효과** $\gamma_i = \exp(\|\boldsymbol{z}_i - \boldsymbol{c}_{k_b}\|_2 - \|\boldsymbol{z}_i - \boldsymbol{c}_{k_a}\|_2)$ s.t. $\gamma_i \geq 1,$
    - 차별 효과를 강화
    - 예시의 난의도를 시사
      - Easy: 구분 명확, 계수 큰 값
      - Hard: 구분 모호, 계수 작은 값



  - 거리정보 사용하기 위해 intent representation에 계수 곱함

$$\boldsymbol{z}_i^{meta} = \gamma_i \cdot \boldsymbol{z}_i.$$

2. Distance-aware Representation Learning



- Representation learning
  - **Cosine similarity classifier**
    - meta-embedding에 포함된 거리 정보를 포착하기 위해 사용 $\phi(\cdot)$: cosine classifier
    - "hard" 예시에 더 집중할 수 있도록 도움 $\phi(\boldsymbol{z}_i^{meta})^k = \alpha \cdot \cos\left(\boldsymbol{z}_i^{meta}, w_k^\star\right) = \alpha \cdot \overline{\boldsymbol{z}_i^{meta}}^\top \overline{w_k^\star},$
    - **Non-linear squashing function**
      - 벡터 크기 반영하는데 도움
      - 메타 임베딩의 크기가 크면 1보다 작은 길이로 수축시키고 작으면 0으로 수축

$$\overline{\boldsymbol{z}_i^{meta}} = \frac{\|\boldsymbol{z}_i^{meta}\|^2}{1+\|\boldsymbol{z}_i^{meta}\|^2} \frac{\boldsymbol{z}_i^{meta}}{\|\boldsymbol{z}_i^{meta}\|},$$

$$\overline{w_k^\star} = \frac{w_k^\star}{\|w_k^\star\|},$$

  - Softmax loss 사용
    - 초기 intent representation은 보정되고, 결정경계학습에 이용

$$\mathcal{L}_s = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(\phi(\boldsymbol{z}_i^{meta})^{y_i})}{\sum_{j=1}^{K} \exp(\phi(\boldsymbol{z}_i^{meta})^{j})},$$

3. Adaptive decision boundary learning

	Empirical risk	Open space risk	결과
원 밖에 known intent	↑	↓	Boundary 늘리기
원 안에 Known intent	↓	↑	Open intent 포함

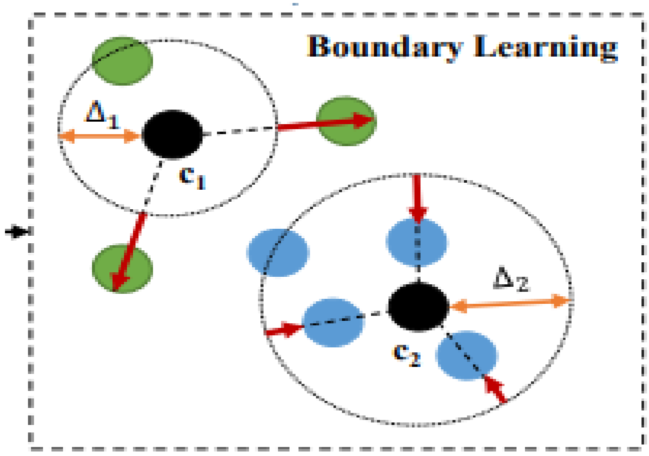
- **Decision boundary formulation**
  - Softplus activation function
    - 반지름과 학습 파라미터를 mapping

$$\Delta_k = \log\left(1 + \exp(\widehat{\Delta_k})\right).$$

- **Boundary learning**
  - Empirical risk와 open space risk 사이의 balance 중요

$$\mathcal{L}_b = \frac{1}{N}\sum_{i=1}^{N}\left[\delta_i\left(\|\boldsymbol{z}_i - \boldsymbol{c}_{y_i}\|_2 - \Delta_{y_i}\right) + \left(1-\delta_i\right)\left(\Delta_{y_i} - \|\boldsymbol{z}_i - \boldsymbol{c}_{y_i}\|_2\right)\right],$$

$$\delta_i := \begin{cases} 1, \text{ if } & \|\boldsymbol{z}_i - \boldsymbol{c}_{y_i}\|_2 > \Delta_{y_i}, \\ 0, \text{ if } & \|\boldsymbol{z}_i - \boldsymbol{c}_{y_i}\|_2 \le \Delta_{y_i}. \end{cases}$$

