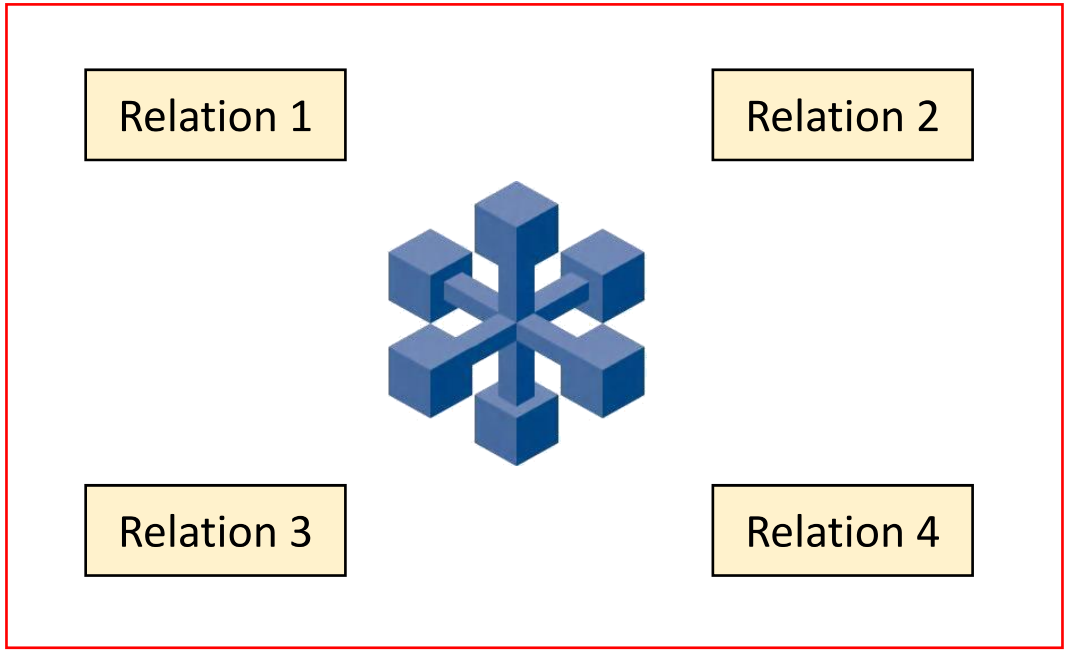
DP-CRE: Continual Relation Extraction via Decoupled Contrastive Learning and Memory Structure Preservation

Mengyi Huang[1,2,†], Meng Xiao[1,†], Ludi Wang[1,*], Yi Du[1,2,3,*]

[1]Computer Network Information Center, Chinese Academy of Sciences

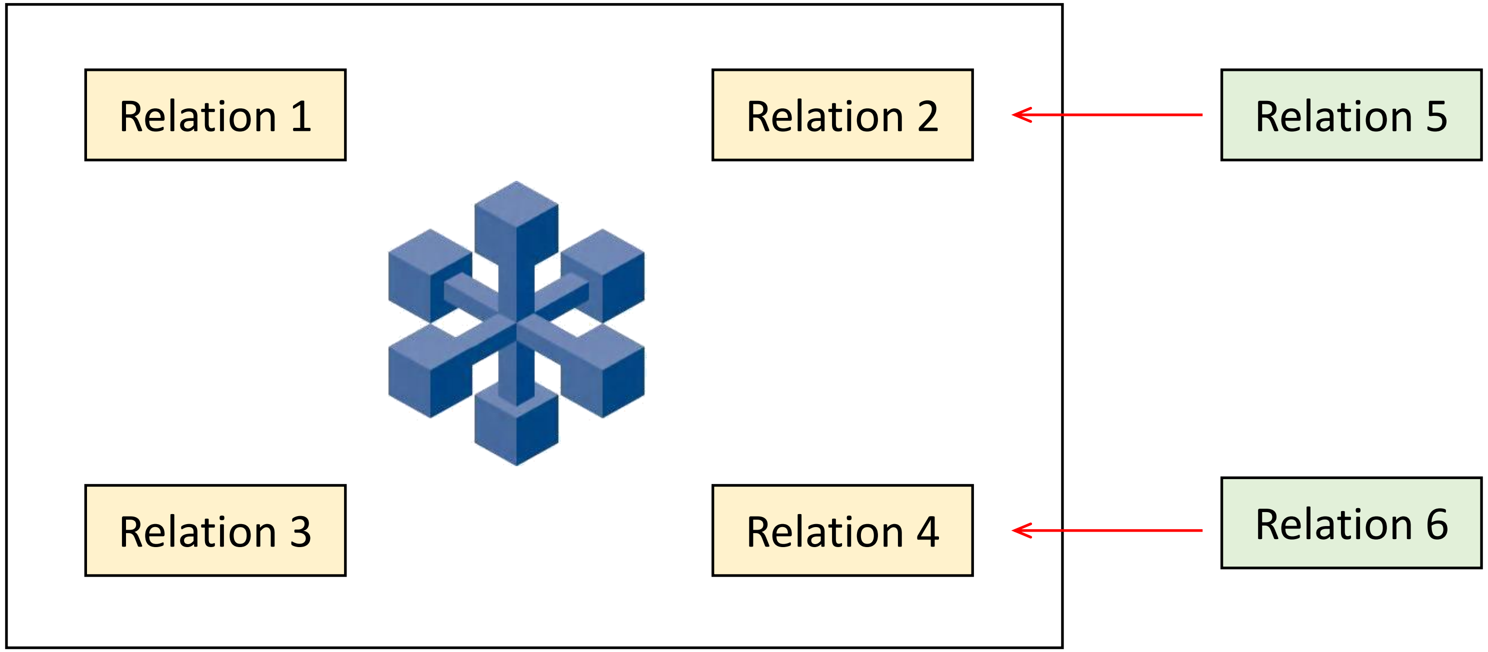
[2]University of Chinese Academy of Sciences

