Method

Self-Evolution Learning for Mixup: Enhance Data Augmentation on Few-Shot Text Classification Tasks

Task

Advisor : Jia-Ling, Koh

Speaker : Yu-Zhi, Liu

Source : ACL '23

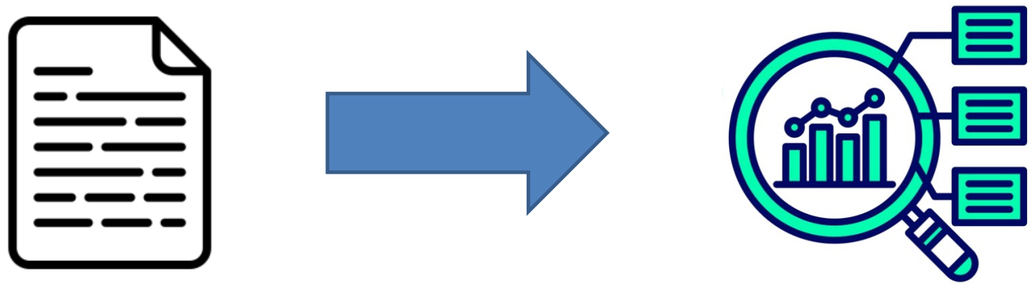
Date : 2024/3/26

Outline

- Introduction
- Method
