Effective Open Intent Classification with K-Center Contrastive Learning and Adjustable Decision Boundary

Xiaokang Liu, Jianquan Li, Jingjing Mu , Min Yang, Ruifeng Xu , Benyou Wang

전북대학교 IT정보공학과
202018392
박나현

AAAI 2023

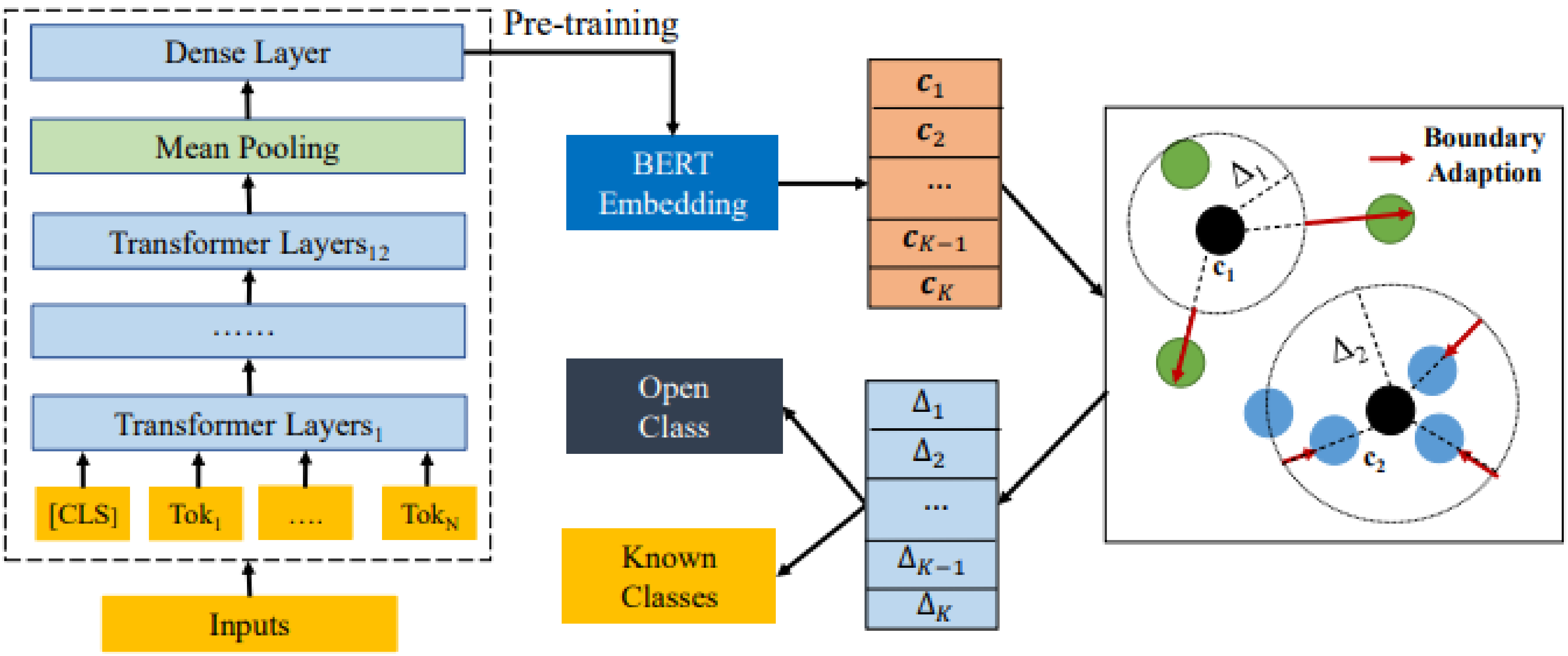
Open intent classification

OOD: out-of-domain
IND: In-domain

- Intent classification
  - 대화시스템에서 중요한 task
  - Closed world 가정하여 미리 정의된 known class 에 대해서만 출력
- Open intent classification
  - OOD를 IND라고 판별하는 에러를 줄이기 위해 노력
  - 최근 방법
    - Known intent 의 decision boundary 를 학습하여 open intent 를 찾도록 제안
    - Ex) ADB