[3]Hangzhou Institute for Advanced Study, UCAS



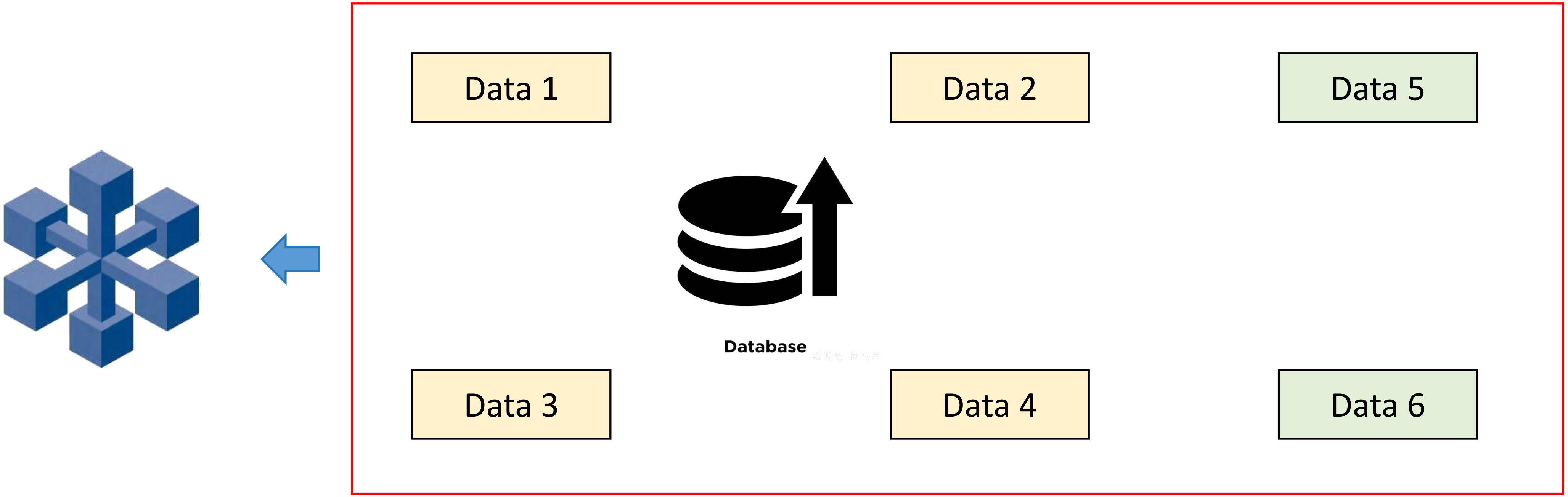
Trained model for Relation 1&2&3&4

01 Target and Motivation



Expand the model Relation 1&2&3&4 and 5&6

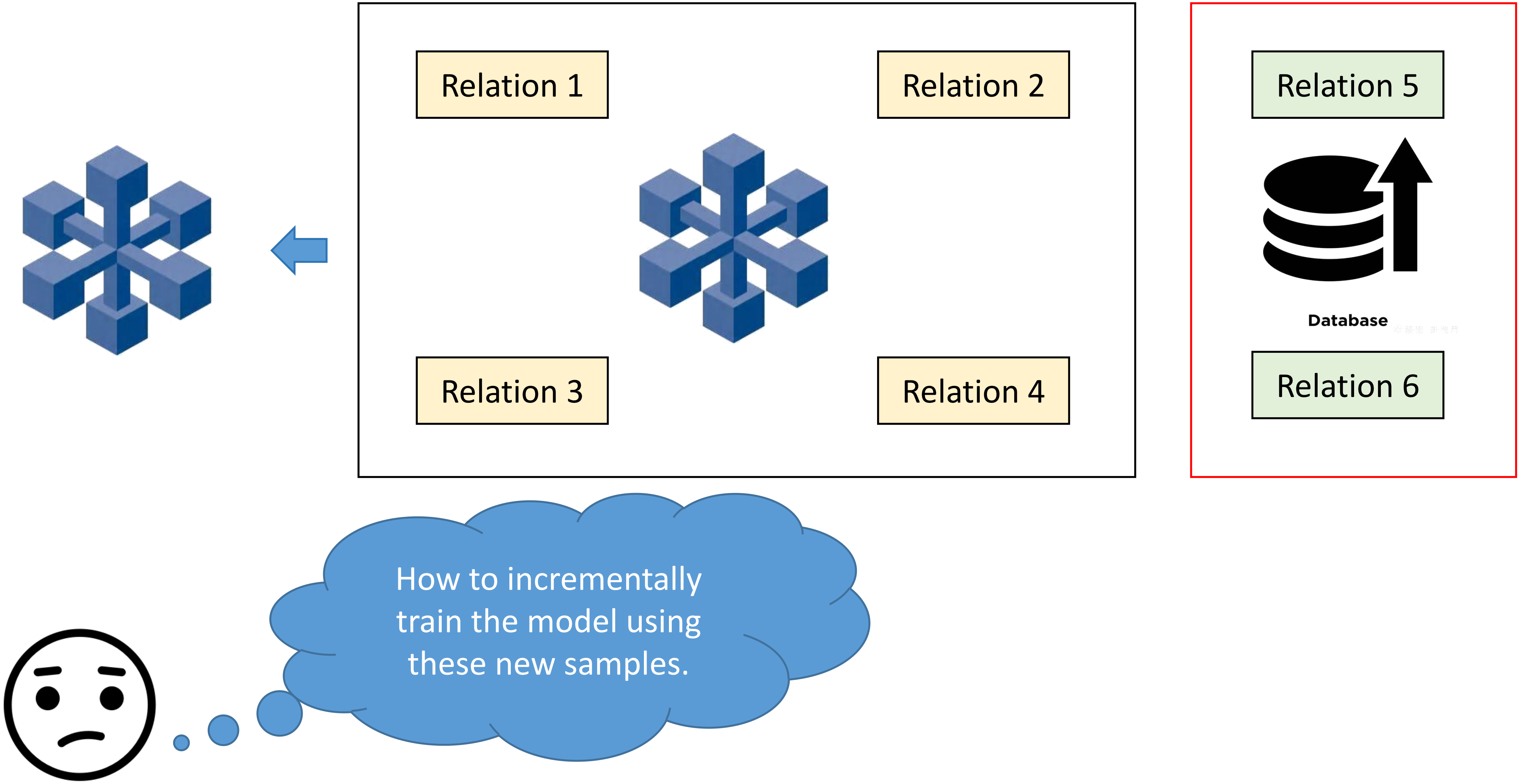
01 Target and Motivation



Retraining the model

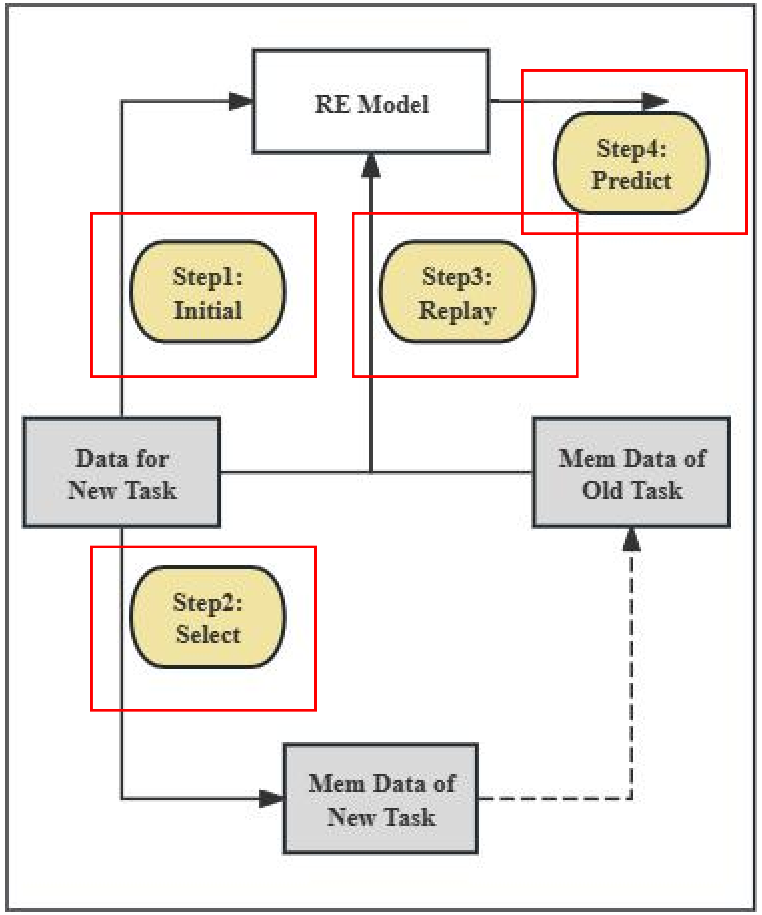
Constraints in **storage** and **computational** resources

01 Target and Motivation



Summary of Existing Approaches

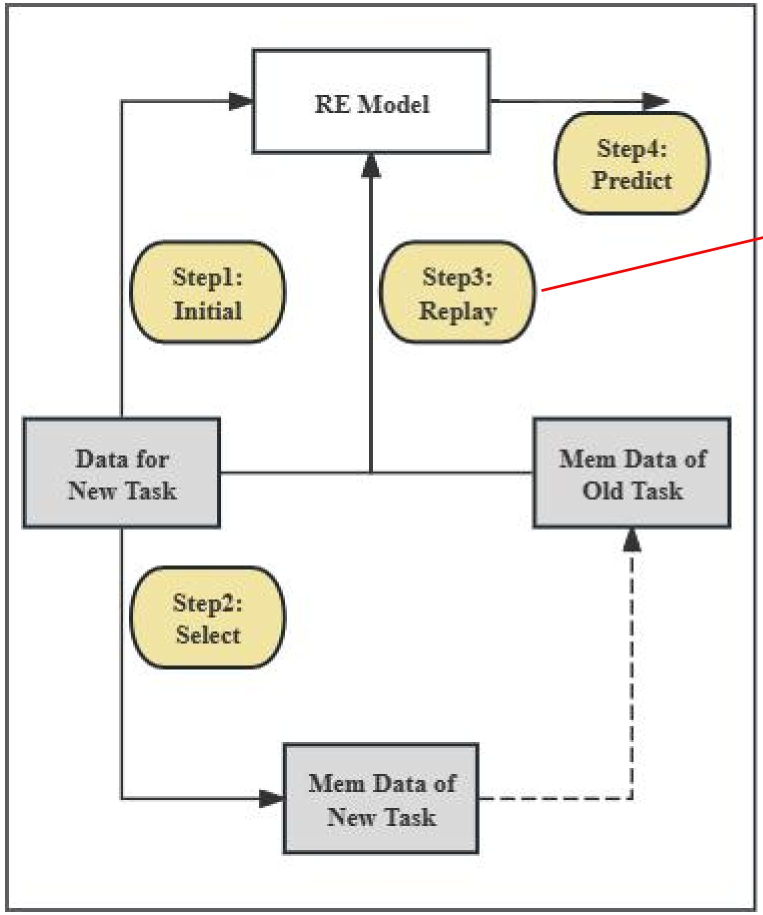
The Existed Approaches of Continual Relation Extraction



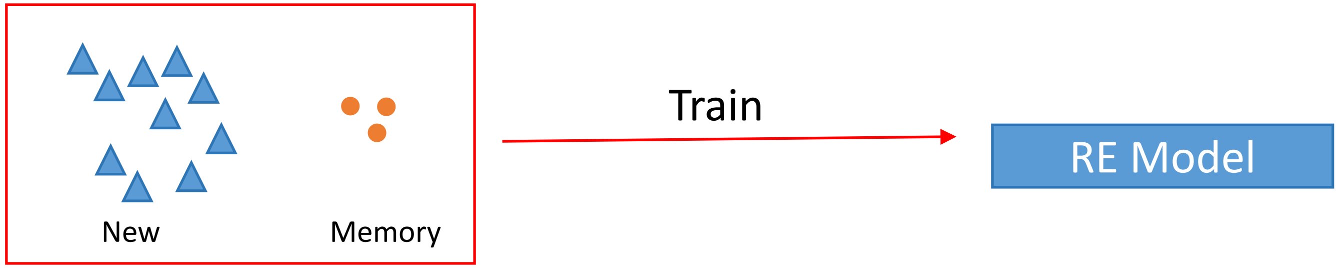
The **Memory-Based method** is widely used in current Continual Relation Extraction work.

02 Summary of Existing Approaches

The Existed Approaches of Continual Relation Extraction



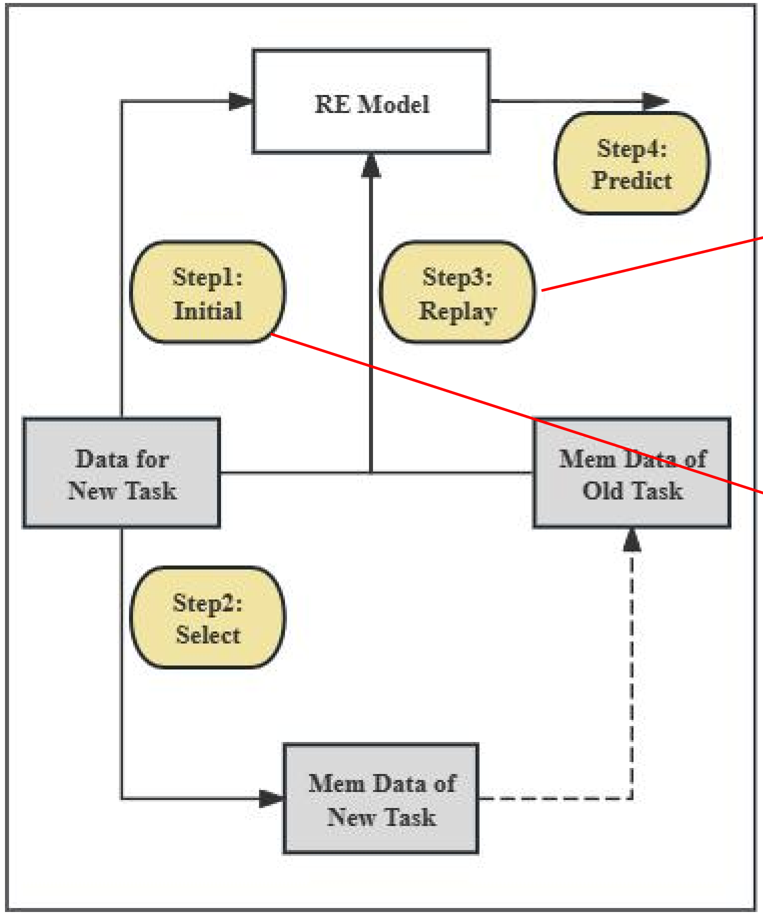
The **Memory-Based method** is widely used in current Continual Relation Extraction work.