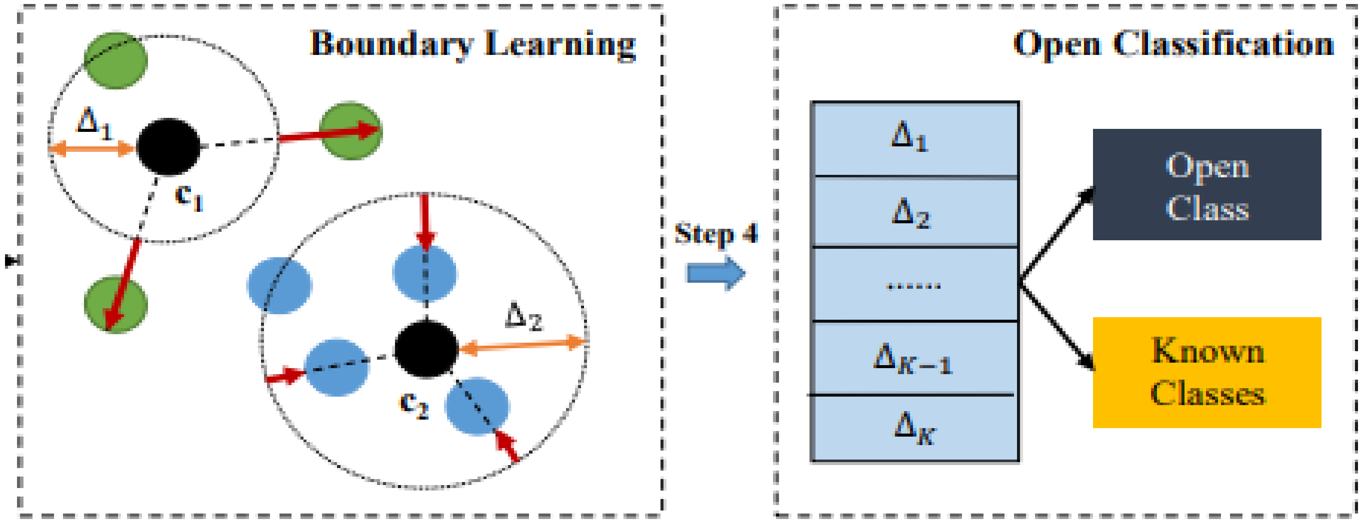


4. Open classification with decision boundary

- Boundary 밖에 위치하면 open, 안에 위치하면 known

$$k_a = \underset{k}{\operatorname{argmin}} \{d(\boldsymbol{z}_i, \boldsymbol{c}_k)\}_{k \in I^{\text{Known}}},$$

$$\hat{y} = \begin{cases} \text{Open, if } d(\boldsymbol{z}_i, \boldsymbol{c}_k) > \Delta_{k_a}; \\ k_a, \text{ if } d(\boldsymbol{z}_i, \boldsymbol{c}_k) \leq \Delta_{k_a}, \end{cases}$$



Experiments

- Datasets
  - **BANKING**
    - 은행 도메인
  - **OOS**
    - Open data가 포함된 dataset
  - **StackOverflow**
    - Technical question 포함

Dataset	Classes	#Training	#Validation	#Test	Vocabulary Size	Length (max / mean)
BANKING	77	9,003	1,000	3,080	5,028	79 / 11.91
OOS	150	15,000	3,000	5,700	8,376	28 / 8.31
StackOverflow	20	12,000	2,000	6,000	17,182	41 / 9.18

Table 1: Statistics of BANKING, OOS and StackOverflow datasets. # indicates the total number of sentences.