- Experience
- Conclusion

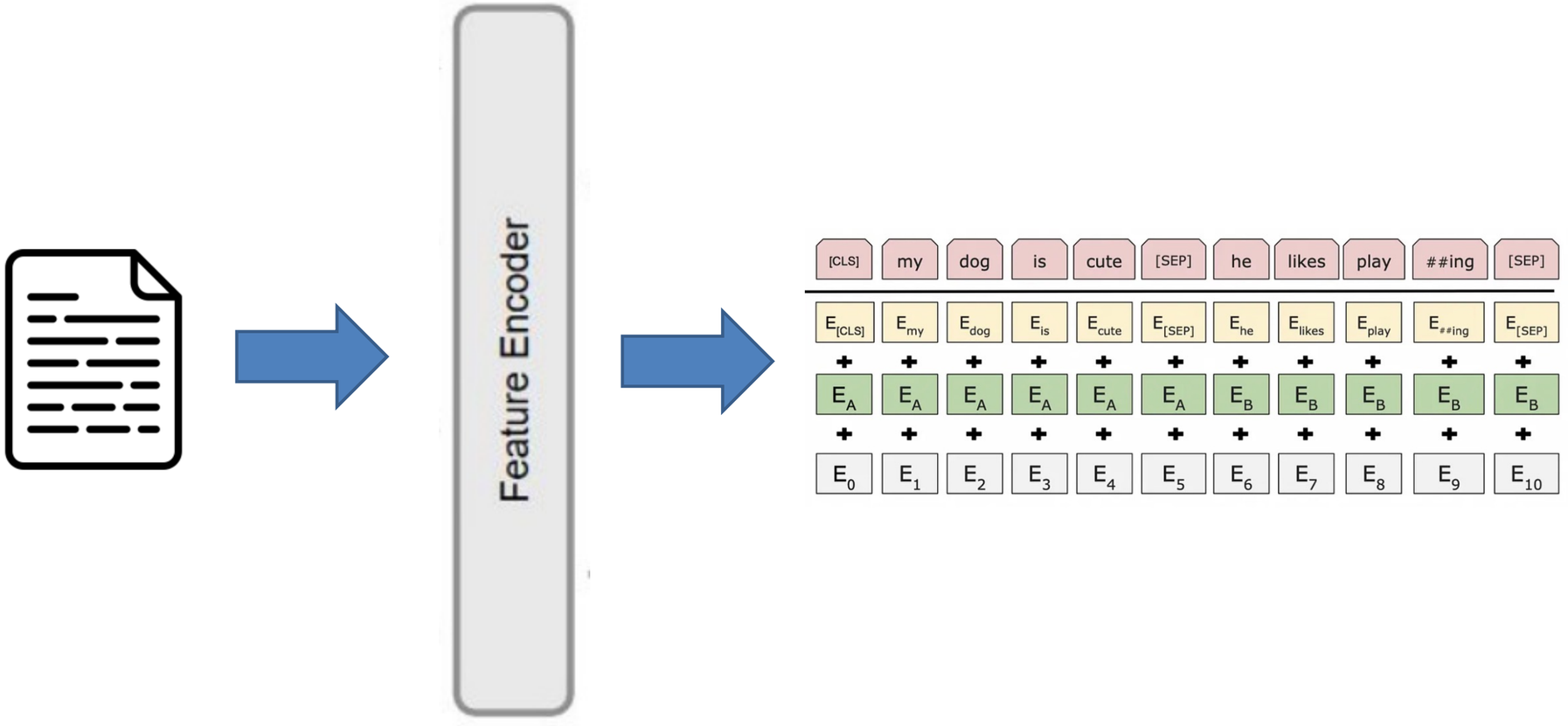
Few-Shot Text Classification Tasks

- Problem: training label limited



Classification

Data Augmentation

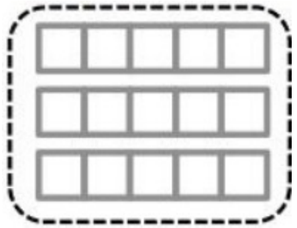


Data Augmentation Method

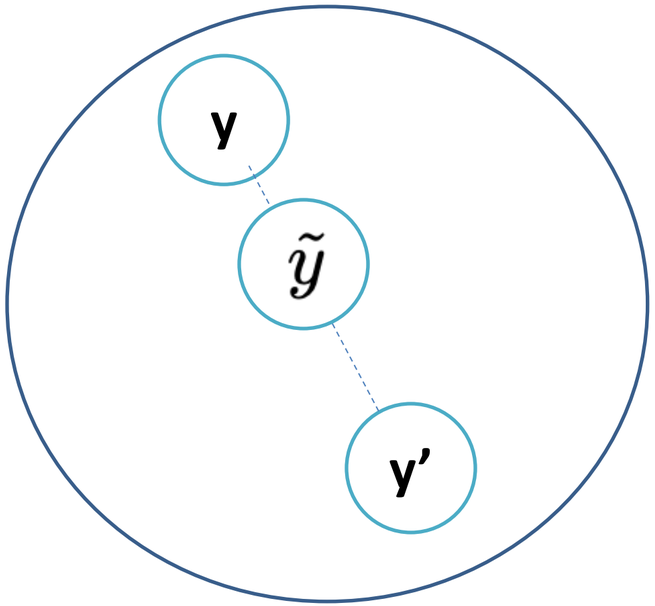