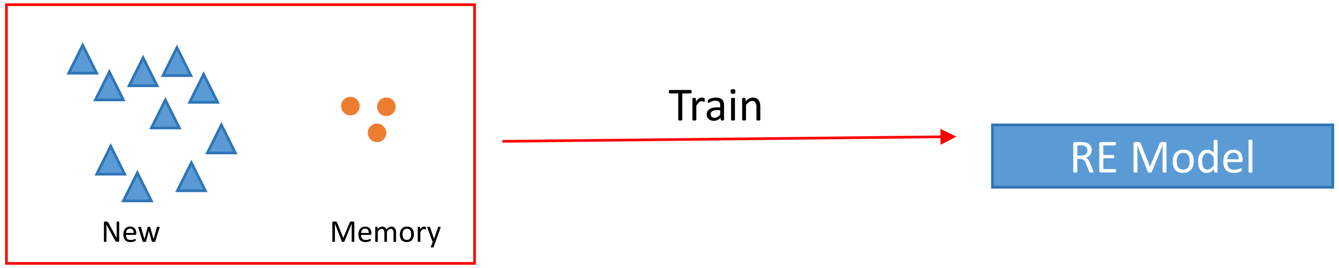
Memory & New Samples Mixed Together

02 Summary of Existing Approaches

The Existed Approaches of Continual Relation Extraction



The Memory-Based method is widely used in current Continual Relation Extraction work.

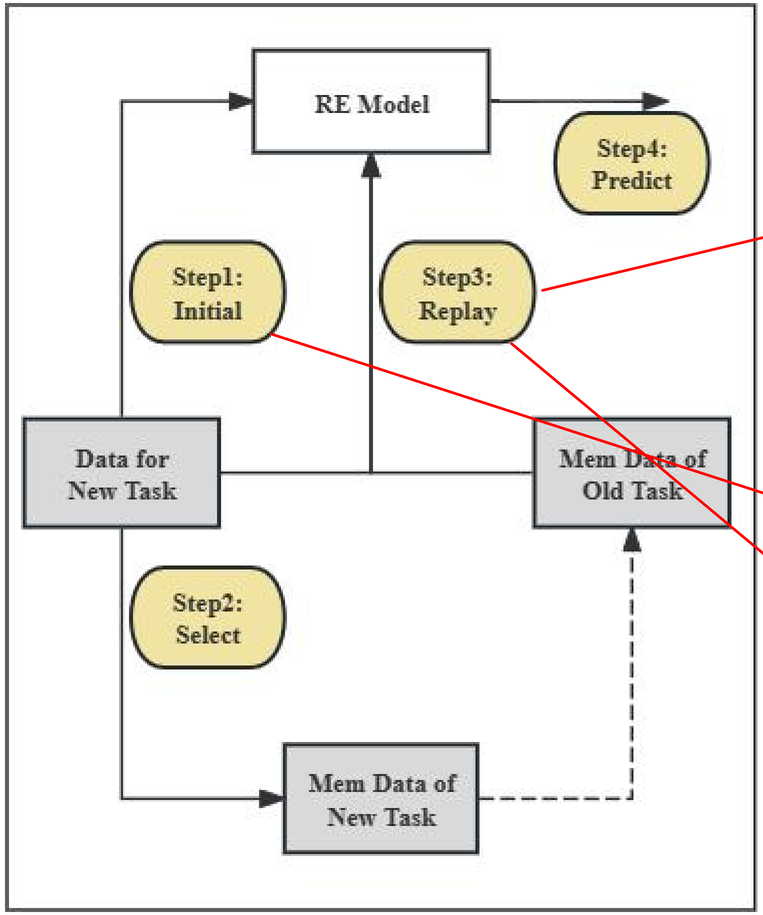


Memory & New Samples Mixed Together

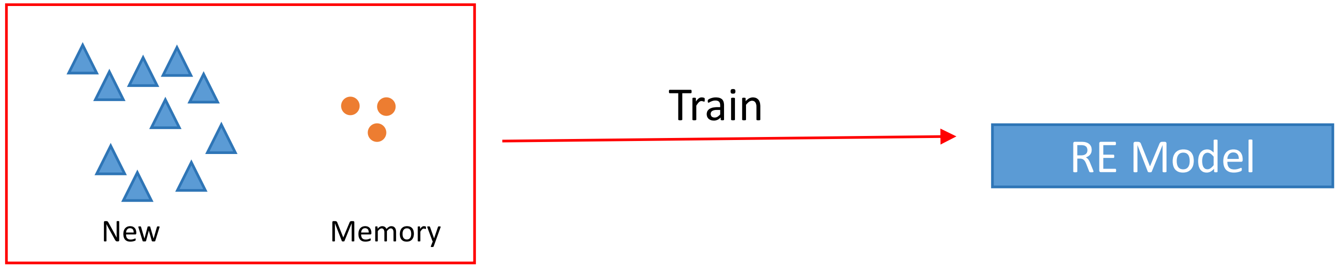
1 **New task** knowledge (Frozen Model Parameters & Class Augmentation)

02 Summary of Existing Approaches

The Existed Approaches of Continual Relation Extraction



The Memory-Based method is widely used in current Continual Relation Extraction work.



Memory & New Samples Mixed Together

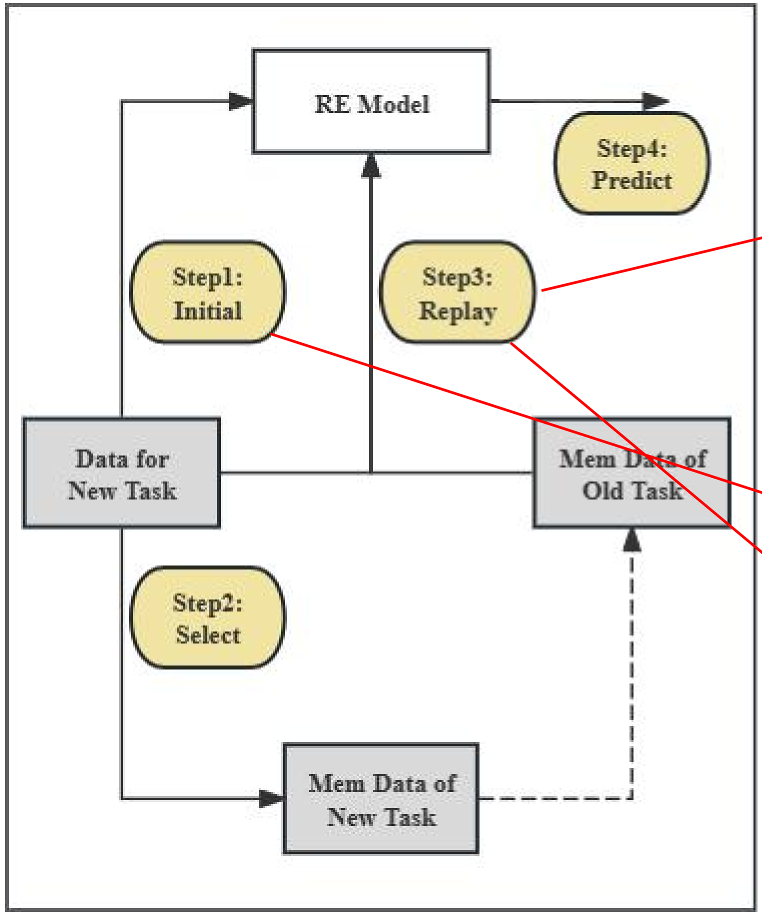
1 **New task** knowledge (Frozen Model Parameters & Class Augmentation)

2 **Replay process** (Contrastive Learning & Data Augmentation)

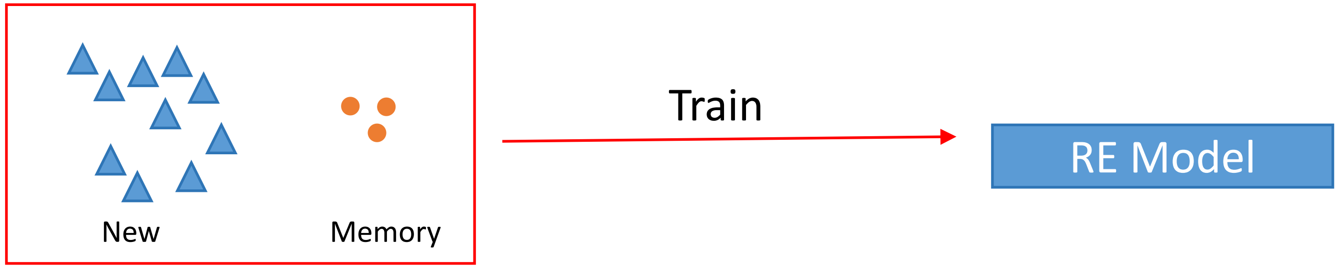
3 **Memory embedding** to the same places

02 Summary of Existing Approaches

The Existed Approaches of Continual Relation Extraction



The Memory-Based method is widely used in current Continual Relation Extraction work.



