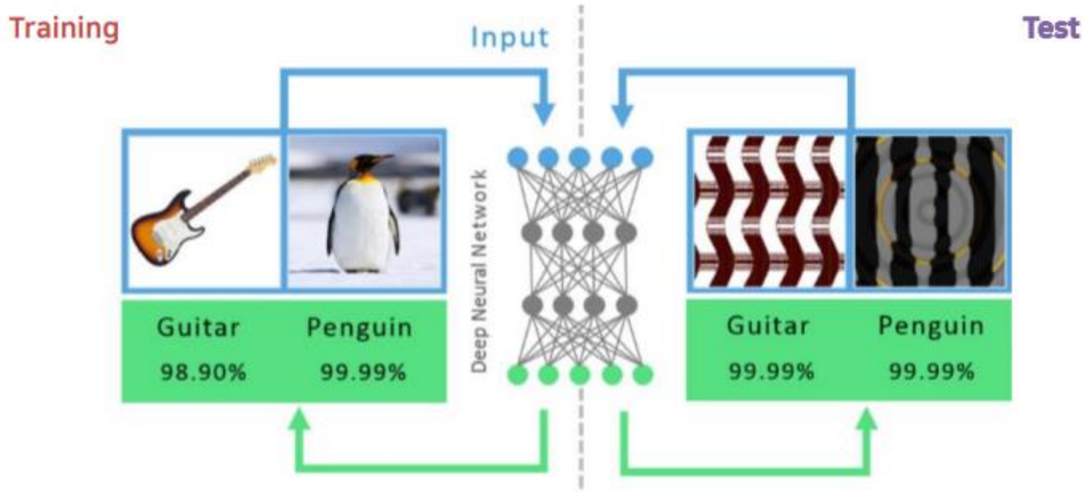
Deep Open Intent Classification with Adaptive Decision Boundary

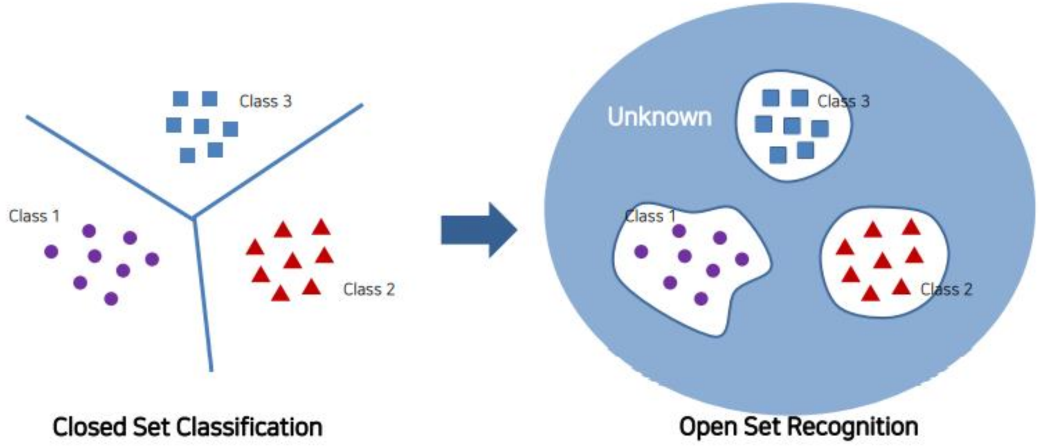
Hanlei Zhang, Hua Xu, Ting-En Lin

전북대학교 IT정보공학과
202018392
박나현

Open world classification

- 기존 심층 신경망에서는 학습한 class에 대한 확률만 출력가능
- **Open set recognition**
  - 모델을 학습할 때 처음 보는 class에 대한 데이터가 나타났을 때 학습 class 중 하나가 아닌 'unknown'으로 분류를 할 수 있는 기능을 수행하는 방법