Results

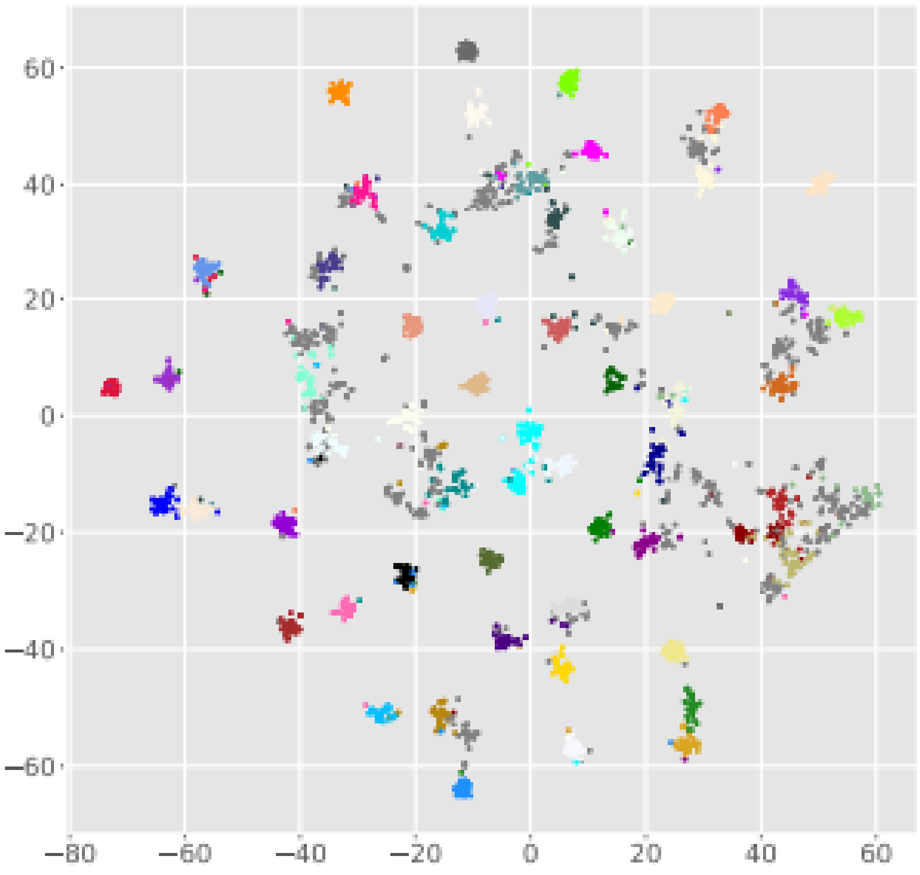
- Student's t-test 실시

p-value < 0.05 (†) and p-value < 0.1 (*)