ADB

- IND 만을 이용하여 학습 후 OOD 분류

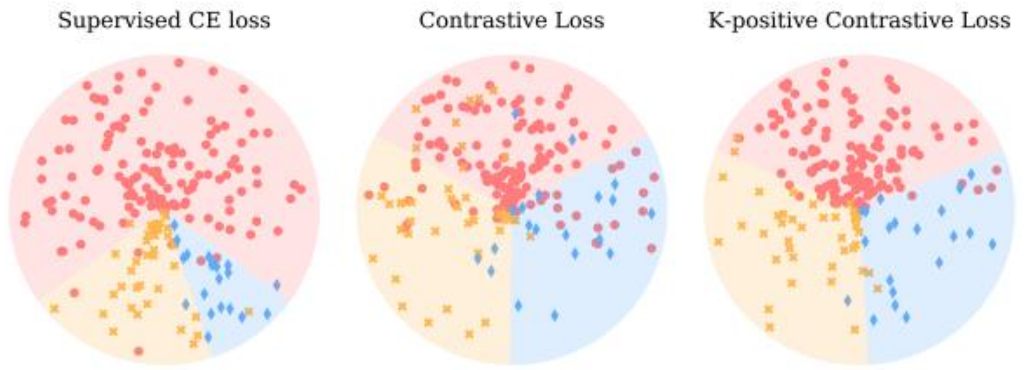


Challenge

- Representation learning 에서
  Training label 에 의존
  - Noise가 있고 불균형한 training data로 인한
    OOD에 대한 모델 성능 감소
- ADB 에서 Decision boundary 방식 보안
  - Known intent 만을 사용하기 때문에
    OOD 에 대한 성능 저하 가능성 존재

- K-center contrastive learning(KCCL) 제안
  - K positive sample 끼리 비교하여
    서로 더 가깝도록 학습
- Expanding and Shrinking (ADBES) 제안
  - OOD 가 경계로부터 멀리 있다면 경계 늘리고
    반대의 상황에서는 작게 만들어 학습

Contrastive learning



- Positive pair 거리 가깝도록 negative pair 거리 멀도록 학습
- 기존 supervised contrastive learning
  - Representation learning 에서 instance-rich class 의 지배 문제 발생
    - 동일한 class의 모든 sample 을 사용하여 positive pair 를 구성

Positive sample

$$\mathcal{L}_{\mathrm{CL}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\mathbf{z}_i\cdot\mathbf{z}_i^{+}/\tau}}{e^{\mathbf{z}_i\cdot\mathbf{z}_i^{+}/\tau}+\sum_{\mathbf{z}_i^{-}\in Z_i^{-}}e^{\mathbf{z}_i\cdot\mathbf{z}_i^{-}/\tau}}$$