Open intent classification

- Goal
  - To classify the **n-class** known intents into their corresponding classes correctly while identifying the $(\mathbf{n}+\mathbf{1})^{th}$ **class open intent**

User utterances	Intent Label
Book a flight from LA to Madrid.	Book flight
Can you get me a table at Steve's?	Restaurant reservation
Book Delta ticket Madison to Atlanta.	Book flight
Schedule me a table at Red Lobster.	Restaurant reservation
...	...
Can you tell me the name of this song?	**Open**
Look up the calories in an apple.	**Open**

n개의 class로 분류

n+1번째 새로운 class = open intent

기존 문제를 해결

기존문제

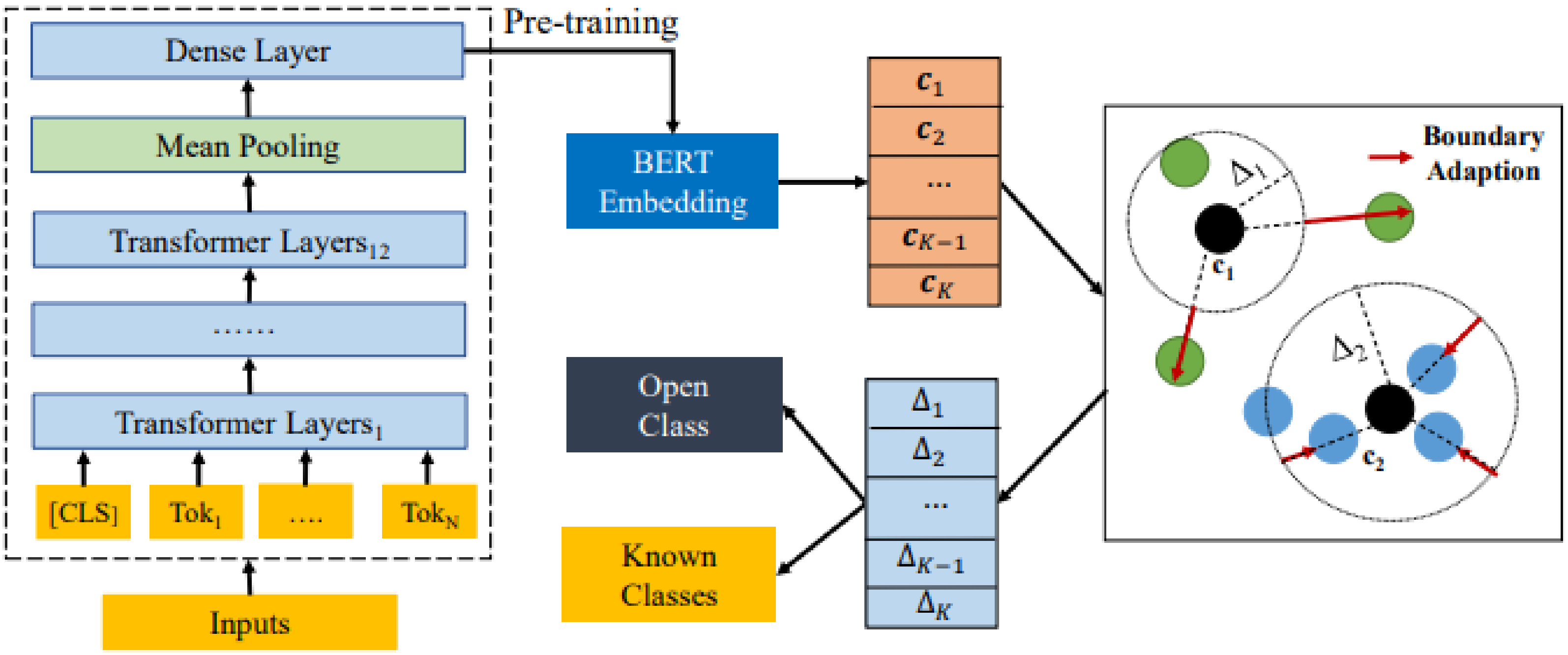
- need to design specific classifiers for identifying the open class and perform poorly with the common classifier.
- It is a complicated and time-consuming process to manually select the optimal decision condition.