Data Augmentation

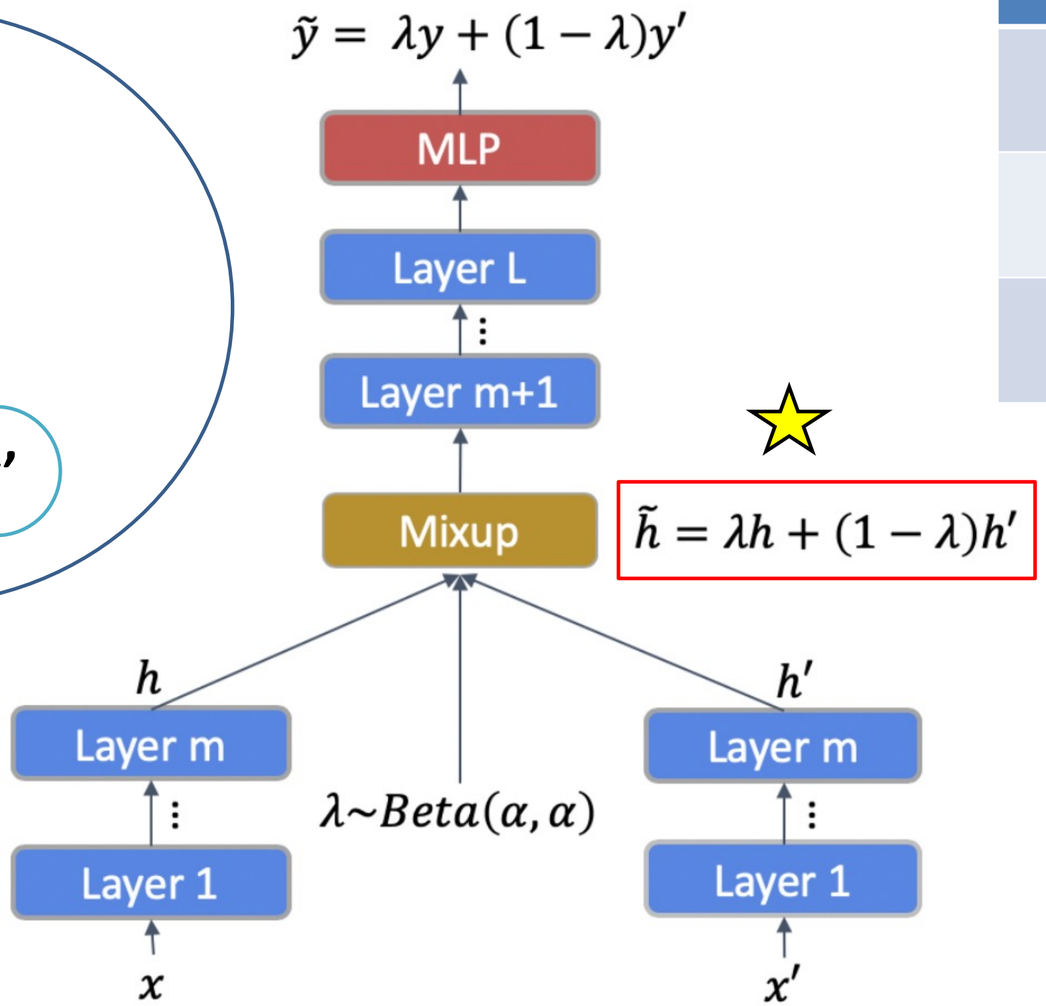
實體data

表示法



Data Augmentation Mixup



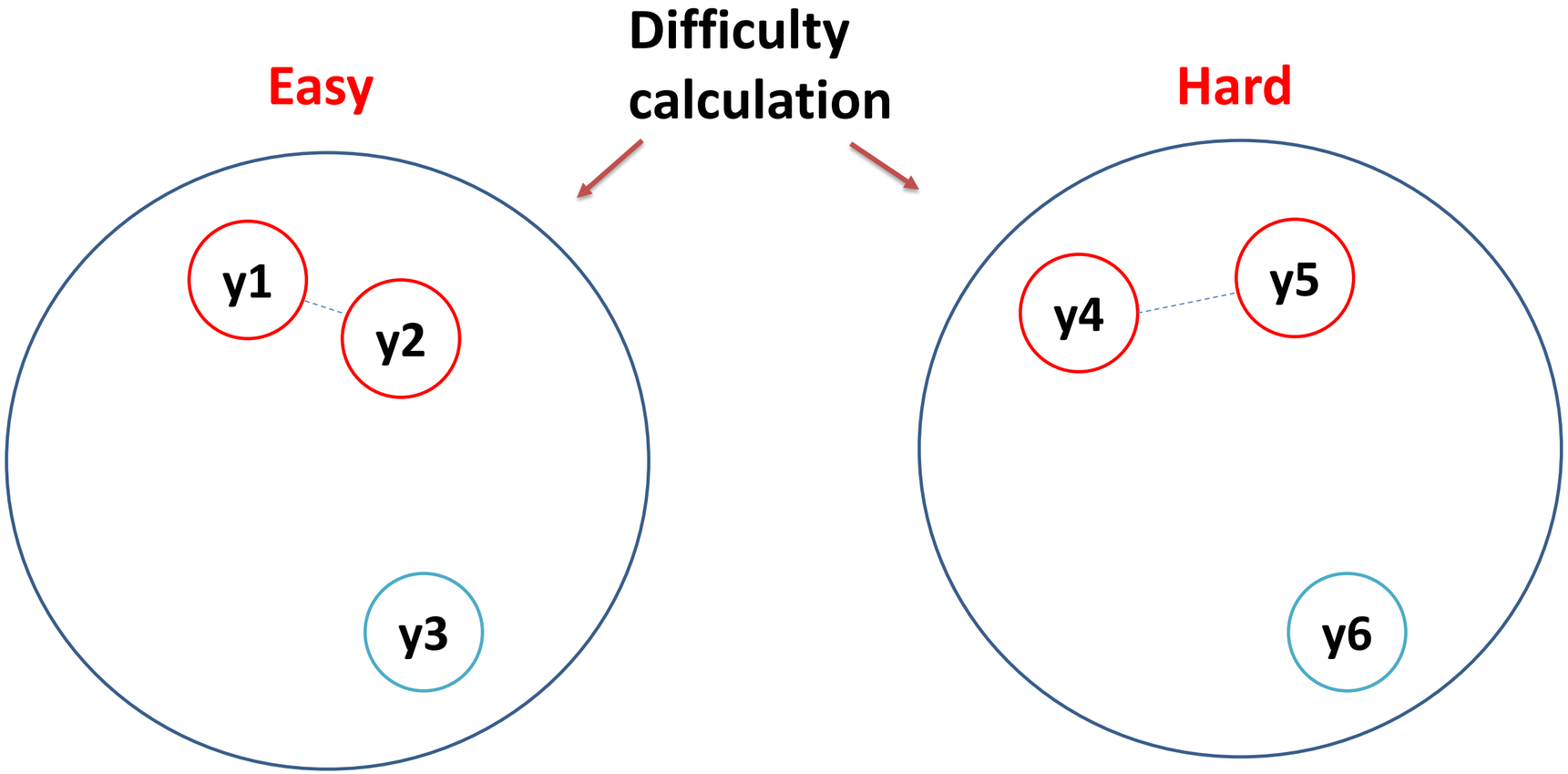


	C1	C2	C3
y	0	0	1
y‘	0	1	0
y~	0	0.2	0.8

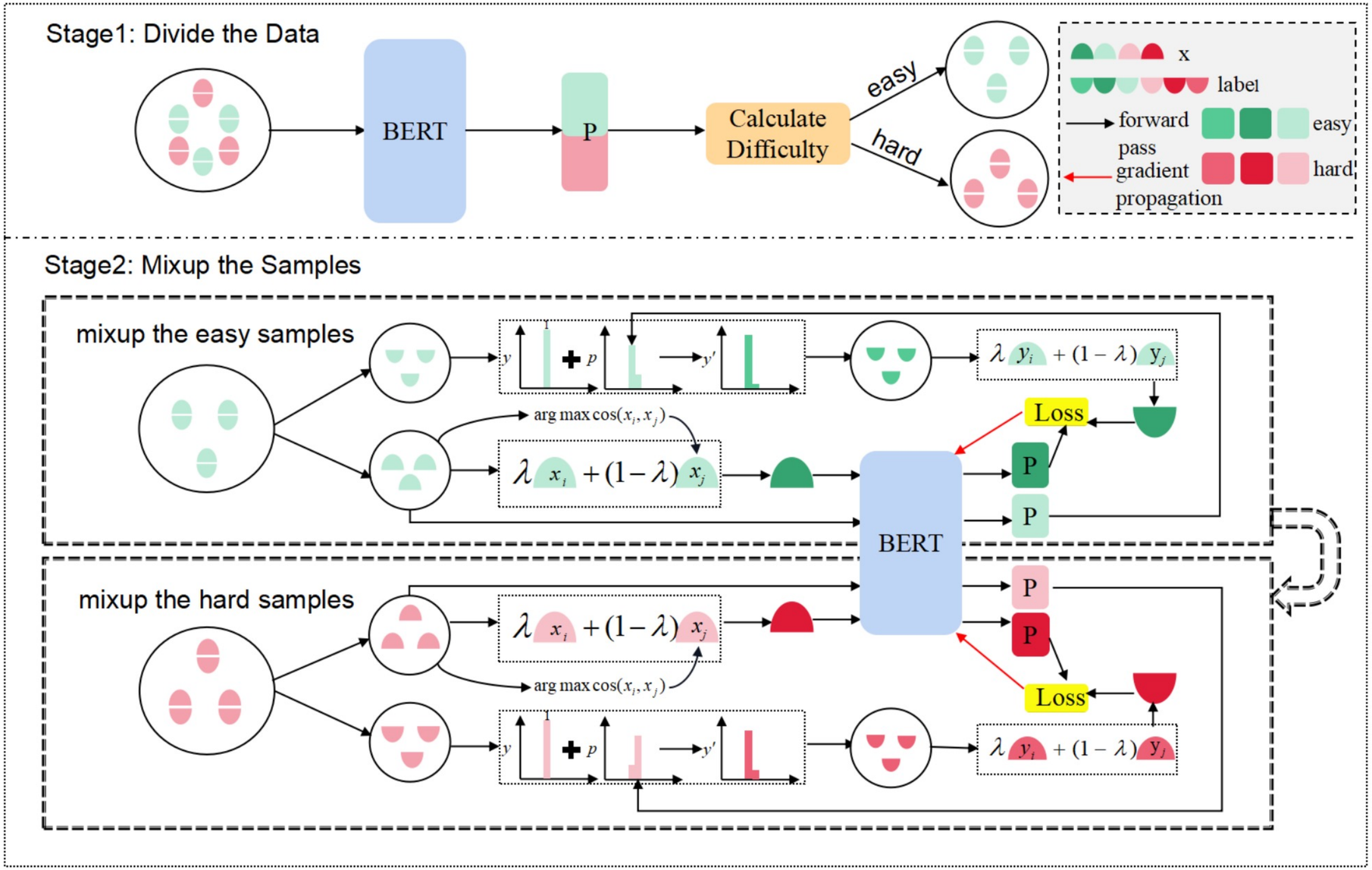
λ=0.2

<-輸入兩個樣本X與X’, y與y’