Temperature hyperparameter:
Hard negative sample의 penalty 강도

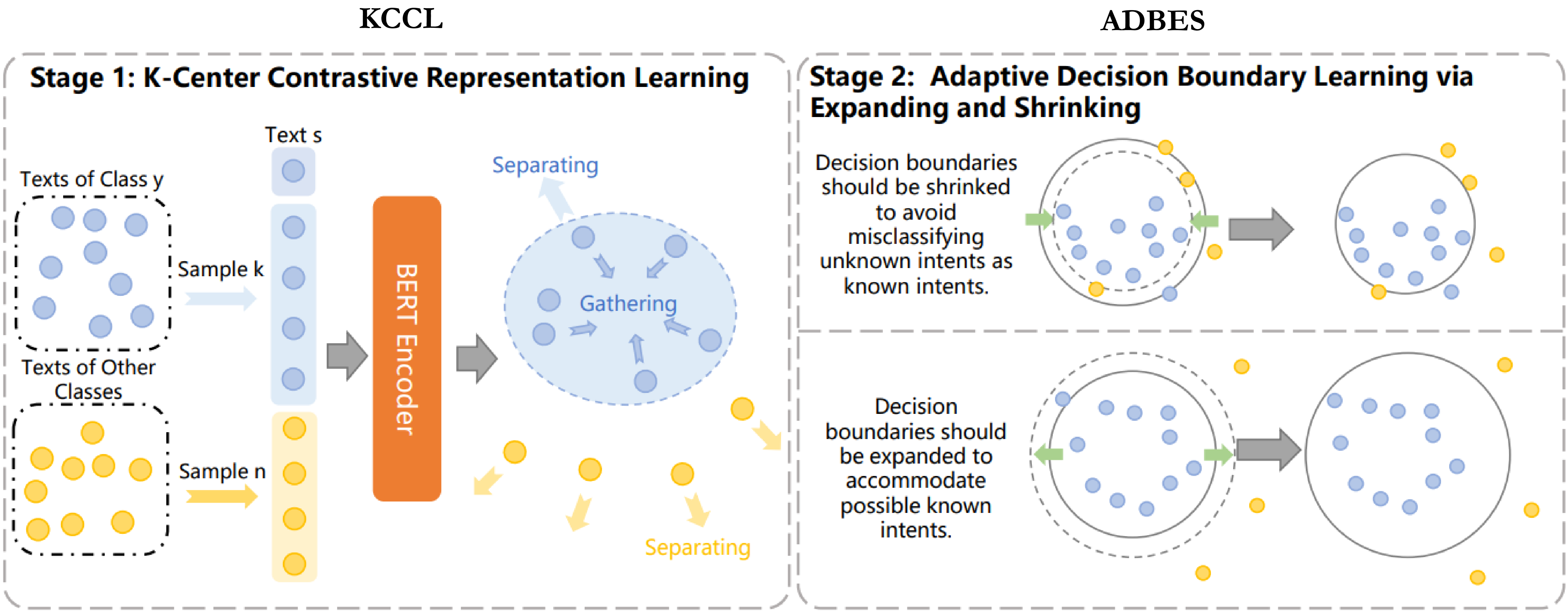
negative sample

$$\mathcal{L}_{\mathrm{KCL}} = \frac{1}{N(k+1)}\sum_{i=1}^{N}\sum_{v_j^{+}\in\{\tilde{v}_i\}\cup V_{i,k}^{+}} -\log\frac{\exp(v_i\cdot v_j^{+}/\tau)}{\exp(v_i\cdot\tilde{v}_i/\tau)+\sum_{v_j\in V_i}\exp(v_i\cdot v_j/\tau)},$$

augmented sample

Model Overview

K-center contrastive learning and adjustable decision boundary learning(CLAB)



Initial Representation Learning

- Feature representation 학습 위해 Pre-train BERT 사용
  - Hidden representation of tokens $[CLS, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, SEP] \in \mathbb{R}^{(n+2)\times H}$
- **Mean pooling**

$$\mathbf{o}_i = \textit{mean-pooling}([CLS, \mathbf{v}_1, \ldots, \mathbf{v}_n, SEP])$$

- **Enhance the feature representation learning**

$$\mathbf{h}_i = \mathrm{ReLU}(W_1\mathbf{o}_i + \mathbf{b}_1), \quad \mathbf{z}_i = \frac{\mathbf{h}_i}{\|\mathbf{h}_i\|_2}$$

Normalize 진행

K-Center Contrastive Learning (KCCL)

- K 개 positive set 내에서 각 instance pair 끼리 비교
  - 기존 KCL: k positive 와 negative 비교

$\mathbf{z}_{i,m}^{+}$ $\mathbf{z}_{i,n}^{+}$: Z_i^{+}에 있는 positive sample

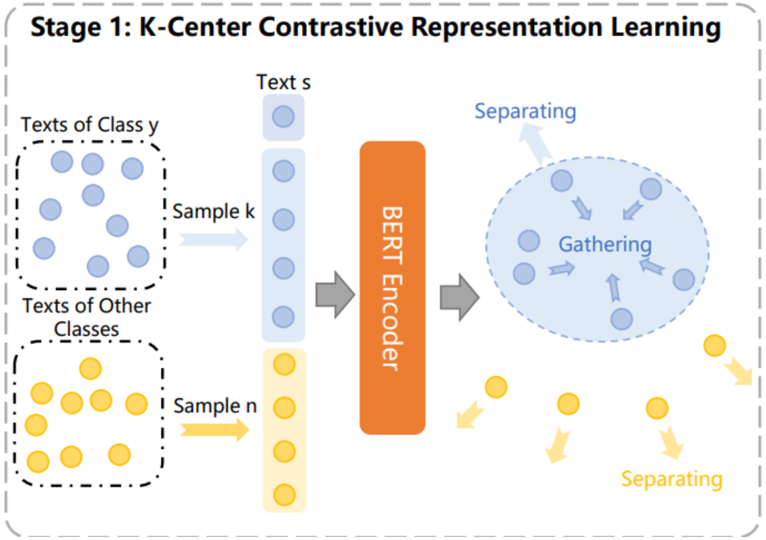
$$\mathcal{L}_{\mathrm{KCCL}} = -\frac{1}{N \cdot K \cdot (K+1)} \sum_{i=1}^{N} \sum_{m=1}^{K+1} \sum_{n=1, n\neq m}^{K+1} \log \frac{e^{\mathbf{z}_{i,m}^{+} \cdot \mathbf{z}_{i,n}^{+}/\tau}}{e^{\mathbf{z}_{i,m}^{+} \cdot \mathbf{z}_{i,n}^{+}/\tau} + \sum_{\mathbf{z}_i^{-} \in \mathbf{Z}_i^{-}} \left(e^{\mathbf{z}_{i,m}^{+} \cdot \mathbf{z}_i^{-}/\tau} + e^{\mathbf{z}_{i,n}^{+} \cdot \mathbf{z}_i^{-}/\tau}\right)}$$