제안한 모델에서 해결

- model architecture 변경 불필요
- 자동적으로 decision boundary 학습

Model architecture

- Use known intents as prior knowledge
- Propose a novel post-processing method to learn the adaptive decision boundary (ADB) for open intent classification



Intent Representation

- Use BERT model to extract deep intent features

Hidden layer size

$$[CLS, T_1, \cdots, T_N] \in \mathbb{R}^{(N+1)\times H}$$

- **Mean pooling**
  - to synthesize the high-level semantic features in one sentence
  - Get the averaged representation $\boldsymbol{x}_i \in \mathbb{R}^H$

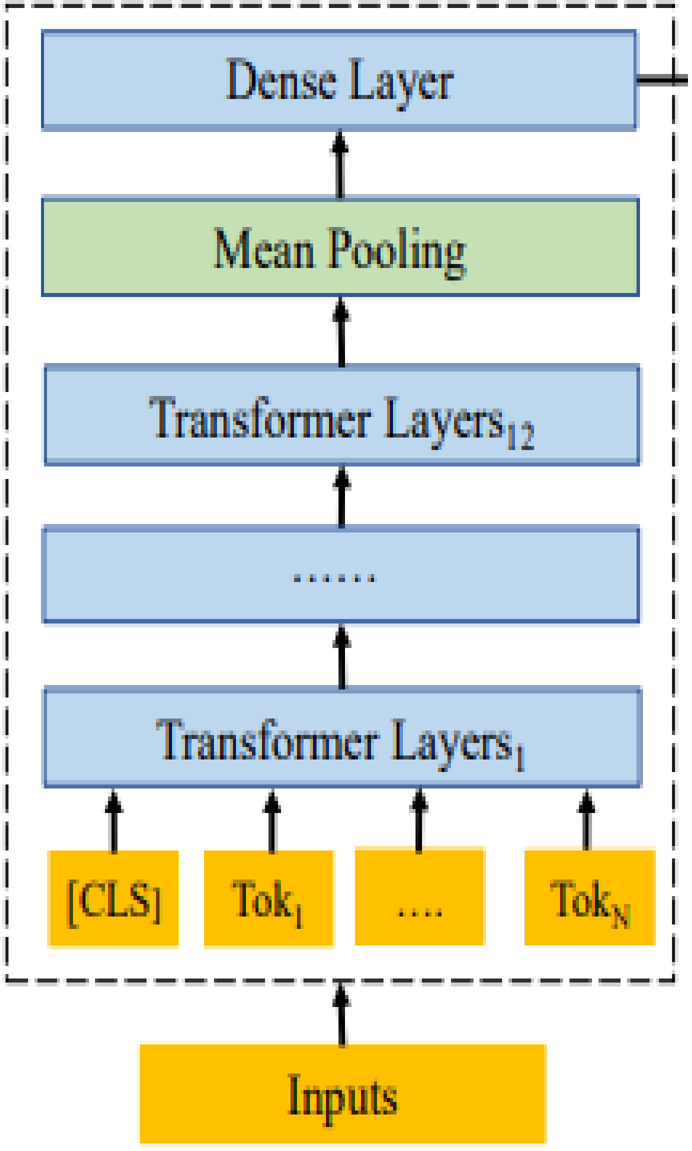
$$\boldsymbol{x}_i = \text{mean-pooling}([CLS, T_1, \cdots, T_N])$$