Data Selection



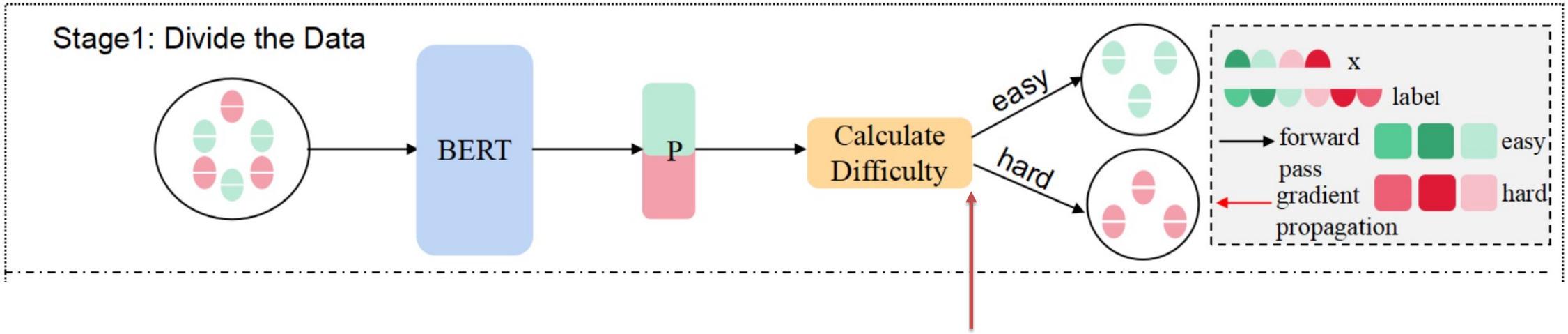
Architecture



Outline

- Introduction
- Method
- Experience
- Conclusion

Dividing the dataset



計算完difficulty
取中間值
區分出easy, hard

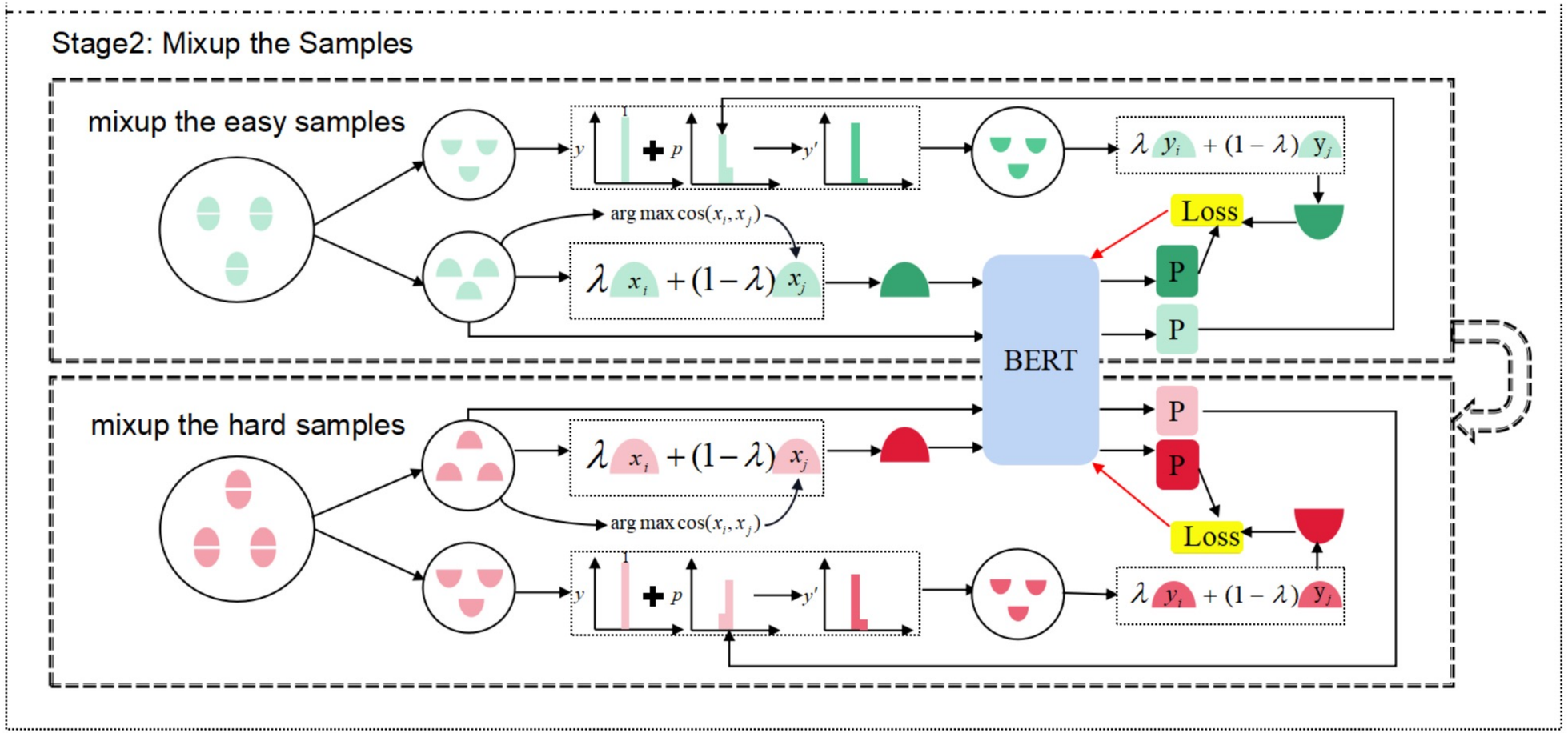
Calculate Difficulty

- dividing the dataset based on the degree of difficulty

$$d(x_i) = 1 - (p(y_i|x_i) - \max_{y \in C, y \neq y_i} p(y|x_i)), \quad (3)$$

C1	C2	C3
0.2	0.45	0.35

Mixup the samples

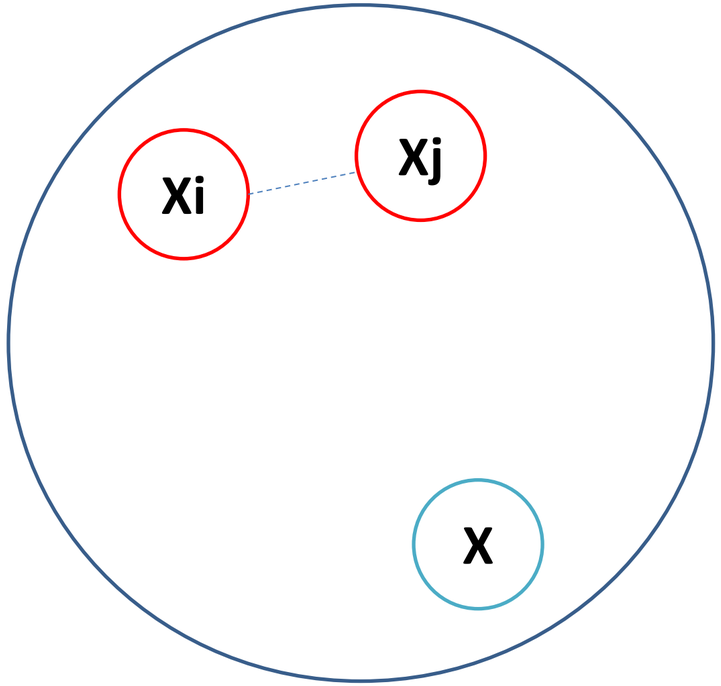


Easy to Hard

Data Selection Method

$$x_j = x_{\arg\max \cos(x_i, x_j)}$$

Given a sample xi, search for the most similar sample xj, where the similarity is measured by cosine similarity.