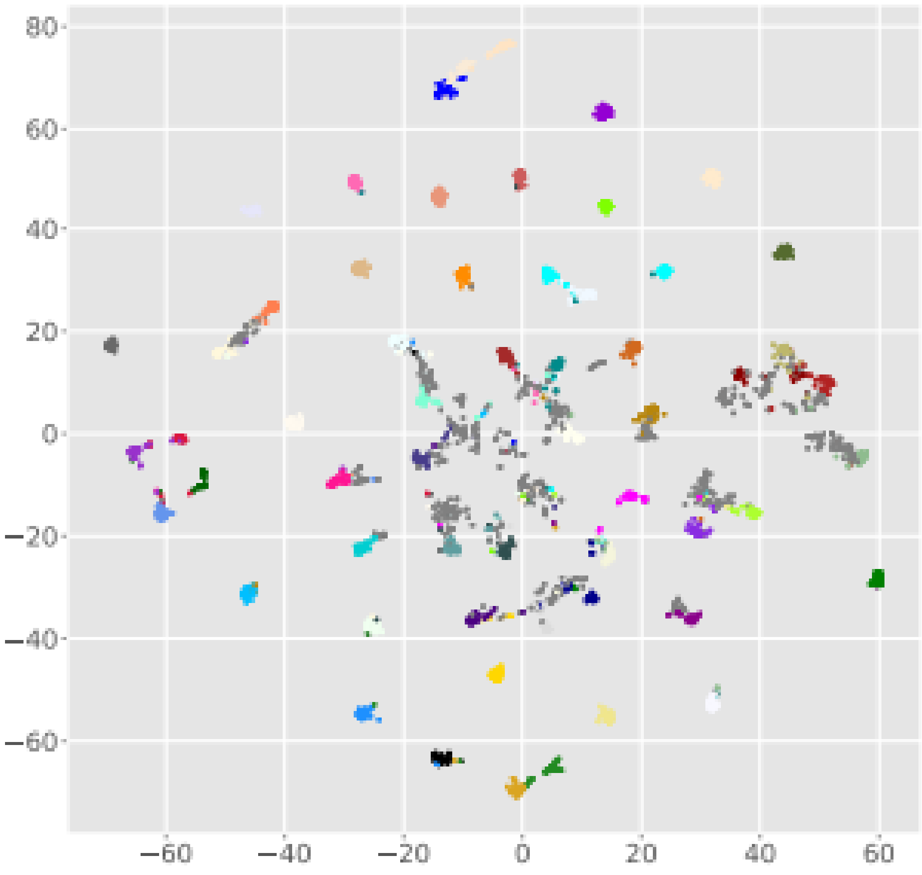
Datasets	Methods	25%		50%		75%		Mean	
		ACC	F1	ACC	F1	ACC	F1	ACC	F1
BANKING	MSP	41.84†	50.03†	59.80†	71.40†	75.90†	83.49†	59.18†	68.31†
	SEG	49.73†	52.03†	54.66†	62.86†	64.54†	69.37†	56.31†	61.42†
	OpenMax	47.76†	53.18†	65.53†	74.64†	78.32†	84.95†	63.87†	70.92†
	LOF	66.73†	63.38†	71.13†	76.26†	77.21†	83.64†	71.69†	74.43†
	DOC	70.31†	65.74†	74.60†	78.24†	78.94†	83.79†	74.62†	75.92†
	DeepUNK	70.68†	65.57†	71.01†	75.41†	74.73†	81.12†	72.14†	74.03†
	$(K+1)$-way	75.43†	68.31†	74.66†	78.13†	79.90†	85.22†	76.66†	77.22†
	ADB	79.94	72.08	79.52†	81.33†	**81.35**	**86.08**	80.27	79.83
	DA-ADB	**81.09**	**73.65**	**81.64**	**82.60**	81.18	85.68	**81.30**	**80.64**
OOS	MSP	49.78†	49.42†	62.71†	70.33†	72.86†	81.61†	61.78†	67.12†
	SEG	53.34†	47.57†	60.54†	62.51†	42.97†	42.49†	52.28†	50.86†
	OpenMax	70.27†	63.03†	80.22†	79.86†	75.36†	71.17†	75.28†	71.35†
	LOF	87.77	78.13	85.22†	83.86†	85.07†	87.20†	86.02†	83.06†
	DOC	86.08†	75.86†	85.19†	83.89†	85.93†	87.87†	85.73†	82.54†
	DeepUNK	87.18*	77.32*	84.95†	83.35†	84.61†	86.53†	85.58†	82.40†
	$(K+1)$-way	86.98†	76.58†	83.71†	82.85†	85.31†	87.90†	85.33†	82.44†
	ADB	88.21	78.14*	86.47†	85.11	86.98*	**88.95**	87.22†	84.07
	DA-ADB	**89.49**	**79.95**	**87.96**	**85.64**	**87.46**	88.47†	**88.30**	**84.69**
StackOverflow	MSP	27.91†	37.49†	53.23†	62.70†	73.20†	78.70†	51.45†	59.63†
	SEG	23.35†	34.59†	43.04†	55.10†	62.63†	69.86†	43.01†	53.18†
	OpenMax	38.97†	45.35†	60.27†	67.72†	75.78†	80.90†	58.34†	64.66†
	LOF	25.02†	35.29†	44.56†	56.57†	65.05†	71.87†	44.88†	54.58†
	DOC	57.75†	57.34†	73.88†	76.80†	80.55†	84.37†	70.73†	72.84†
	DeepUNK	40.03†	45.64†	55.46†	64.78†	71.56†	77.63†	55.68†	62.68†
	$(K+1)$-way	53.05†	53.12†	63.54†	69.26†	74.72†	79.47†	63.77†	67.28†
	ADB	86.70*	79.79†	86.51†	85.55†	82.84*	86.07†	85.35†	83.80†
	DA-ADB	**89.03**	**82.81**	**87.79**	**86.92**	**83.63**	**86.89**	**86.82**	**85.54**