- **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 $b_h \in \mathbb{R}^D$



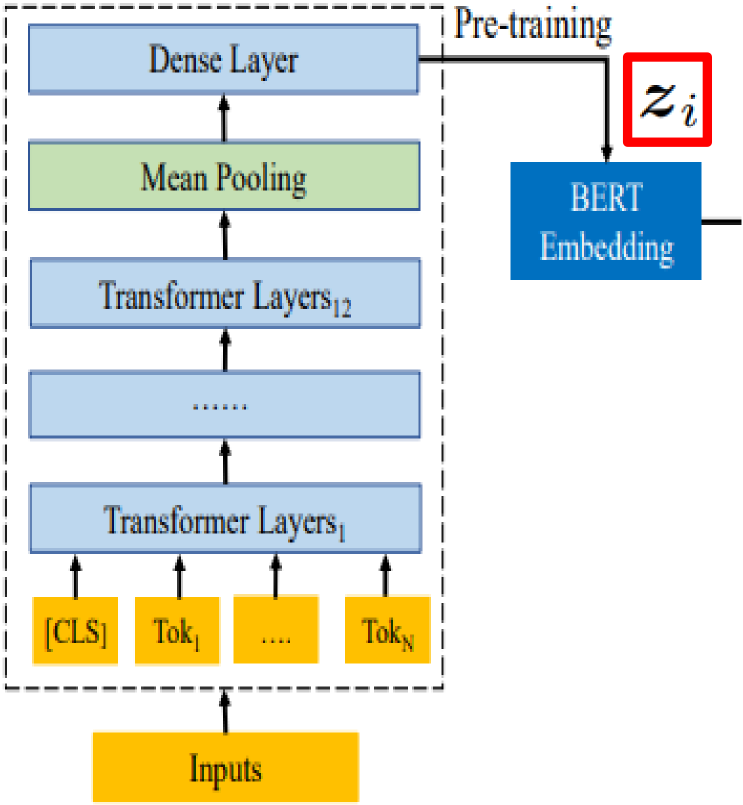
Pre-training

- **Due to lack of open intent samples, we use known intents as prior knowledge to pre-train the model.**
- Use Simple softmax loss to learn the intent feature $\boldsymbol{z}_i$

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

$\phi(\cdot)$: linear classifier

$\phi(\cdot)^j$: j번째 output logits



Adaptive Decision Boundary learning

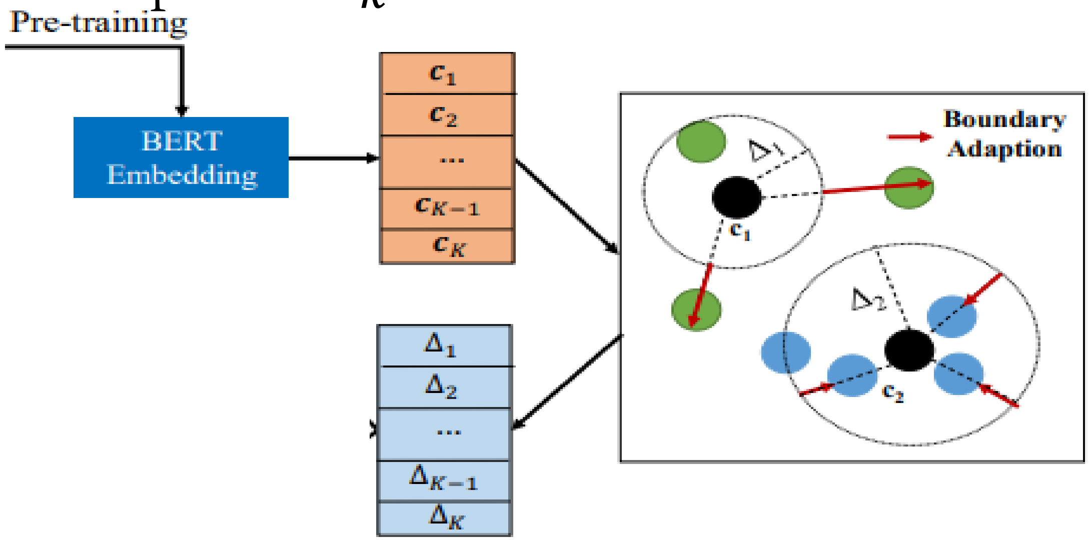
- **Decision Boundary Formulation**
  - Known intent label with their corresponding labels