Text Classification Model and Mixup

input

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \quad (1)$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \quad (2)$$

label

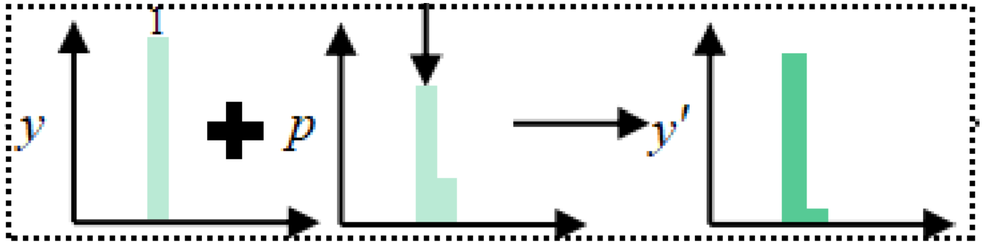
$\lambda \in [0, 1]$

Instance-Specific Label Smoothing for Regularization

$$y_i' = (1 - \alpha) * y_i + \alpha u_i$$

Old : ui (fixed)
α=0.1

	C1	C2	C3
yi	0	0	1
ui	1/3	1/3	1/3
yi’	1/30	1/30	28/30

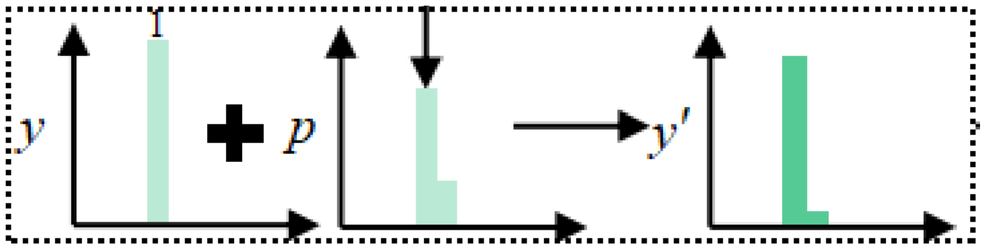


Instance-Specific Label Smoothing for Regularization

$$y'_i = (1 - \alpha) * y_i + \alpha r_i$$

New : ri (Dynamic)
α=0.1

	C1	C2	C3
yi	0	0	1
ri	0.2	0.3	0.5
yi’	0.02	0.03	0.95



Cross-Entropy Loss

$$\mathcal{L}_{LS} = -\frac{1}{m}\sum_{i=1}^{m}\tilde{y}_i' \log p_i \qquad (6)$$

類別

	C1	C2	C3
	0.4	0.4	0.2
P1 ->	0.1	0.6	0.3

Outline

- Introduction
- Method
- Experience
- Conclusion

Dataset

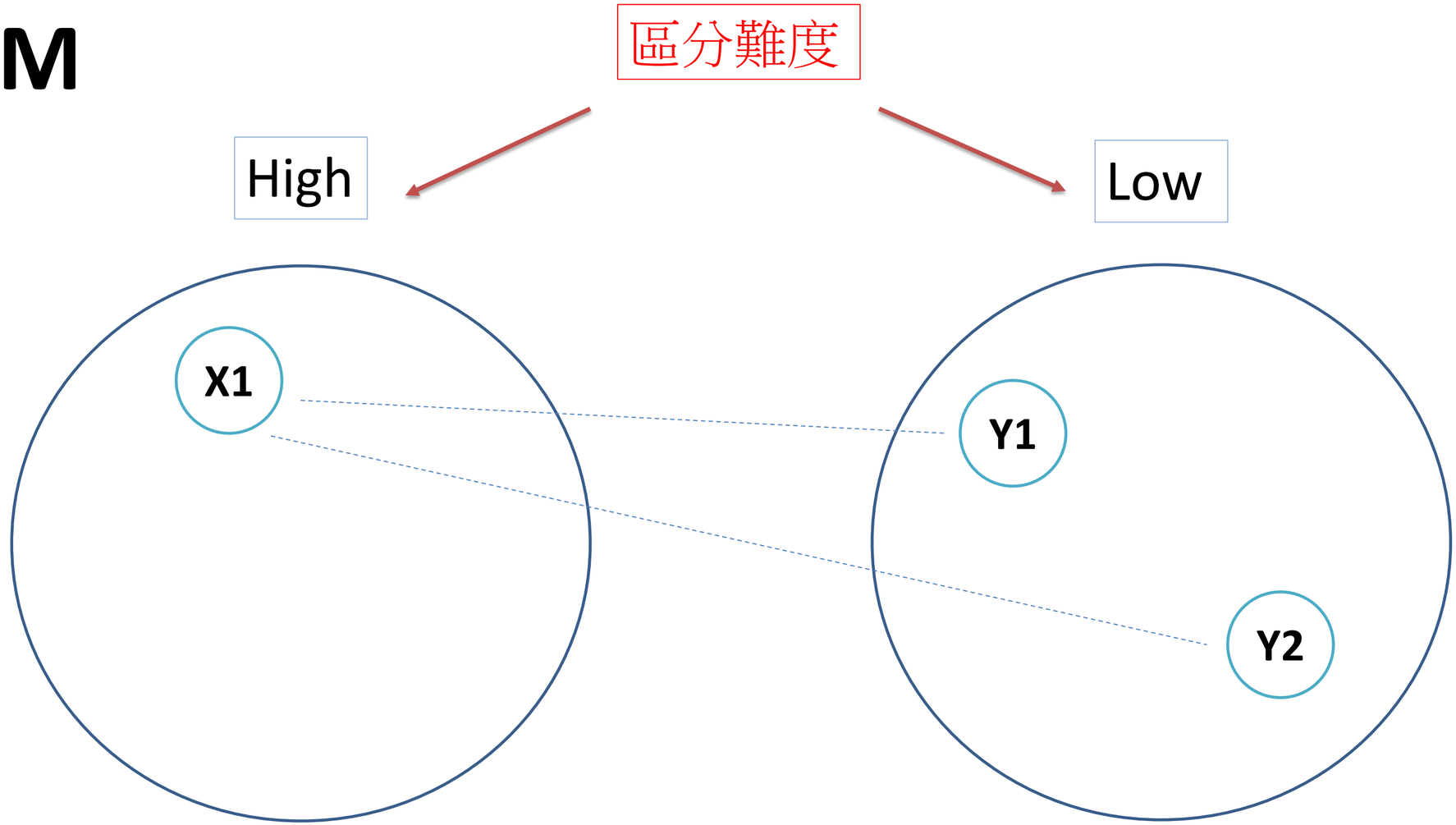
Dataset	Task	# Label	Size
SST-2	Sentiment	2	67k / 1.8k
RTE	NLI	2	2.5k / 3k
MRPC	Paraphrase	2	3.7k / 1.7k
CB	NLI	3	556 / 250
SUBJ	Classification	2	8k / 2k
Rotten tomato	Sentiment	2	8.53k / 1.07k
Amazon counterfactual	Classification	2	5k / 5k