Memory & New Samples Mixed Together

1 **New task** knowledge (Frozen Model Parameters & Class Augmentation)

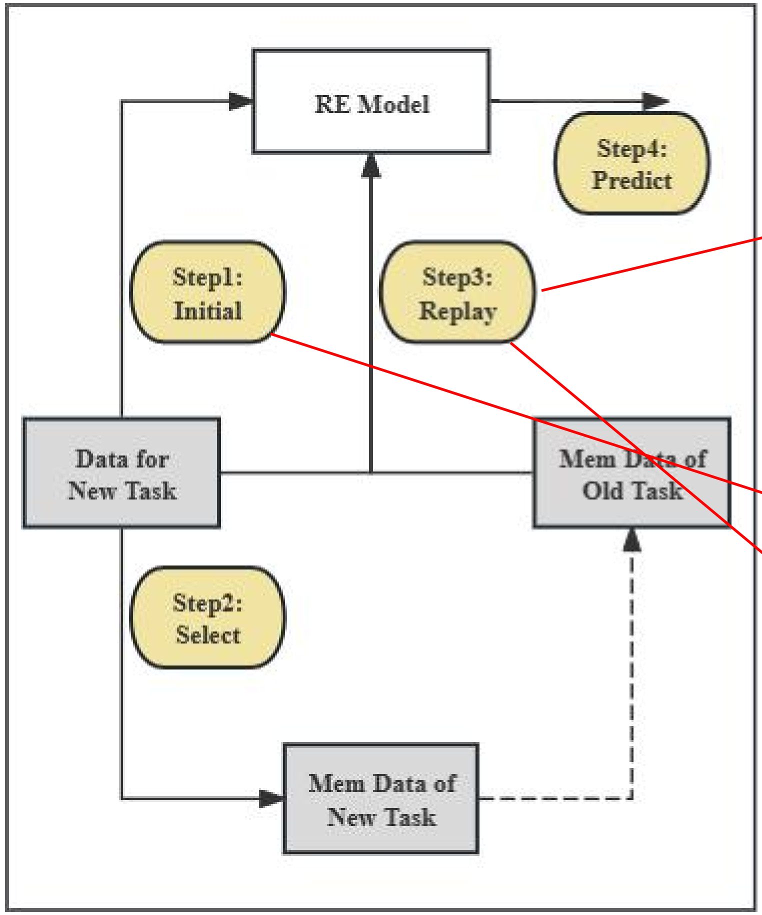
2 **Replay process** (Contrastive Learning & Data Augmentation)

3 **Memory embedding** to the same places

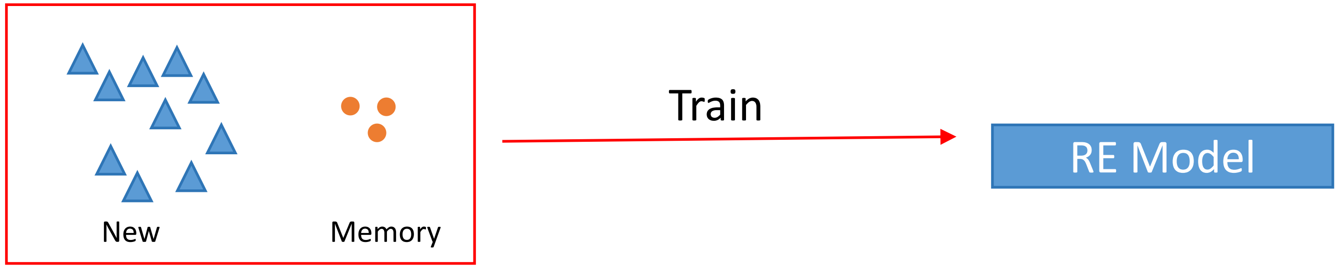
Train Memory & New samples with the same status would bring **model bias**.

02 Summary of Existing Approaches

The Existed Approaches of Continual Relation Extraction



The Memory-Based method is widely used in current Continual Relation Extraction work.



Memory & New Samples Mixed Together

1 **New task** knowledge (Frozen Model Parameters & Class Augmentation)

2 **Replay process** (Contrastive Learning & Data Augmentation)

3 **Memory embedding** to the same places

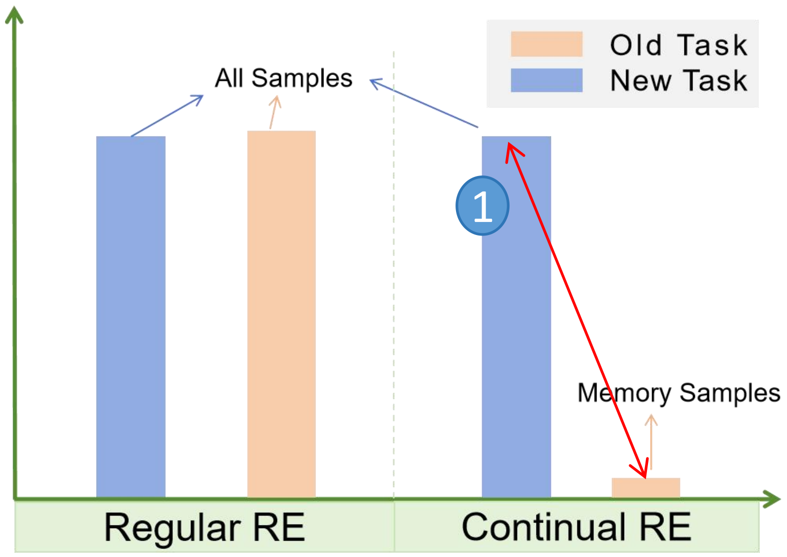
Train Memory & New samples with the same status would bring **model bias**.

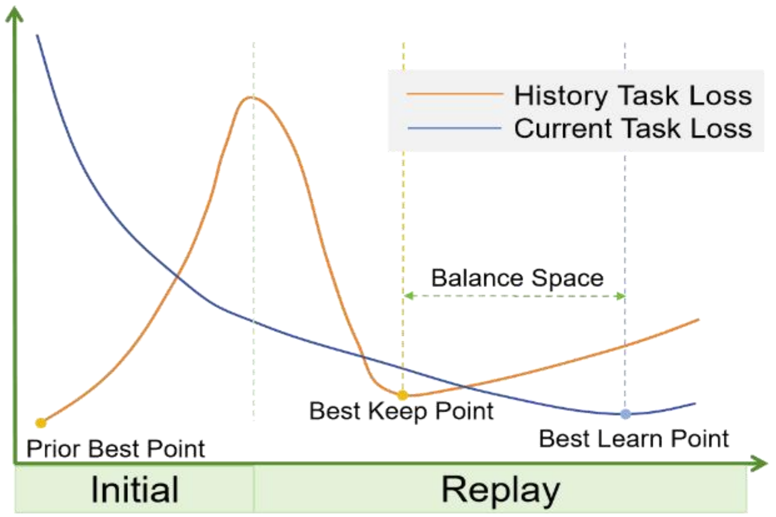
Strict Limitation may **limit the ability** to learn new relations.

03 Discussion the Decoupled Tasks

Data Bias Brings Model Bias.

1 Imbalance in old and new data





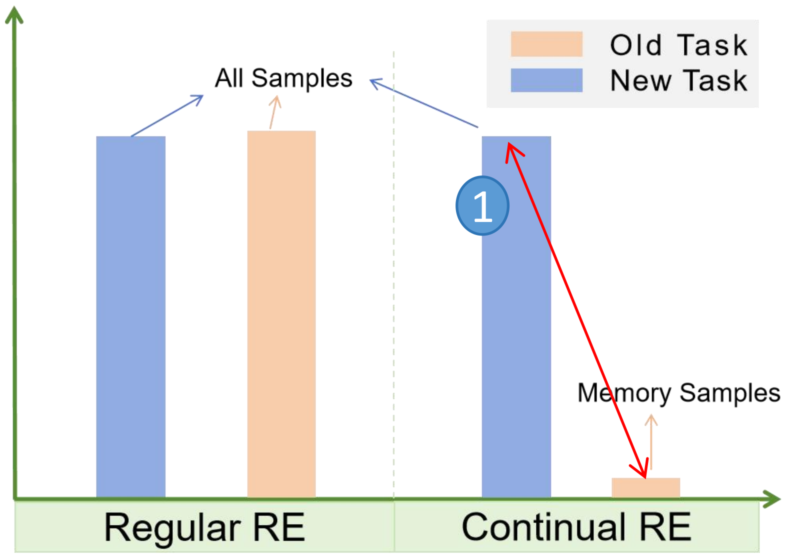
03 Discussion the Decoupled Tasks

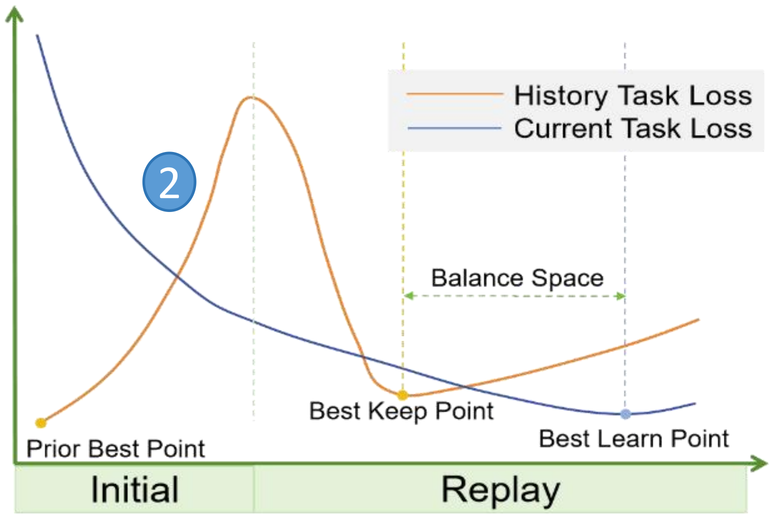
Data Bias Brings Model Bias.

1. Imbalance in old and new data

A unified Task or Decoupled Tasks?

2. One task can influence the other.





03 Discussion the Decoupled Tasks

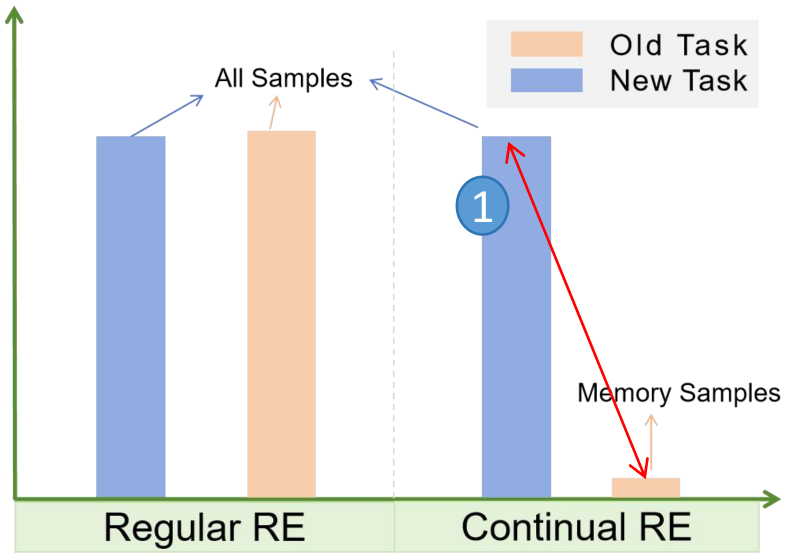
Data Bias Brings Model Bias.