$$S = \{(\boldsymbol{z}_i, y_i), \ldots, (\boldsymbol{z}_N, y_N)\}$$

  - The centroid is the mean vector of embedded samples in S_k
    - S_k: set of examples labeled with class k

centroid $\boldsymbol{c}_k \in \mathbb{R}^D$

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



Adaptive Decision Boundary learning

- For each known intent z_i, we aim to satisfy the following constraints

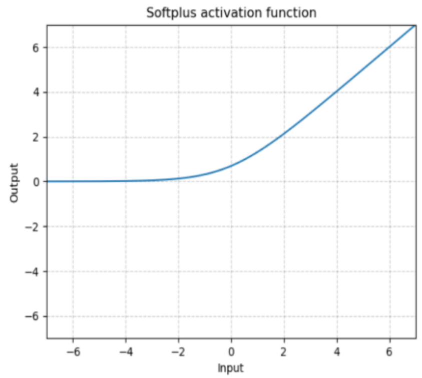
$$\forall \boldsymbol{z}_i \in S_k, \underbrace{\|\boldsymbol{z}_i - \boldsymbol{c}_k\|_2}_{\text{Euclidean distance}} \leq \underbrace{\Delta_k}_{\text{radius}},$$

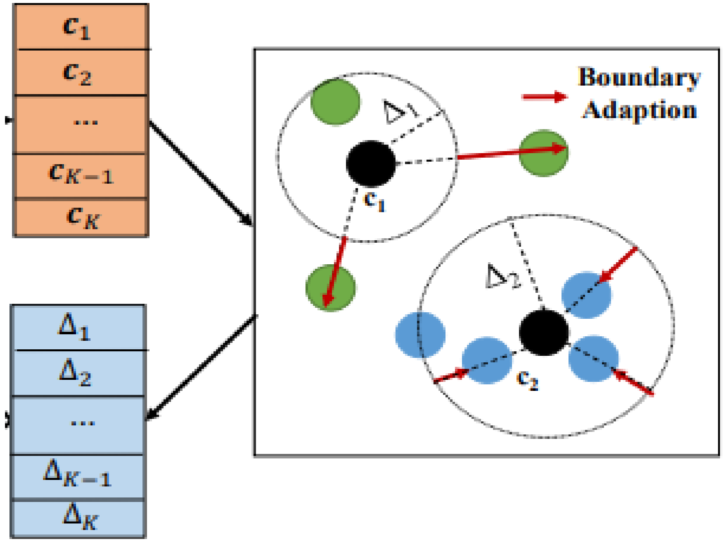
  - we hope examples belonging to class k are constrained in the ball area with centroid and radius.

- We use Softplus activation function as mapping between radius and the learnable parameter.

$$\Delta_k = \log\left(1 + e^{\widehat{\Delta_k}}\right)$$

  - SoftPlus
    - Sigmoid 적분값, relu함수 부드럽게
    - Learned radius 항상 0이상





Adaptive Decision Boundary learning

- The suitable decision boundaries should satisfy **two conditions**
  - They should be broad enough to surround known intent samples as much as possible
  - They need to be tight enough to prevent the open intent samples from being identified as known intents

Adaptive Decision Boundary learning

- **Boundary learning**
  - The decision boundaries should be adaptive to the intent feature space **to balance both empirical and open space risk**
    - **Open space risk : the classifier should not cover too much empty space**
  - **Boundary Loss**