- Cross-entropy

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \log \mathrm{softmax}(W_2 \mathbf{z}_i + \mathbf{b}_2)$$

- Final loss function

$$\mathcal{L}_{S_1} = \lambda \cdot \mathcal{L}_{\mathrm{KCCL}} + (1-\lambda) \cdot \mathcal{L}_{\mathrm{CE}}$$



Adaptive Decision Boundary via Expanding and Shrinking

- **Decision Boundary Learning**
  - Decision center
    - Representation 의 평균

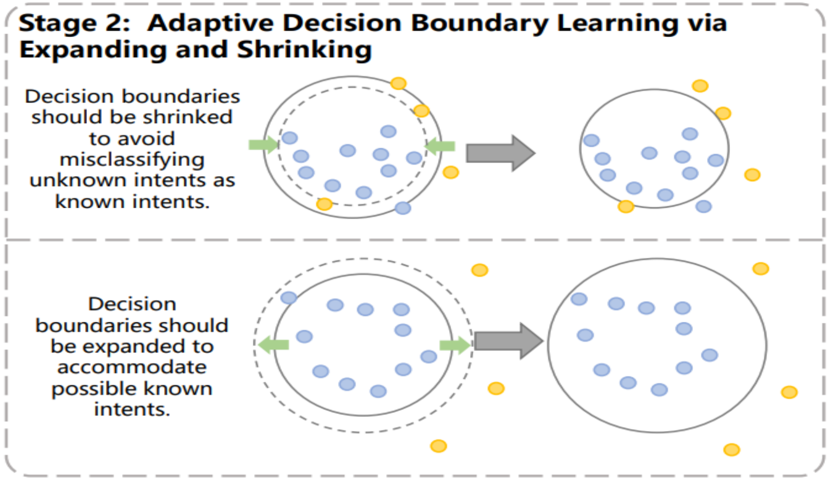
$$\mathbf{c}_k = \frac{1}{|\mathcal{S}_k|} \sum_{\mathbf{z}_i \in \mathcal{S}_k} \mathbf{z}_i$$

Instance 수

  - Known instance 의 제약

$$\|\mathbf{z}_i - \mathbf{c}_k\|_2 \leq \Delta_k$$

  - 기존 ADB Radius
    - Empirical risk 를 줄어들도록 known intent 포함할 만큼 boundary가 넓고, boundary를 축소하여 open space risk 를 줄이는 balance 필요

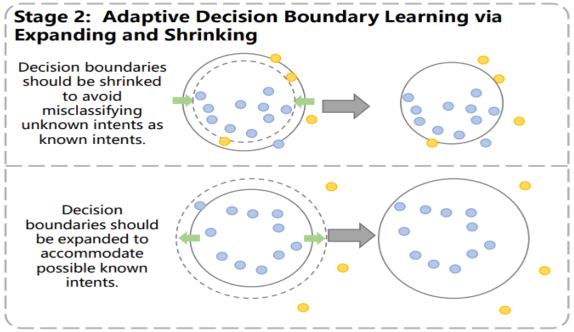


$$\mathcal{L}_{\mathrm{ADB}} = \frac{1}{N} \sum_{i=1}^{N} \gamma_i \cdot (\|\mathbf{z}_i - \mathbf{c}_{y_i}\|_2 - \Delta_{y_i}) + (1-\gamma_i) \cdot (\Delta_{y_i} - \|\mathbf{z}_i - \mathbf{c}_{y_i}\|_2)$$

$$\gamma_i = \begin{cases} 1, & \text{if } \|\mathbf{z}_i - \mathbf{c}_{y_i}\|_2 > \Delta_{y_i}, \\ 0, & \text{if } \|\mathbf{z}_i - \mathbf{c}_{y_i}\|_2 \leq \Delta_{y_i}. \end{cases}$$