Datasets	Methods	25%		50%		75%		Mean	
		Open	Known	Open	Known	Open	Known	Open	Known
BANKING	MSP	38.84†	50.62†	42.13†	72.17†	41.64†	84.21†	40.87†	69.00†
	SEG	52.97†	51.98†	42.35†	63.40†	37.58†	69.92†	44.30†	61.77†
	OpenMax	48.52†	53.42†	55.03†	75.16†	53.02†	85.50†	52.19†	71.36†
	LOF	72.64†	62.89†	66.81†	76.51†	54.19†	84.15†	64.55†	74.52†
	DOC	76.64†	65.16†	72.66†	78.38†	63.51†	84.14†	70.94†	75.89†
	DeepUNK	76.98†	64.97†	67.80†	75.61†	50.57†	81.65†	65.12†	74.08†
	$(K+1)$-way	81.52†	67.61†	72.38†	78.29†	62.13†	85.62	72.01†	77.17†
	ADB	85.57	71.37	79.32†	81.38†	67.32*	**86.40**	77.40†	79.72
	DA-ADB	**86.49**	**72.97**	**82.10**	**82.61**	**69.51**	85.69*	**79.37**	**80.51**
OOS	MSP	54.74†	49.28†	57.49†	70.50†	56.26†	81.83†	56.16†	67.20†
	SEG	60.59†	47.23†	61.13†	62.52†	41.60†	42.50†	54.44†	50.75†
	OpenMax	77.51†	62.65†	82.15†	79.83†	75.18†	71.14†	78.28†	71.21†
	LOF	91.96	77.77	87.57†	83.81†	82.81†	87.24†	87.45†	82.94†
	DOC	90.78†	75.46†	87.45†	83.84†	83.87†	87.91†	87.37†	82.40†
	DeepUNK	91.61†	76.95†	87.48†	83.30†	82.67†	86.57†	87.25†	82.27†
	$(K+1)$-way	91.44†	76.19†	85.84†	82.82†	82.39†	87.95*	86.56†	82.32†
	ADB	92.30	77.77*	88.54†	85.06	84.81†	**88.99**	88.55†	83.94
	DA-ADB	**93.20**	**79.60**	**90.14**	**85.58**	**86.09**	88.49†	**89.81**	**84.56**
StackOverflow	MSP	11.66†	42.66†	26.94†	66.28†	37.86†	81.42†	25.49†	63.45†
	SEG	4.36†	40.63†	4.72†	60.14†	6.38†	74.09†	5.15†	58.29†
	OpenMax	34.52†	47.51†	46.11†	69.88†	49.69†	82.98†	43.44†	66.79†
	LOF	7.14†	40.92†	5.18†	61.71†	5.22†	76.31†	5.85†	59.65†
	DOC	62.50†	56.30†	71.18†	77.37†	65.32†	85.64†	66.33†	73.10†
	DeepUNK	36.87†	47.39†	35.80†	67.67†	34.38†	77.63†	35.68†	65.19†
	$(K+1)$-way	56.31†	52.48†	53.68†	70.81†	47.57†	81.60†	52.52†	68.30†
	ADB	90.91*	77.56†	87.72†	85.33†	74.02	86.87†	84.22†	83.25†
	DA-ADB	**92.61**	**80.84**	**88.86**	**86.72**	**74.66**	**87.71**	**85.38**	**85.09**

Discussion



(a) Vanilla Intent Representations



(b) Learned with Distance-aware Concepts

Conclusions

- DA-ADB 제안
  - distance-aware 개념으로 차별적인 feature학습
  - Empirical risk와 open space risk사이 균형을 맞추며 decision boundary 학습

Thank you