Natural language inference

Baseline

- AUM
- Dmix
- EmbedMix
- SSMix
- TreeMix

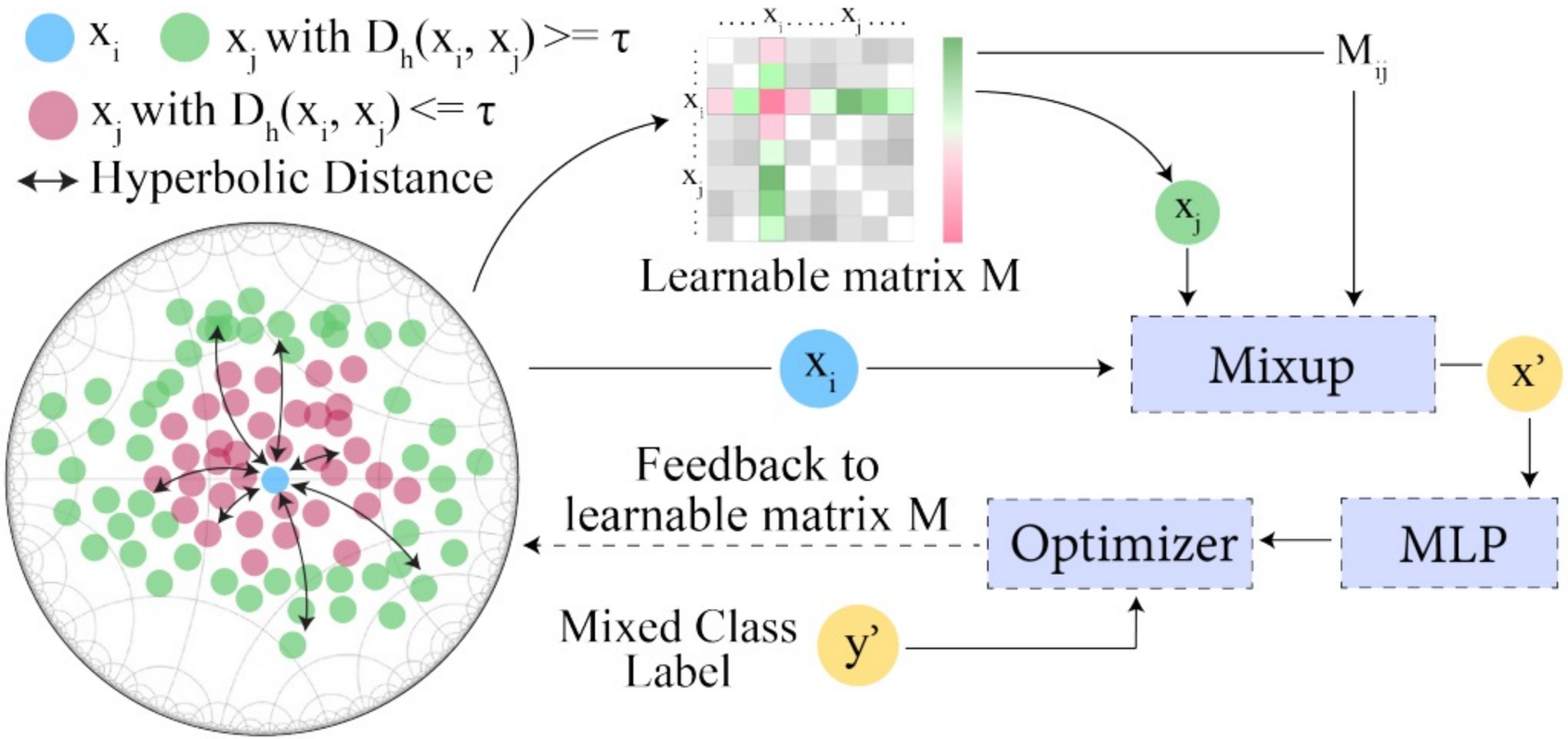
AUM



Find the most similar/dissimilar samples

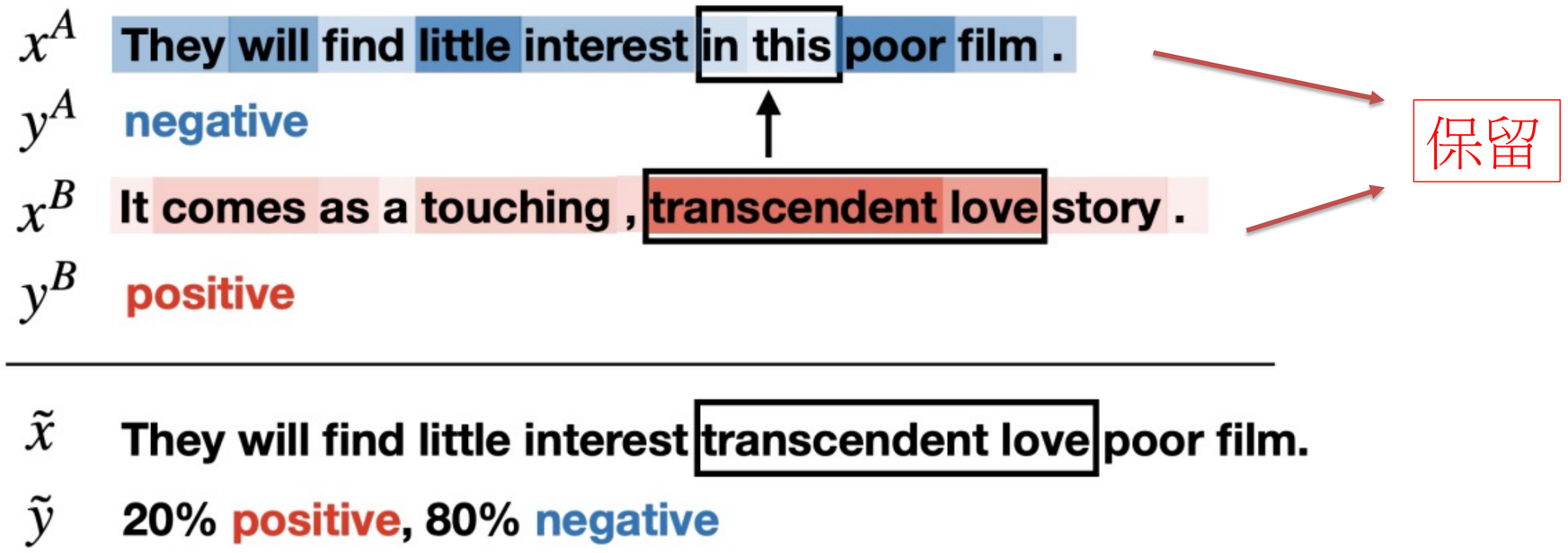
Dmix

利用distance
來調整λ



計算距離

SSMix


x^A They will find little interest in this poor film .
y^A negative
x^B It comes as a touching , transcendent love story .
y^B positive
保留
$\tilde{x}$ They will find little interest transcendent love poor film.
$\tilde{y}$ 20% positive, 80% negative

Experiment

Performance of Different Methods

NLI

Method	SST2	RTE	MRPC	CB	Rott.	SUBJ	Amazon	Score	
	Acc.	*Acc.*	*Acc.*	*Acc.*	*Acc.*	*Acc.*	*Acc.*	*Avg.*	Δ (↑)
TMix	54.94	49.60	61.90	41.06	56.95	83.16	58.14	57.95	–
-w/ AUM	56.60	49.81	62.10	42.35	58.94	83.30	65.22	59.75	+1.80
-w/ DMix	53.68	54.40	46.40	56.80	41.80	51.76	88.66	56.21	-1.74
-w/ SE (Ours)	57.56	49.99	62.69	42.85	58.23	83.87	68.58	60.53	+2.58

Experiment

Performance upon Different Mixup Methods