Adaptive Decision Boundary via Expanding and Shrinking

- **Expanding and Shrinking**
  - OOD가 경계에서 멀리 있다면 IND 를 더 많이 포함하도록 반지름을 늘림
  - OOD가 경계로 부터 가까이 있다면 반지름을 줄임
  - 수학적
    - OOD 에서 중심 사이의 거리가 skewed normal distribution 따름
    - 양끝 분포에 따라 반지름 조절



$$\alpha_i = \begin{cases} 1, & \left\|\mathbf{z}_i^- - \mathbf{c}_{y_i}\right\|_2 > \Delta_{y_i} + e, \\ 0, & \left\|\mathbf{z}_i^- - \mathbf{c}_{y_i}\right\|_2 \le \Delta_{y_i} + e. \end{cases}$$

$$\beta_i = \begin{cases} 1, & \left\|\mathbf{z}_i^- - \mathbf{c}_{y_i}\right\|_2 < \Delta_{y_i} + s, \\ 0, & \left\|\mathbf{z}_i^- - \mathbf{c}_{y_i}\right\|_2 \ge \Delta_{y_i} + s. \end{cases}$$

$$\mathcal{L}_{\text{ADBES}} = \mathcal{L}_{\text{ADB}} + \frac{1}{N}\sum_{i=1}^{N}\left\{\eta \cdot \alpha_i \cdot \left[\left\|\mathbf{z}_i^- - \mathbf{c}_{y_i}\right\|_2 - \left(\Delta_{y_i} + e\right)\right] + \eta \cdot \beta_i \cdot \left[\left(\Delta_{y_i} + s\right) - \left\|\mathbf{z}_i^- - \mathbf{c}_{y_i}\right\|_2\right]\right\}$$

Expansion parameter

Shrink parameter

Open Intent Classification during Inference

$$\hat{y}_j = \begin{cases} \text{Unknown, if } d(\mathbf{z}_j, \mathbf{c}_k) > \Delta_k, k \in \mathcal{Y}; \\ \text{arg min}_{k \in \mathcal{Y}} d(\mathbf{z}_j, \mathbf{c}_k), \text{ otherwise.} \end{cases}$$

$$\mathcal{Y} = \{1, \ldots, K\}$$

Experiments

Dataset	Class	Train/Valid/Test	Length (max/mean)
BANKING	77	9003 / 1000 / 3080	79 / 11.91
OOS	150	15000 / 3000 / 5700	28 / 8.31
StackOverflow	20	12000 / 2000 / 6000	41 / 9.18

Table 1: Statistics of experimental datasets.

- Dataset
  - **BANKING**
    - 은행 관련 dataset
  - **OOS**
    - OOD detection 위한 dataset
  - **StackOverflow**
    - Technical question dataset

- **Experimental settings**
  - Model
    - BERT
    - 마지막 layer 제외 freeze
  - KCCL
    - Positive sample 수 : 1에서 10개
    - Negative sample 수 : 1
    - 감마 : 0.25
  - 2 번째 stage
    - Decision boundary 만 학습
  - Batch size : 32
  - e: 0.5~1.2
  - s: 0~0.5

Main Results

F1-score

	Methods	BANKING ALL	BANKING Unknown	BANKING Known	OOS ALL	OOS Unknown	OOS Known	StackOverflow ALL	StackOverflow Unknown	StackOverflow Known
25%	MSP	50.09	41.43	50.55	47.62	50.88	47.53	37.85	13.03	42.82
	DOC	58.03	61.42	57.85	66.37	81.98	65.96	47.73	41.25	49.02
	OpenMax	54.14	51.32	54.28	61.99	75.76	61.62	45.98	36.41	47.89
	DeepUnk	61.36	70.44	60.88	71.16	87.33	70.73	52.05	49.29	52.60
	ADB	71.62	84.56	70.94	77.19	91.84	76.80	80.83	90.88	78.82
	CLAB	**75.87**	**88.73**	**75.20**	**81.26**	**94.49**	**80.91**	**86.03**	**95.12**	**84.21**
50%	MSP	71.18	41.19	71.97	70.41	57.62	70.58	63.01	23.99	66.91
	DOC	73.12	55.14	73.59	78.26	79.00	78.25	62.84	25.44	66.58
	OpenMax	74.24	54.33	74.76	80.56	81.89	80.54	68.18	45.00	70.49
	DeepUnk	77.53	69.53	77.74	82.16	85.85	82.11	68.01	43.01	70.51
	ADB	80.90	78.44	80.96	85.05	88.65	85.00	85.83	87.34	85.68
	CLAB	**83.08**	**81.82**	**83.36**	**87.03**	**90.63**	**86.98**	**87.68**	**89.30**	**87.52**
75%	MSP	83.60	39.23	84.36	82.38	59.08	82.59	77.95	33.96	80.88
	DOC	83.34	50.60	83.91	83.59	72.87	83.69	75.06	16.76	78.95
	OpenMax	84.07	50.85	84.64	73.16	76.35	73.13	79.78	44.87	82.11
	DeepUnk	84.31	58.54	84.75	86.23	81.15	86.27	78.28	37.59	81.00
	ADB	85.96	66.47	86.29	88.53	83.92	88.58	85.99	73.86	86.80
	CLAB	**88.12**	**70.74**	**88.42**	**90.53**	**86.91**	**90.57**	**88.11**	**77.59**	**88.81**