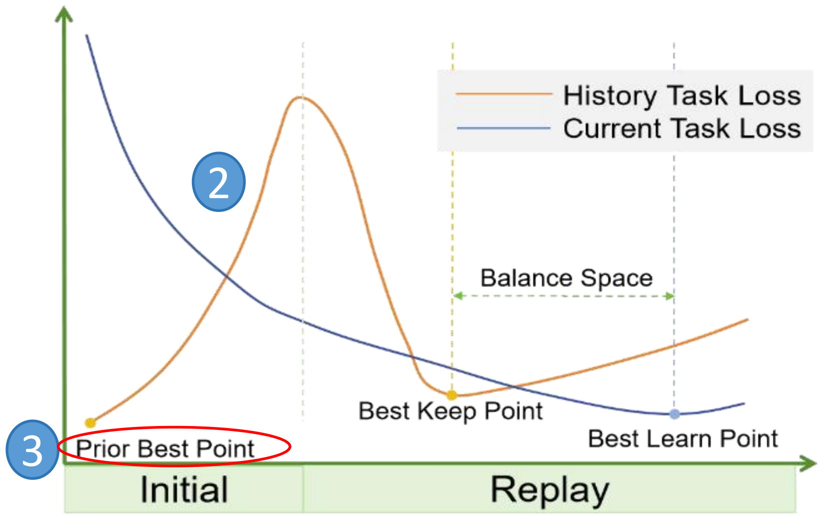
1. Imbalance in old and new data

A unified Task or Decoupled Tasks?

2. One task can influence the other.

3. Historical tasks BEST POINT when unaffected by new data types.





03 Discussion the Decoupled Tasks

Data Bias Brings Model Bias.

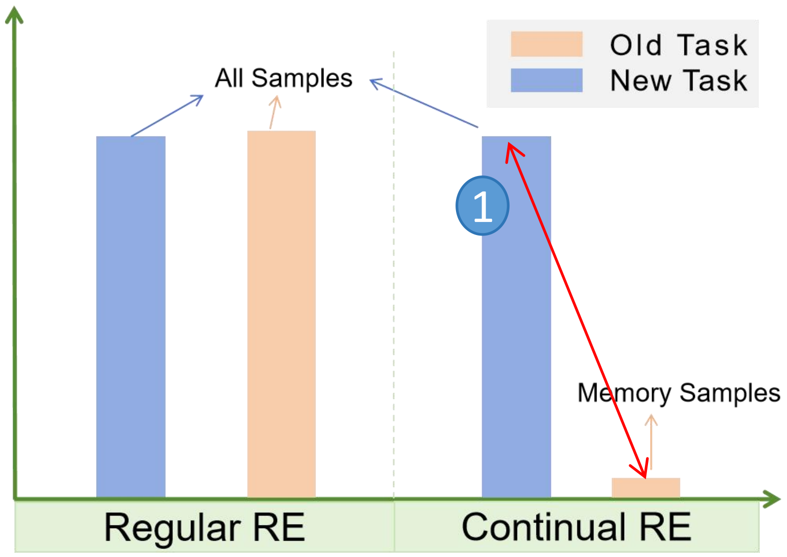
1 Imbalance in old and new data

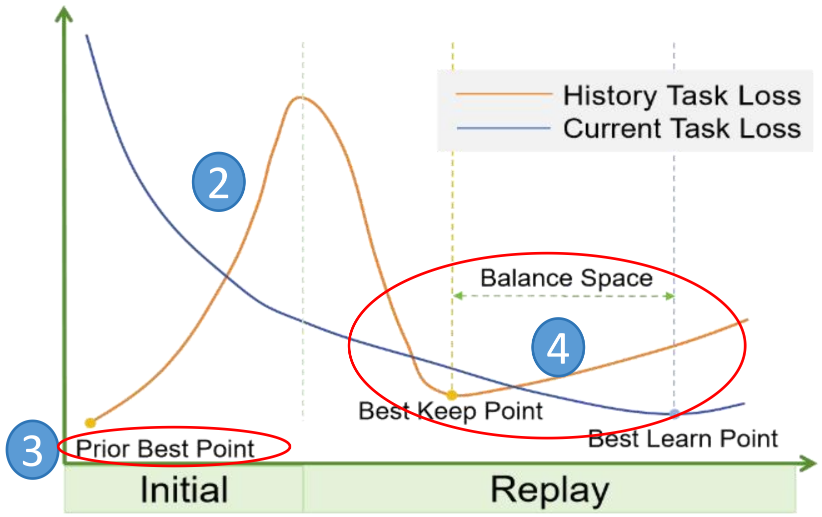
A unified Task or Decoupled Tasks?

2 One task can influence the other.

3 Historical tasks BEST POINT when unaffected by new data types.

4 Balance Space of Best Keep and Best learn because of imbalances in the replay set.





04 Our Contributions

Our DecouPled Contributions for Continual Relation Extraction:

[Balancing CRE with Multi-task Learning]

Prior Information Preservation and New Knowledge Acquisition

[Decoupling to Mitigate Overfitting]

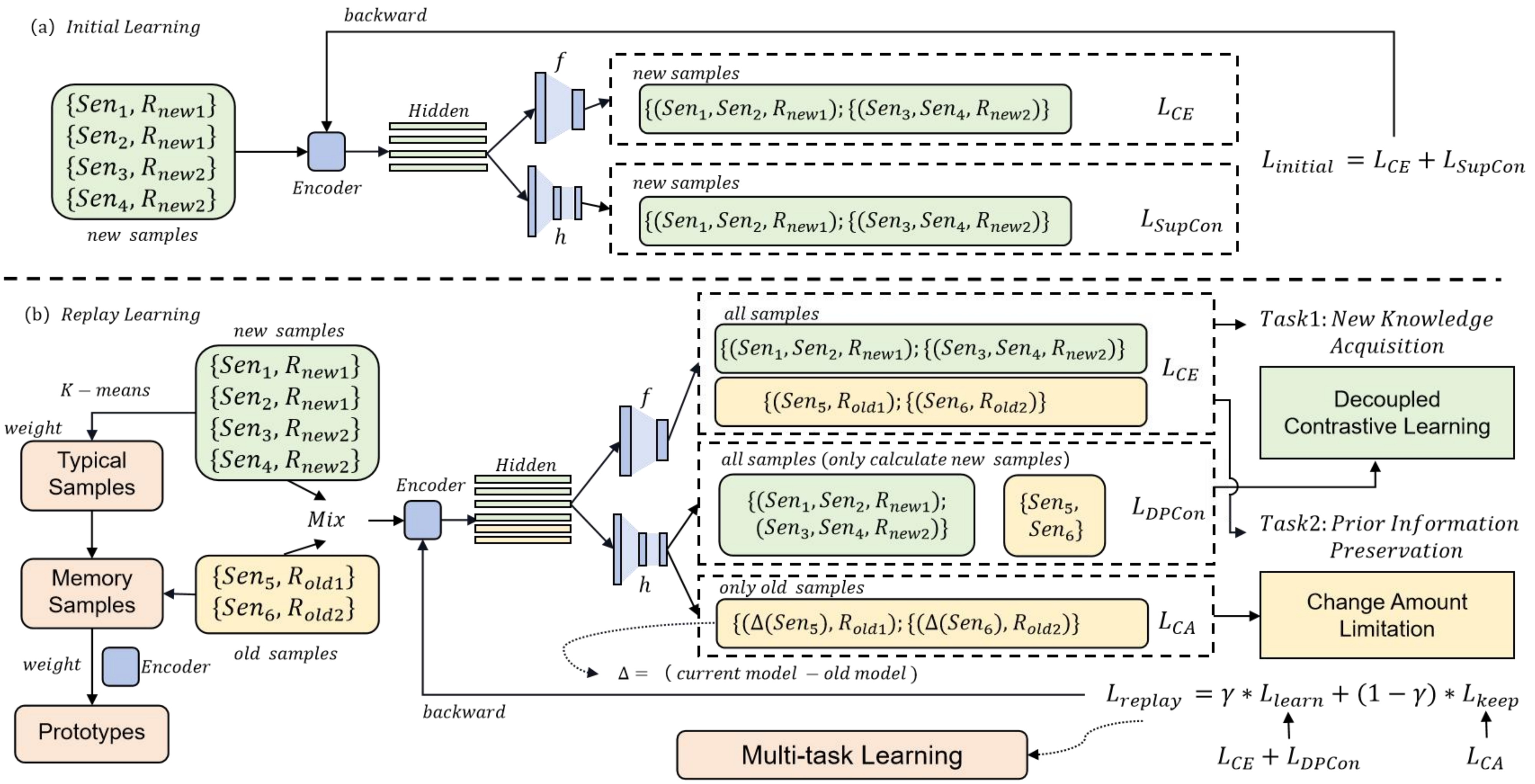
Conserve the memory structural information

[Empirical Validation of DP-CRE]

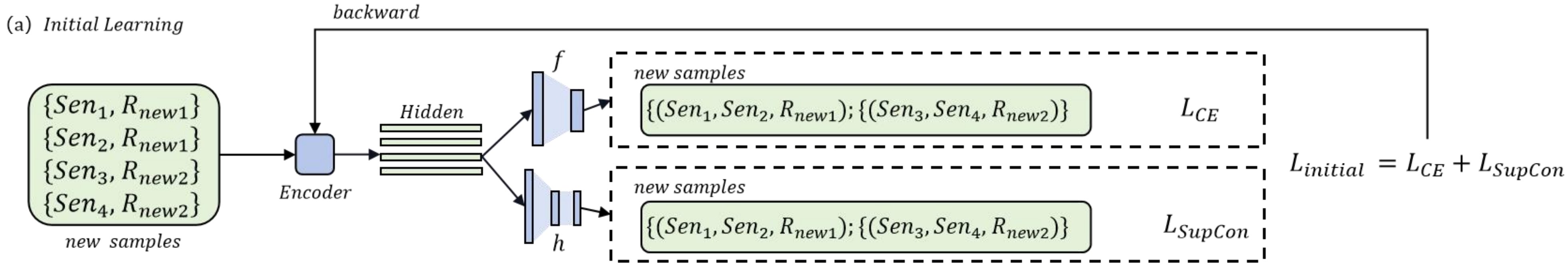
Experiment results demonstrate the SOTA accuracy.