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

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

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

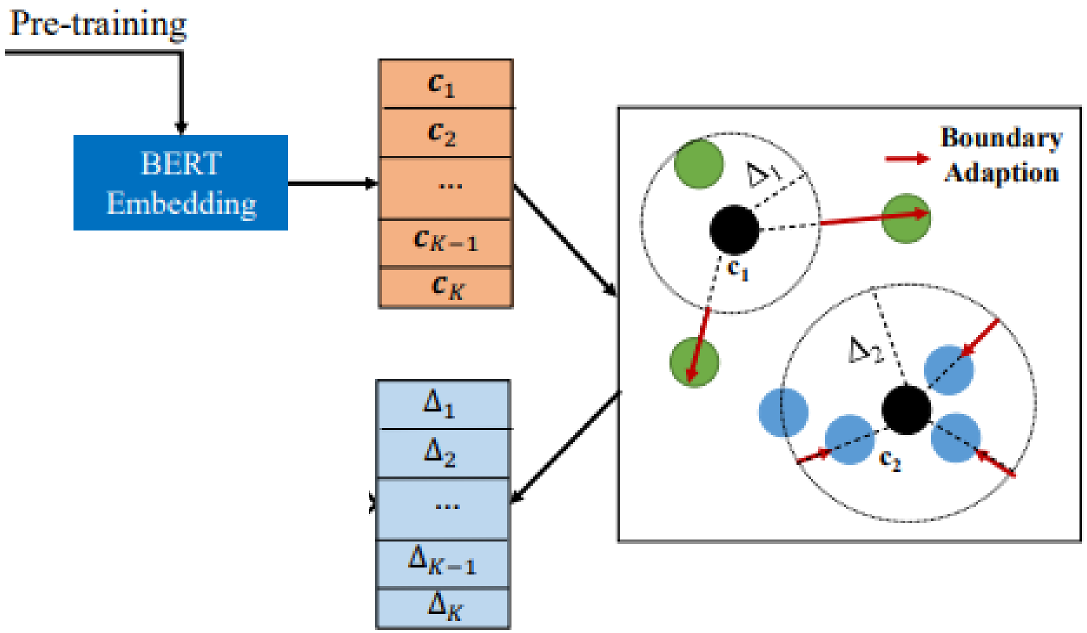
Adaptive Decision Boundary learning

- **Update the boundary parameters**

Learning rate of the boundary parameter

$$\widehat{\Delta_k} := \widehat{\Delta_k} - \eta \frac{\partial \mathcal{L}_b}{\partial \widehat{\Delta_k}}$$

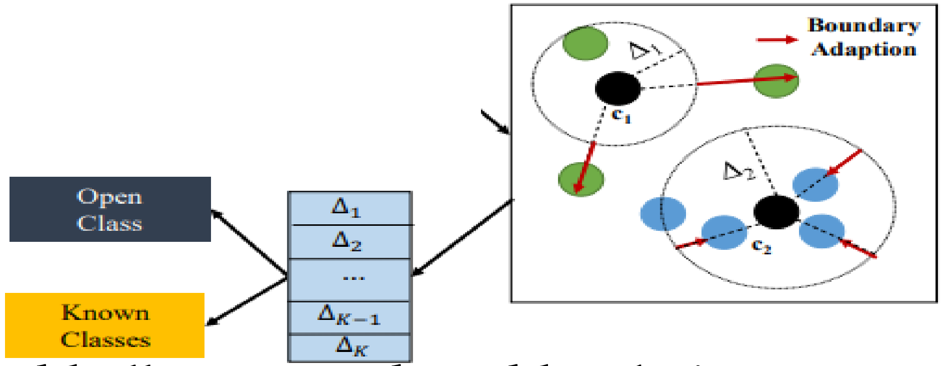
$$\frac{\partial \mathcal{L}_b}{\partial \widehat{\Delta_k}} = \frac{\sum_{i=1}^{N} \delta'\left(y_i = k\right) \cdot (-1)^{\delta_i}}{\sum_{i=1}^{N} \delta'\left(y_i = k\right)} \cdot \frac{1}{1 + e^{-\widehat{\Delta_k}}},$$