Ablation Study

25% dataset

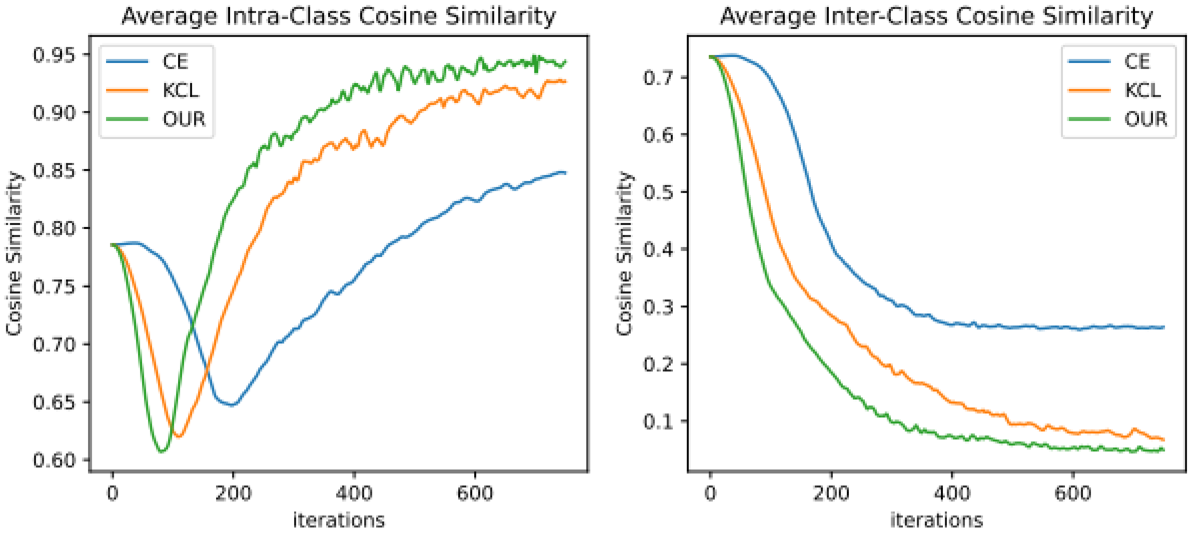
Dataset	Method	ALL	Unknown	Known
BANKING	ADB	71.62	84.56	70.94
	CLAB	**75.87**	**88.73**	**75.20**
	w/o ADBES	74.11	86.17	73.47
	w/o KCCL	73.10	86.28	72.41
OOS	ADB	77.19	91.84	76.80
	CLAB	**81.26**	**94.49**	**80.91**
	w/o ADBES	79.25	92.85	78.89
	w/o KCCL	78.86	92.61	78.50
StackOverflow	ADB	80.83	90.88	78.82
	CLAB	**86.03**	**95.12**	**84.21**
	w/o ADBES	84.41	94.15	82.47
	w/o KCCL	83.35	92.88	81.44

Analysis of K-Center Contrastive Learning

- **Performance Comparison between KCCL and KCL**

BANKING dataset

Percent	Method	ALL	Unknown	Known
25%	ADBES	73.10	86.28	72.41
	KCL+ADBES	74.16	86.31	73.52
	KCCL+ADBES	**75.87**	**88.73**	**75.20**
50%	ADBES	81.46	**81.86**	81.45
	KCL+ADBES	81.95	81.12	82.19
	KCCL+ADBES	**83.08**	81.82	**83.36**
75%	ADBES	86.32	69.02	86.59
	KCL+ADBES	87.12	70.14	87.40
	KCCL+ADBES	**88.12**	**70.74**	**88.42**