05 Methodology of RAFT

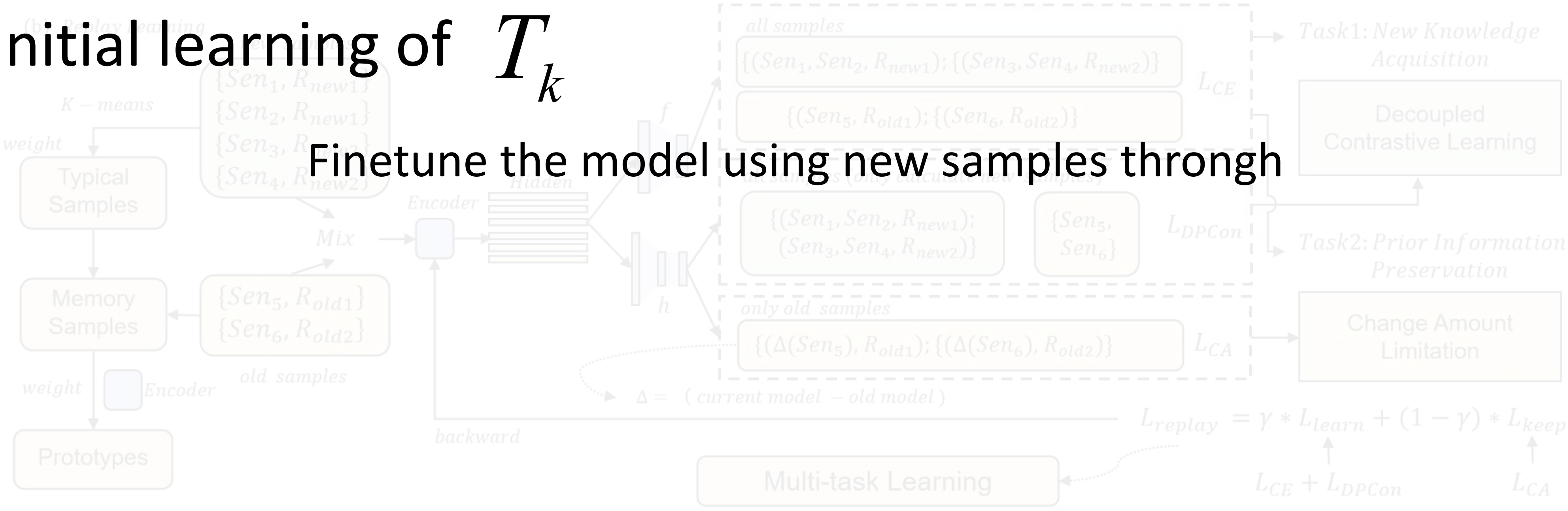
DecouPled Continual Relation Extraction



DP CRE: Initial Learning with New Samples



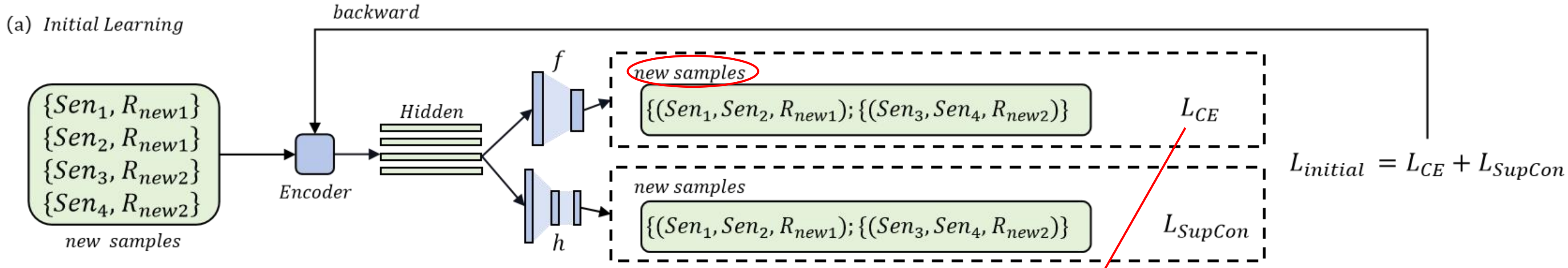
Initial learning of T_k



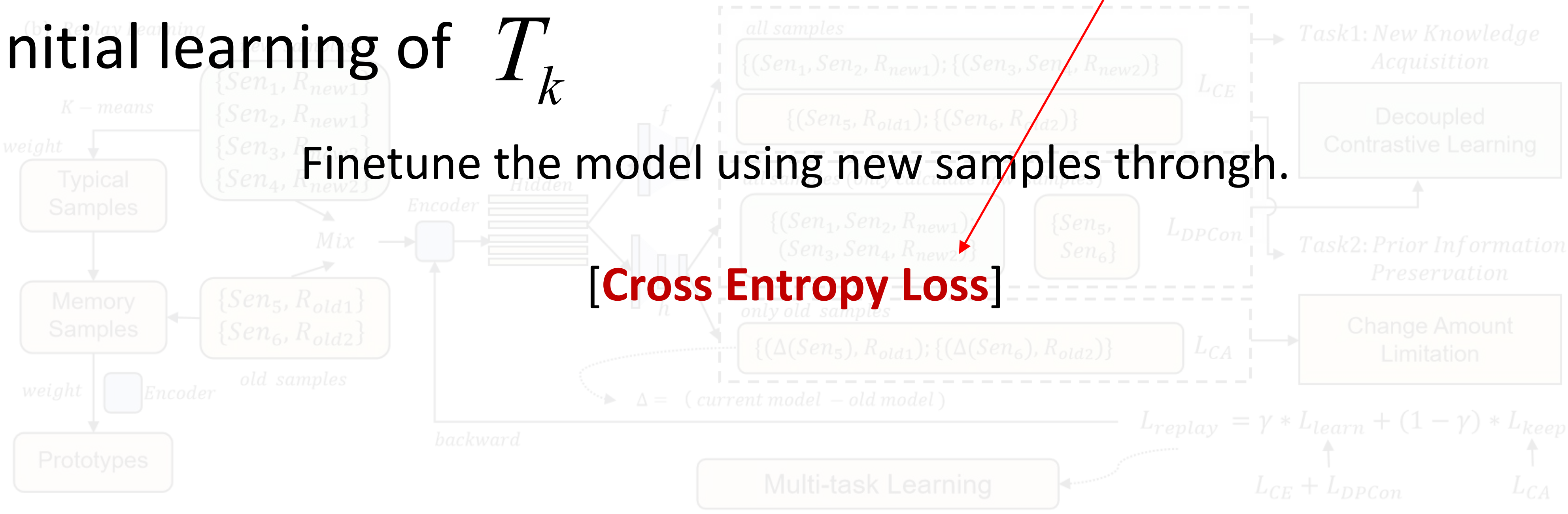
Finetune the model using new samples throngh

05 Methodology of DP-CRE

DP CRE: Initial Learning with New Samples



Initial learning of T_k

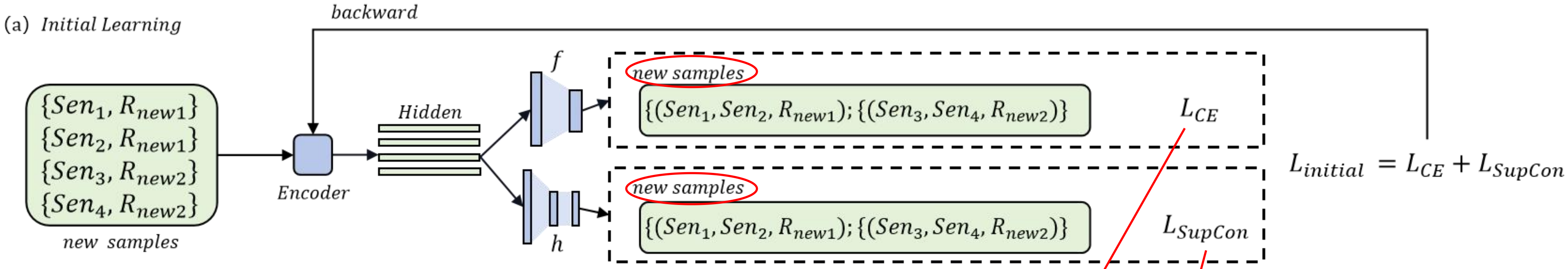


Finetune the model using new samples throngh.

[**Cross Entropy Loss**]

05 Methodology of DP-CRE

DP CRE: Initial Learning with New Samples



Initial learning of T_k

Finetune the model using new samples throngh.

[**Cross Entropy Loss**]

&

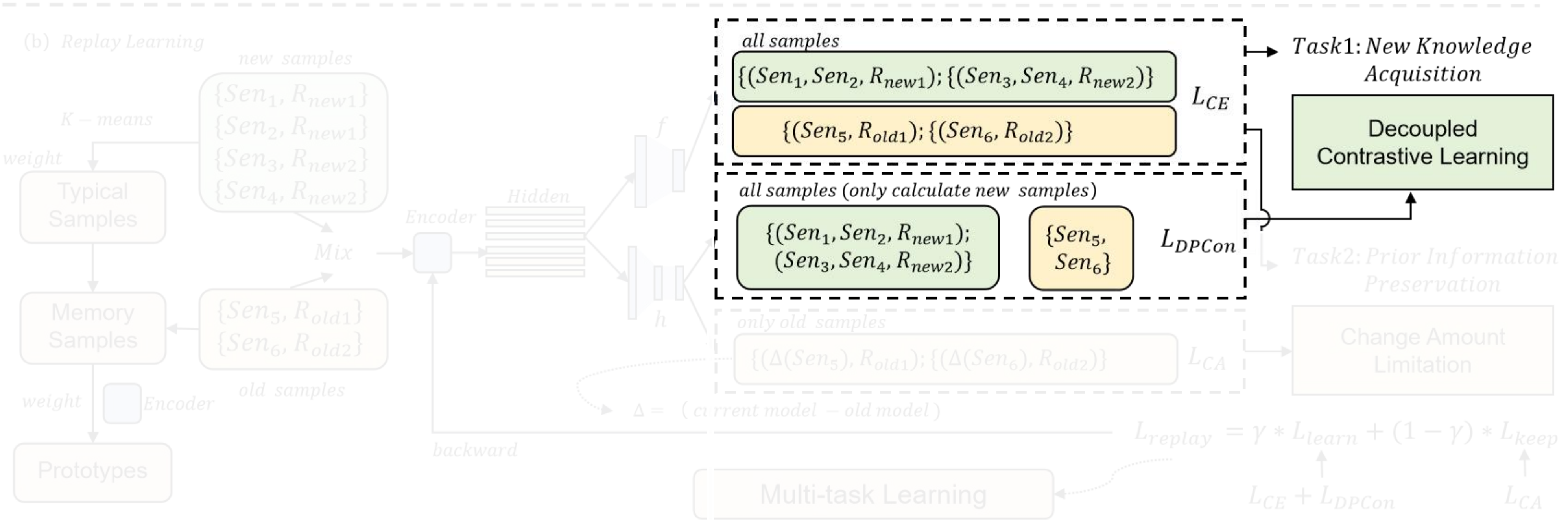
[**Supervised Contrastive Loss**]