Method	SST2	RTE	MRPC	CB	Rott.	SUBJ	Amazon	Score	
	Acc.	*Acc.*	*Acc.*	*Acc.*	*Acc.*	*Acc.*	*Acc.*	*Avg.*	Δ (↑)
Performance upon Different Mixup Methods									
SSMix	55.70	49.52	60.10	37.13	59.86	83.76	62.63	58.08	–
-w/ SE (Ours)	56.96	49.96	61.41	39.63	61.27	84.06	65.60	59.83	+1.45
EMbedMix	53.11	49.52	61.61	37.49	58.83	83.10	63.34	58.14	–
-w/ SE (Ours)	55.89	49.88	63.28	41.07	60.10	83.86	69.22	60.46	+2.32
TreeMix	55.70	49.52	60.04	37.13	59.86	83.76	62.63	58.37	–
-w/ SE (Ours)	56.96	49.96	61.17	39.63	61.27	84.06	65.60	59.80	+1.43

+1.75

Experiment

Comparison with BERT-Large all values

Model	CB	RTE	Rott.	*Avg.*	Δ (↑)
Baseline	37.84	48.51	58.55	48.30	–
-w/ SSMix	42.49	48.37	59.67	50.17	+1.87
-w/ SE (Ours)	**47.49**	**49.16**	**62.26**	**52.97**	+4.67

Learning Strategy

Comparison with different learning strategy

Learning Strategy	SST2	Rott.	Amazon	*Avg.*
Random	55.70	59.86	60.12	58.56
★ Easy-to-hard	**55.81**	**61.17**	**65.37**	**60.78**
Hard-to-easy	55.79	61.13	64.64	60.52