Intra-Class : IND 분포 내에 있는 data
Inter-Class: 다른 IND 사이에 분포된 data

Analysis of K-Center Contrastive Learning

- **Effect of Positive Samples in KCCL**

25% dataset

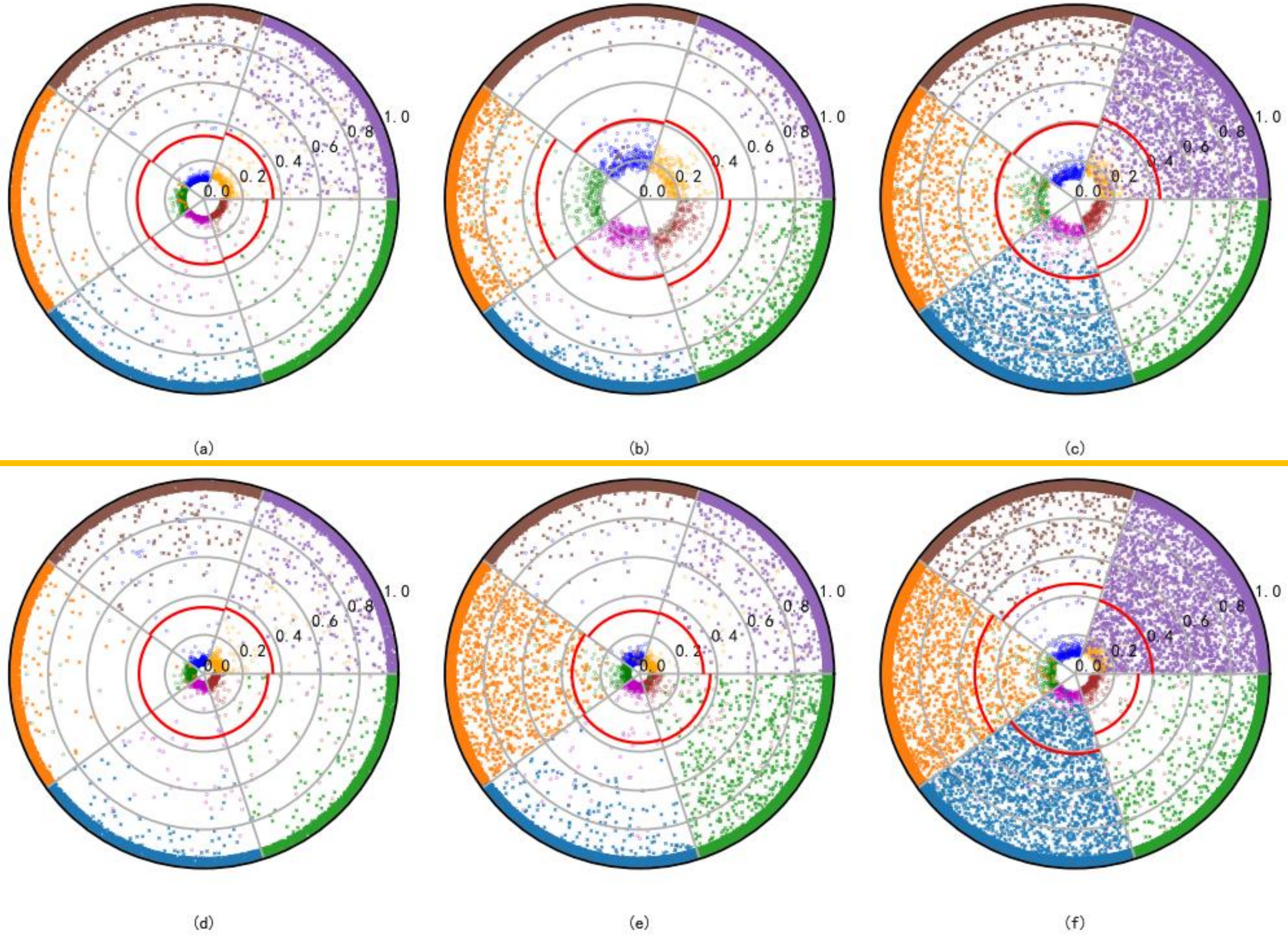
K	**BANKING**	**OOS**	**StackOverflow**
1	72.58	80.18	86.25
2	73.31	80.27	86.39
3	73.21	**80.86**	86.37
4	74.83	80.58	**86.61**
5	**75.01**	80.49	86.26
6	74.55	80.64	85.99
7	72.99	80.75	85.31
8	74.32	80.74	84.31
9	73.01	80.34	84.60
10	72.97	80.74	85.18