05 Methodology of DP-CRE

DP CRE: Decoupled Contrastive Learning to Acquire New Knowledge

Repaly learning of T_k

Task1 : New Knowledge Acquisition



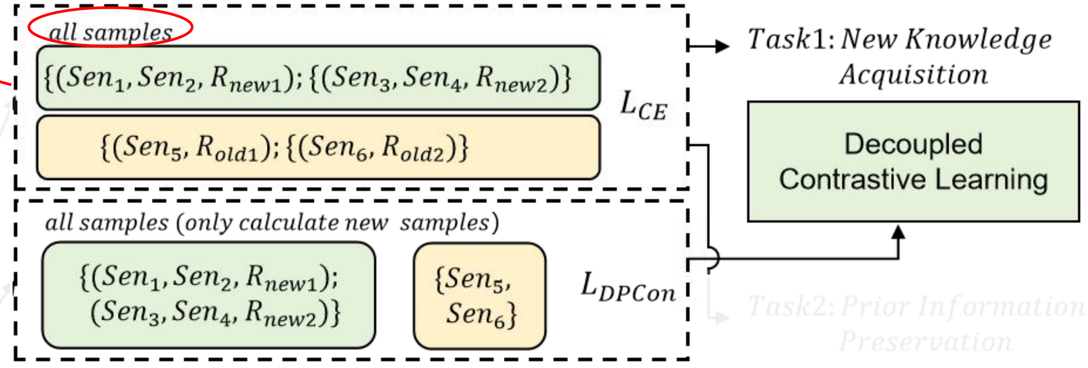
05 Methodology of DP-CRE

DP CRE: Decoupled Contrastive Learning to Acquire New Knowledge

Repaly learning of T_k

Task1 : New Knowledge Acquisition

[**Cross Entropy Loss**]



DP CRE: Decoupled Contrastive Learning to Acquire New Knowledge

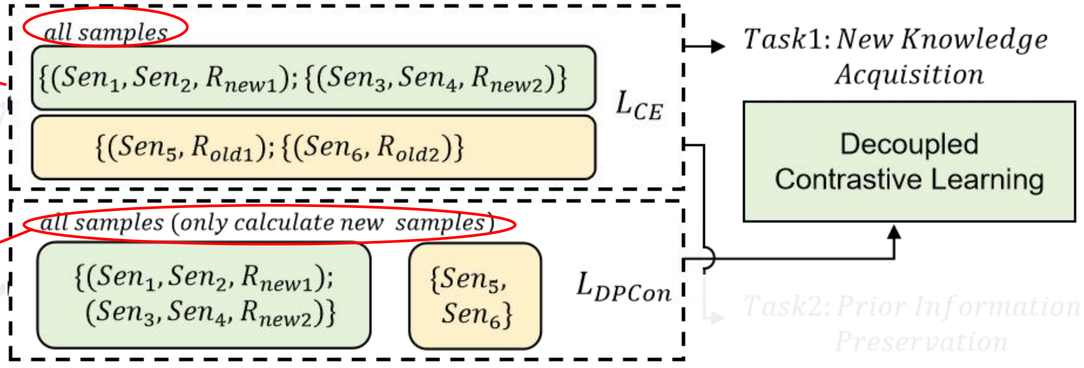
Repaly learning of T_k

Task1 : New Knowledge Acquisition

[Cross Entropy Loss]

&

[Decoupled Contrastive Loss]



$$L_{DPCon} = \sum_{i \in D^k} \frac{-1}{|D^k|} \sum_{j \in D^k} \delta_{y_i = y_j} \times \log \frac{exp\left(h(z_{i,r}) \cdot h(z_{j,r}) / \tau\right)}{\sum_{j \in D_k \cup M} exp\left(h(z_{i,r}) \cdot h(z_{j,r}) / \tau\right)}, \quad (7)$$

DP CRE: Decoupled Contrastive Learning to Acquire New Knowledge

Repaly learning of T_k

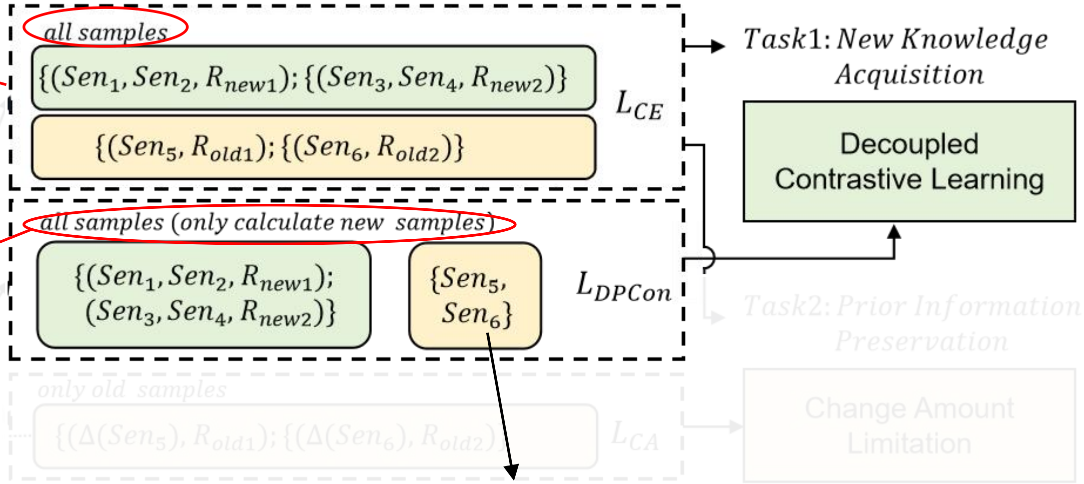
Task1 : New Knowledge Acquisition

[Cross Entropy Loss]

&

[Decoupled Contrastive Loss]

$$L_{DPCon} = \sum_{i \in D^k} \frac{-1}{|D^k|} \sum_{j \in D^k} \delta_{y_i = y_j} \times \log \frac{exp\left(h(z_{i,r}) \cdot h(z_{j,r}) / \tau\right)}{\sum_{j \in D_k \cup M} exp\left(h(z_{i,r}) \cdot h(z_{j,r}) / \tau\right)}, \quad (7)$$