Experiment

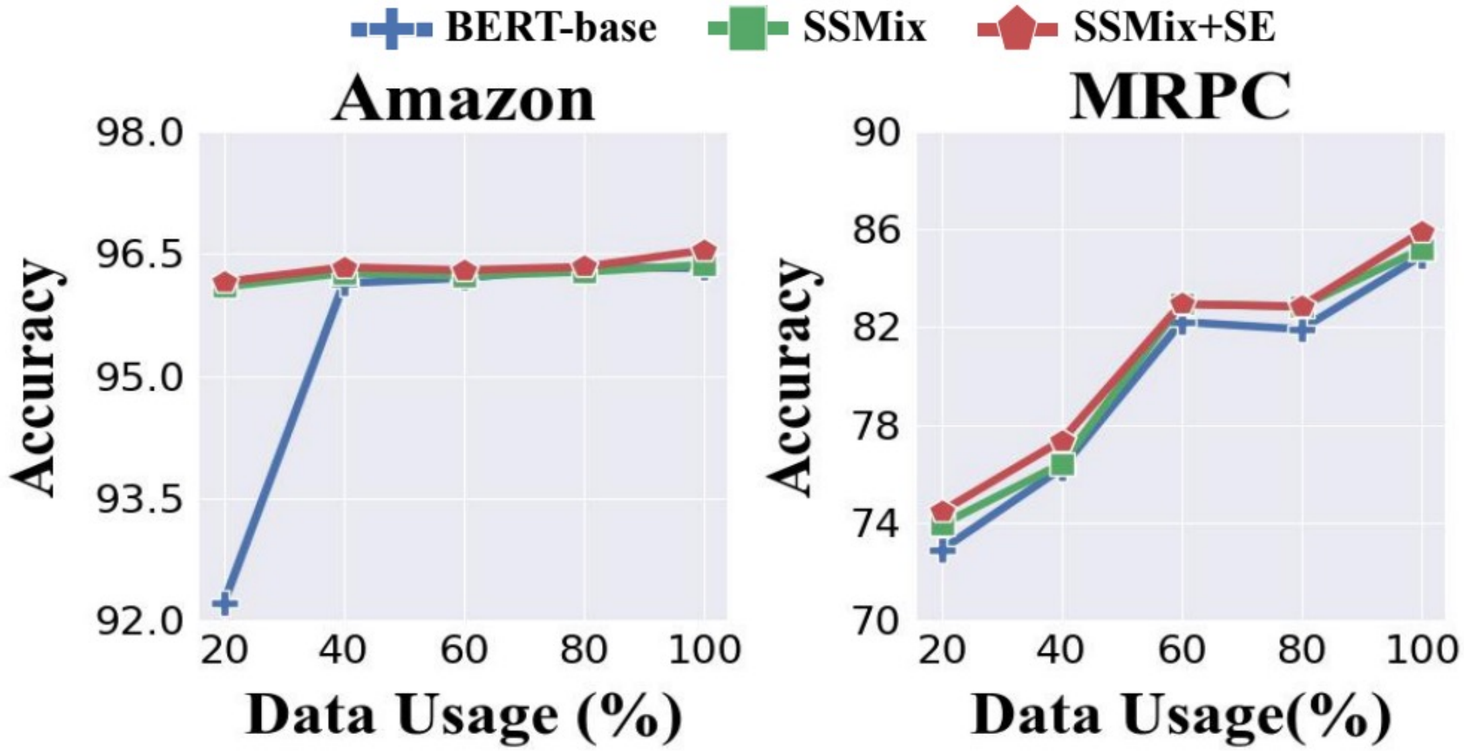
Comparison with different label smoothing

Method	SST2	RTE	Amazon	*Avg.*	Δ (↑)
SSMix	55.81	49.73	65.37	56.97	–
-w/ Vanilla LS	56.12	49.81	65.11	57.01	+0.04
-w/ ILS (Ours)	**56.88**	**49.96**	**65.52**	**57.45**	+0.48

證明smoothing的必要性

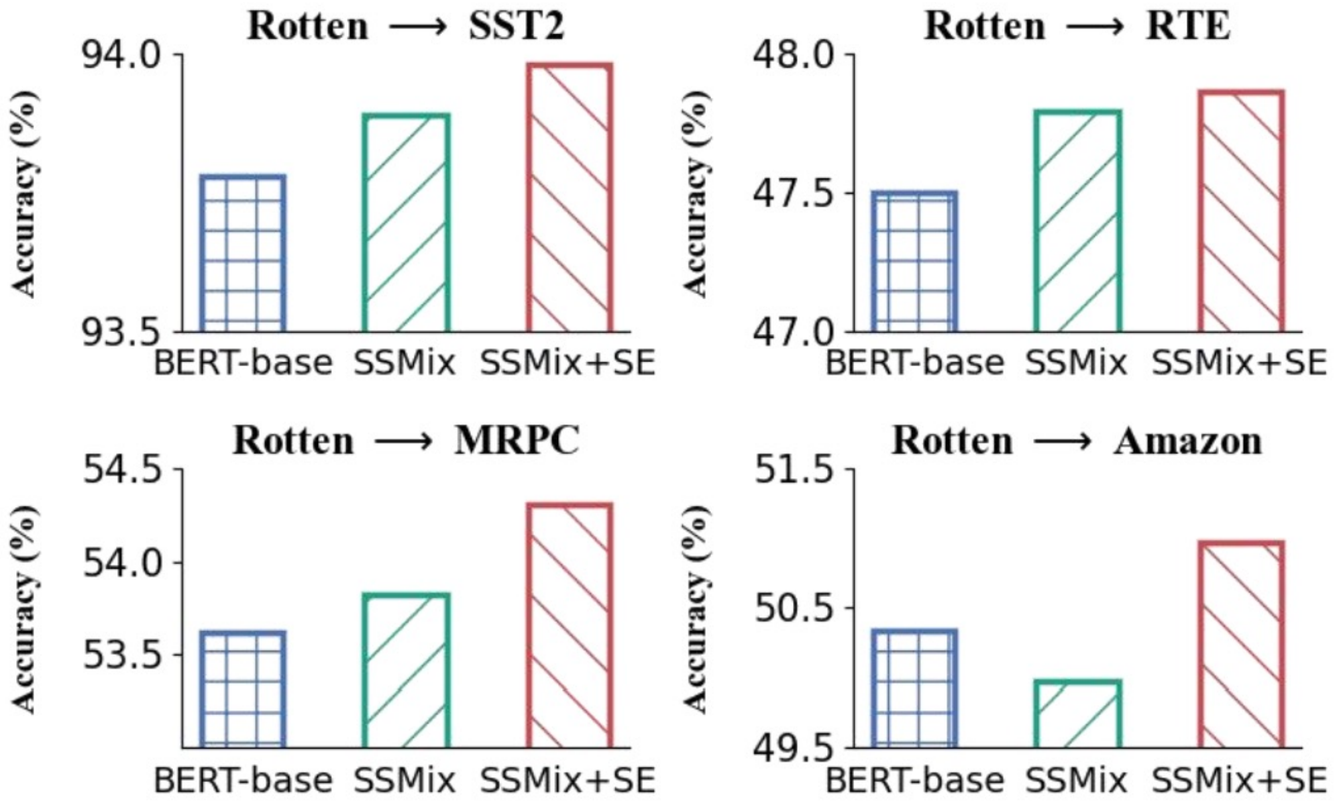
Experiment

Results at various training data sampling rates



Experiment

Analysis of task generalization



Outline

- Introduction
- Method
- Experience
- Conclusion

Conclusion

- Propose a self-evolution (SE) learning mechanism
  - –conducting data division based on the degree of difficulty
  - –mixup based on the order from easy to hard
  - –Instance-specific label smoothing approach
- Improve the existing mixup methods on text classification tasks