- Boundary loss로 boundary를 intent feature space에 적용하고 적절한 decision boundary를 학습 가능

Open Classification with Decision Boundary

- After training, we use the centroids and the learned decision boundaries of each known class



- 가정
  - Known intent samples are constrained in the closed ball area produced by their corresponding centroids and decision boundaries.
  - The open intent samples are outside any of the bounded spherical areas

Intent label $\mathcal{Y} = \{1, 2, \cdots, K\}$

$$\hat{y} = \begin{cases} \text{open, if } \overbrace{d(\boldsymbol{z}_i, \boldsymbol{c}_k)}^{\text{Euclidean distance}} > \Delta_k, \forall k \in \mathcal{Y}; \\ \arg\min_{k \in \mathcal{Y}} d(\boldsymbol{z}_i, \boldsymbol{c}_k), \text{otherwise}, \end{cases}$$

Experiments

- Dataset

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.

  - **BANKING**
    - A fine-grained dataset in the banking domain
  - **OOS**
    - A dataset for intent classification and out-of-scope prediction
  - **StackOverflow**
    - Processed dataset, which has 20 different classes and 1000 samples for each class

Results

- Experimental setting
  - The number of known classes are varied with the proportions of 25%, 50%, and 75% in the training set.

	Methods	BANKING Accuracy	BANKING F1-score	OOS Accuracy	OOS F1-score	StackOverflow Accuracy	StackOverflow F1-score
25%	MSP	43.67	50.09	47.02	47.62	28.67	37.85
	DOC	56.99	58.03	74.97	66.37	42.74	47.73
	OpenMax	49.94	54.14	68.50	61.99	40.28	45.98
	DeepUnk	64.21	61.36	81.43	71.16	47.84	52.05
	ADB	**78.85**	**71.62**	**87.59**	**77.19**	**86.72**	**80.83**
50%	MSP	59.73	71.18	62.96	70.41	52.42	63.01
	DOC	64.81	73.12	77.16	78.26	52.53	62.84
	OpenMax	65.31	74.24	80.11	80.56	60.35	68.18
	DeepUnk	72.73	77.53	83.35	82.16	58.98	68.01
	ADB	**78.86**	**80.90**	**86.54**	**85.05**	**86.40**	**85.83**
75%	MSP	75.89	83.60	74.07	82.38	72.17	77.95
	DOC	76.77	83.34	78.73	83.59	68.91	75.06
	OpenMax	77.45	84.07	76.80	73.16	74.42	79.78
	DeepUnk	78.52	84.31	83.71	86.23	72.33	78.28
	ADB	**81.08**	**85.96**	**86.32**	**88.53**	**82.78**	**85.99**

Results

Macro F1-score

	Methods	BANKING		OOS		StackOverflow	
		Open	Known	Open	Known	Open	Known
25%	MSP	41.43	50.55	50.88	47.53	13.03	42.82
	DOC	61.42	57.85	81.98	65.96	41.25	49.02
	OpenMax	51.32	54.28	75.76	61.62	36.41	47.89
	DeepUnk	70.44	60.88	87.33	70.73	49.29	52.60
	ADB	**84.56**	**70.94**	**91.84**	**76.80**	**90.88**	**78.82**
50%	MSP	41.19	71.97	57.62	70.58	23.99	66.91
	DOC	55.14	73.59	79.00	78.25	25.44	66.58
	OpenMax	54.33	74.76	81.89	80.54	45.00	70.49
	DeepUnk	69.53	77.74	85.85	82.11	43.01	70.51
	ADB	**78.44**	**80.96**	**88.65**	**85.00**	**87.34**	**85.68**
75%	MSP	39.23	84.36	59.08	82.59	33.96	80.88
	DOC	50.60	83.91	72.87	83.69	16.76	78.95
	OpenMax	50.85	84.64	76.35	73.13	44.87	82.11
	DeepUnk	58.54	84.75	81.15	86.27	37.59	81.00
	ADB	**66.47**	**86.29**	**83.92**	**88.58**	**73.86**	**86.80**