Distance Distribution Visualization

Test set에서 수행
StackOverflow dataset

CLAB

ADB

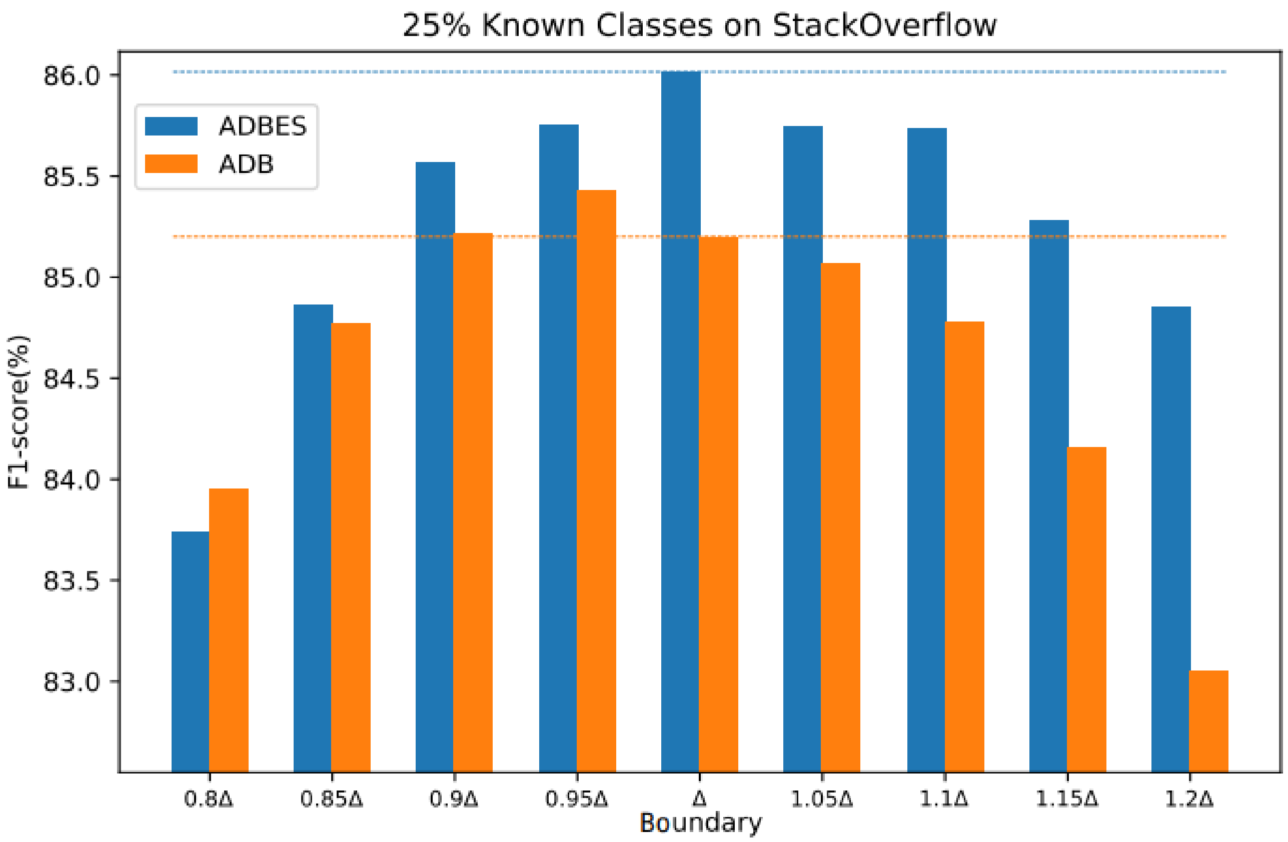


Conclusion

- Open intent 분류 위한 Two-stage method **CLAB** 제안
  - **K**-**C**enter **C**ontrastive **L**earning algorithm 제안
    - 차별적이고 균형 있는 intent feature 학습 위해
    - Open intent 인식 위한 모델의 일반화를 개선
  - **A**djustable **D**ecision **B**oundary learning method with **E**xpanding and **S**hrinking 제안
    - 정확한 decision 상황 결정위해
    - OOD 가 중심으로 부터 멀리 있다면 반지름을 늘리고 그 반대 상황에선 축소

Thank you

Analysis of Adjustable Decision Boundary



Cosine similarity

$$similarity = cos(\Theta) = \frac{A \cdot B}{||A||\ ||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}}$$