Discussion

Boundary Learning Process

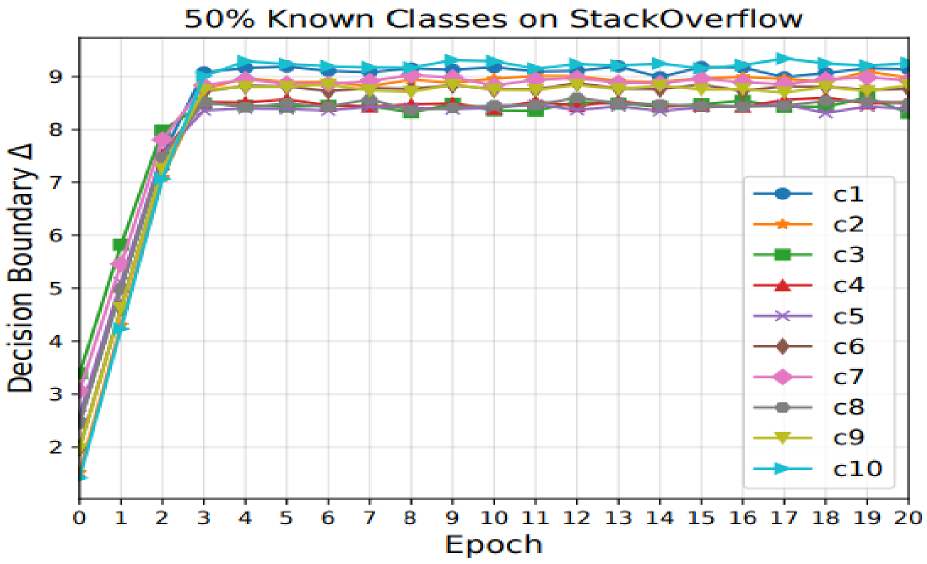


Figure 3: The boundary learning process.

Balance between empirical risk and open space risk

Effect of Decision Boundary

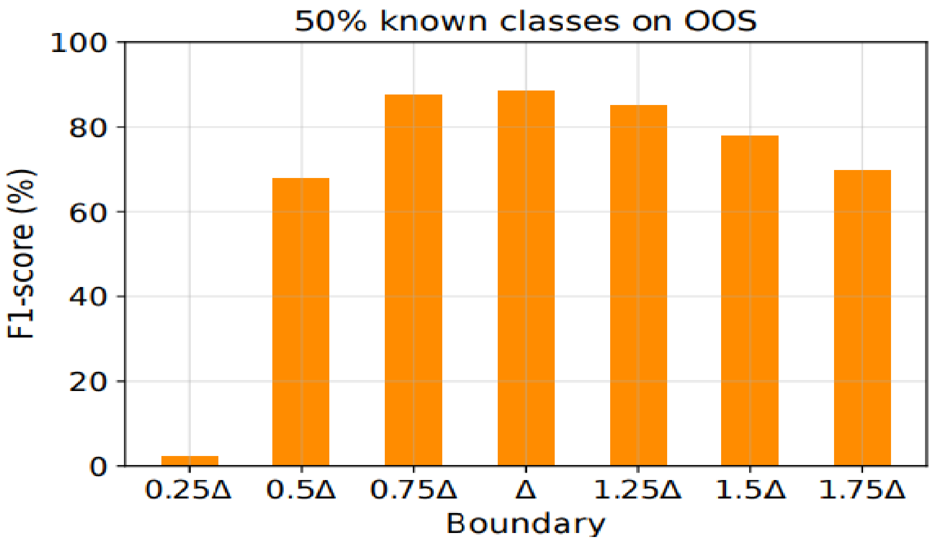


Figure 4: Influence of the learned decision boundary.

Performance of open classification is sensitive to the size of the decision boundaries

Thank you