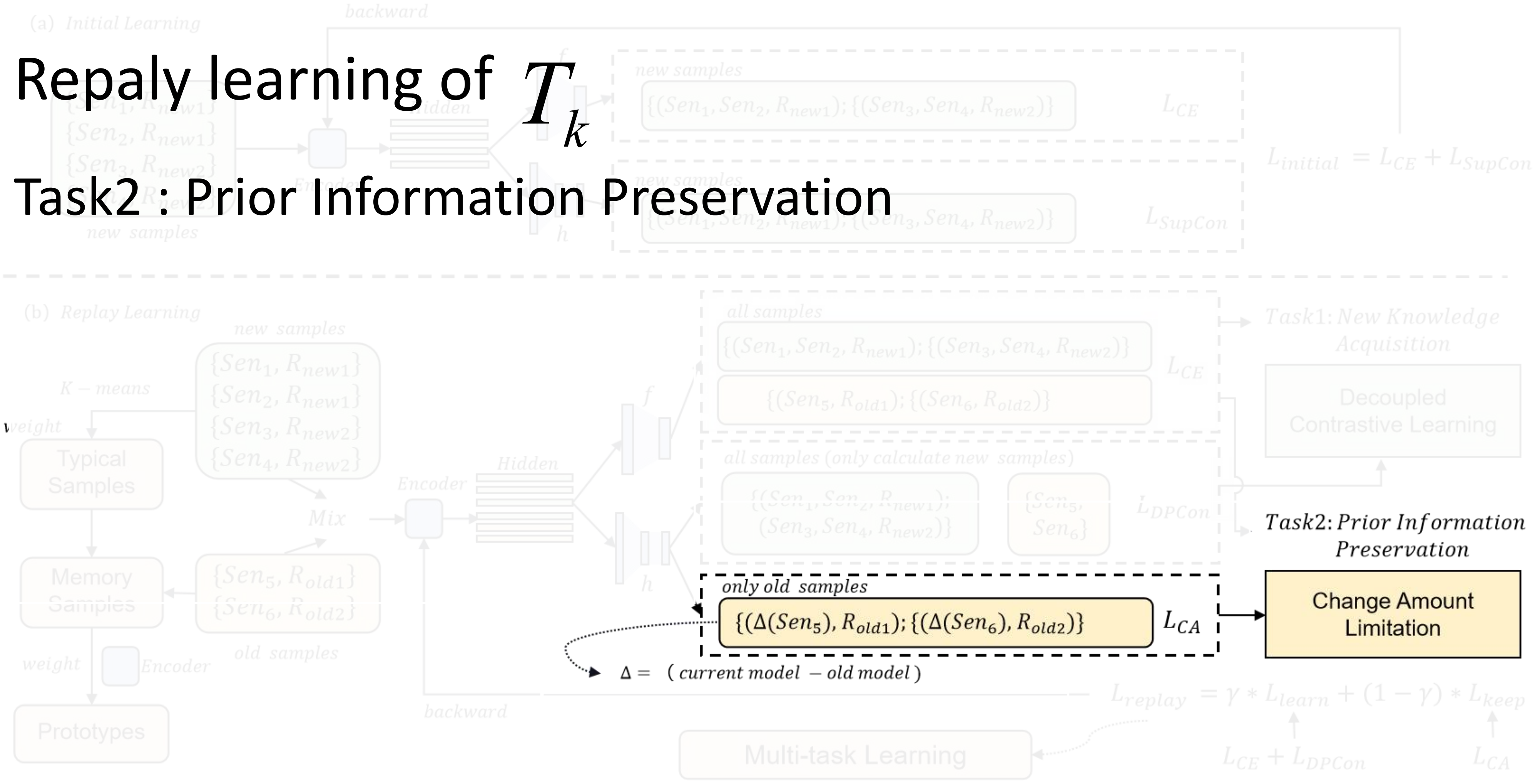
ONLY as Negative Samples

05 Methodology of DP-CRE

DP CRE: Change Amount Limitation to Preserve Prior Information

Repaly learning of T_k

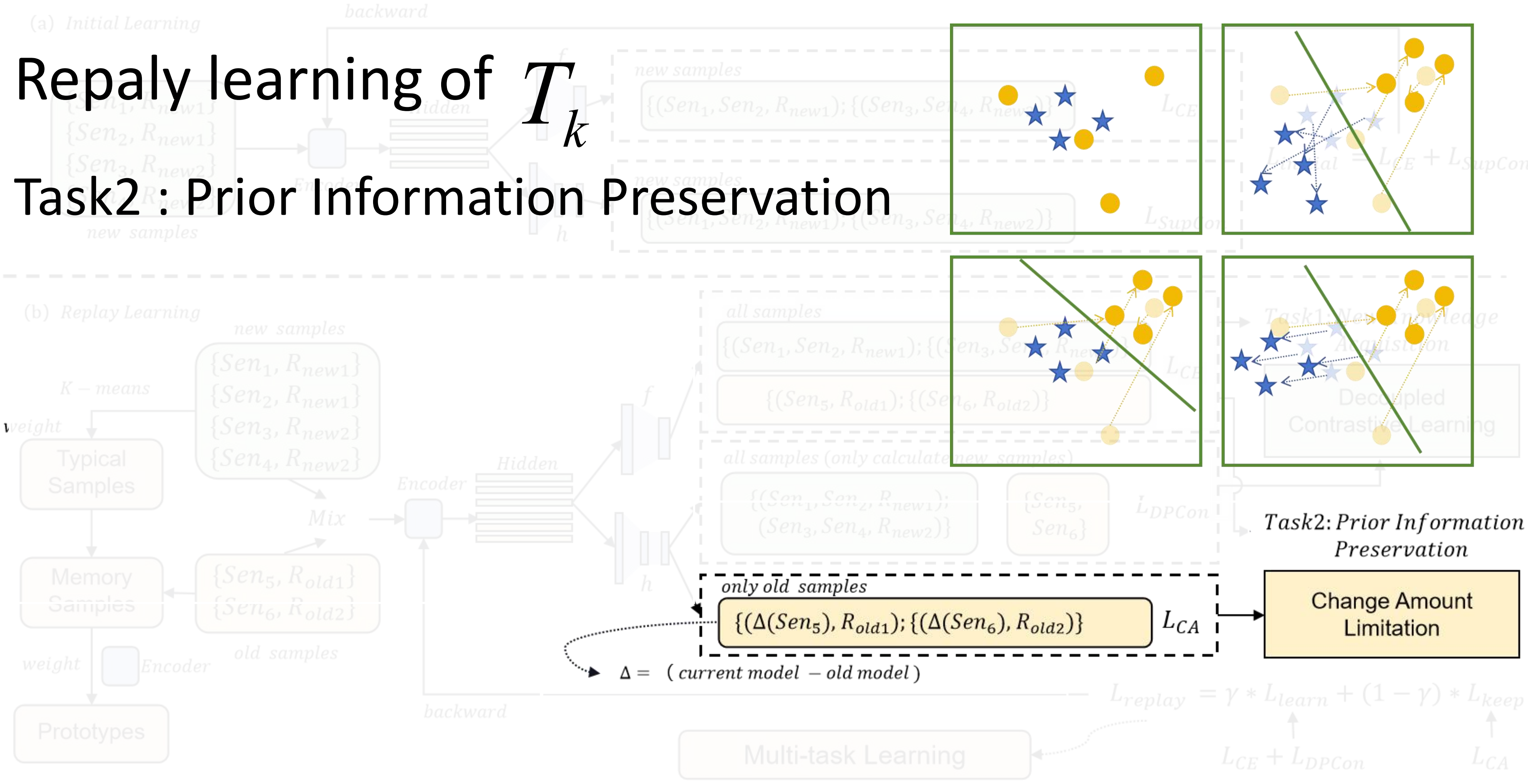
Task2 : Prior Information Preservation



DP CRE: Change Amount Limitation to Preserve Prior Information

Repaly learning of T_k

Task2 : Prior Information Preservation



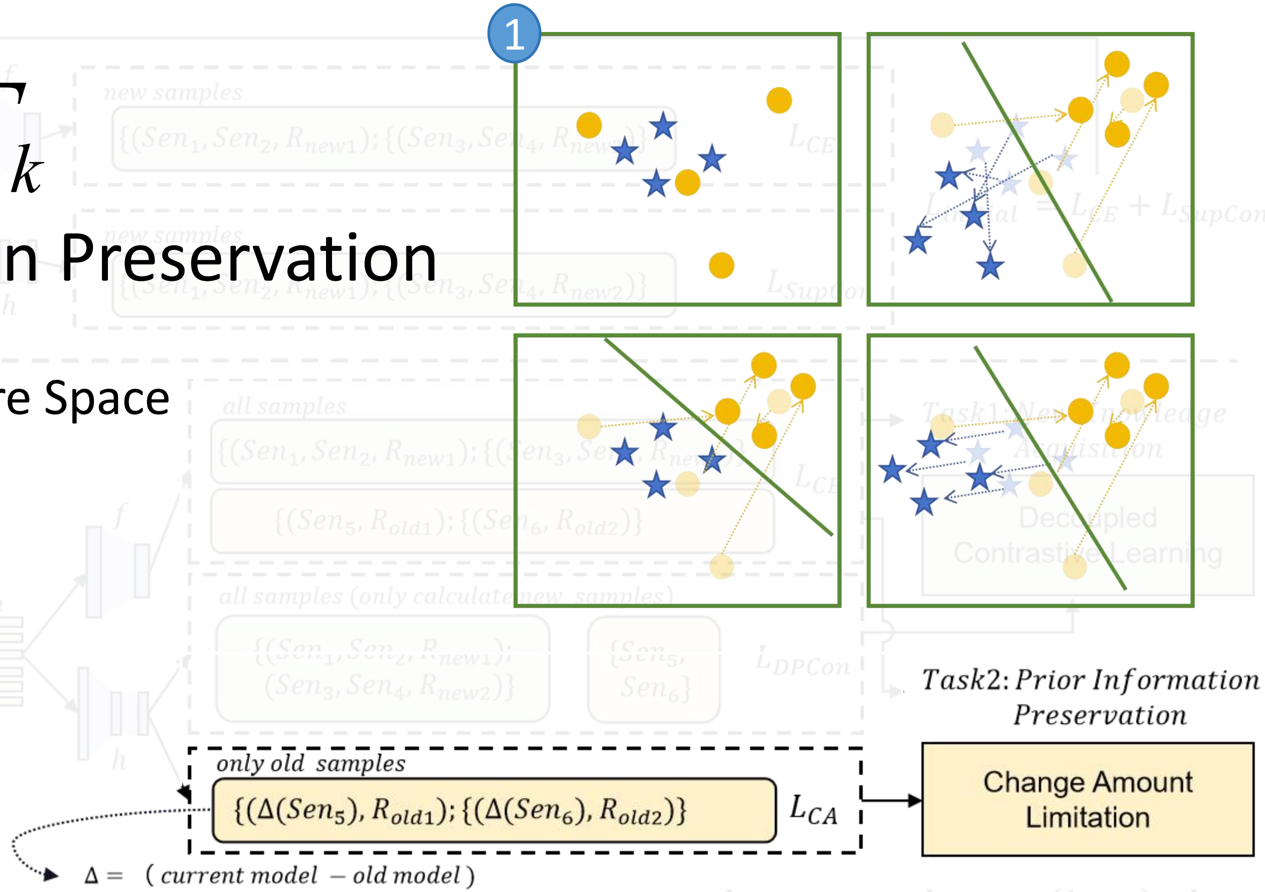
Methodology of DP-CRE

DP CRE: Change Amount Limitation to Preserve Prior Information

Repaly learning of T_k

Task2 : Prior Information Preservation

1 Untrained Samples in Feature Space



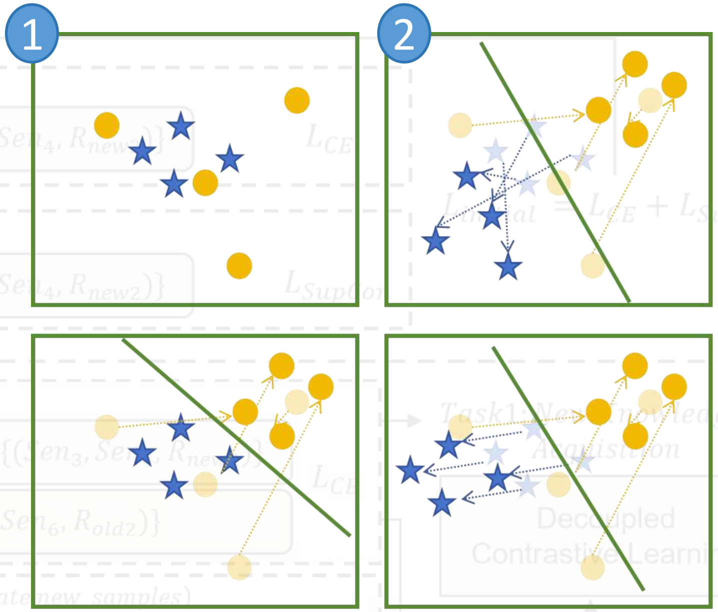
DP CRE: Change Amount Limitation to Preserve Prior Information

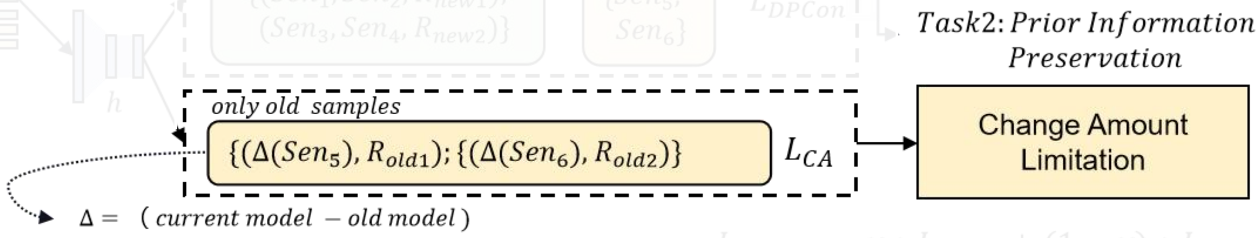
Repaly learning of T_k

Task2 : Prior Information Preservation

1. Untrained Samples in Feature Space

2. Contrastive Learning to All Data





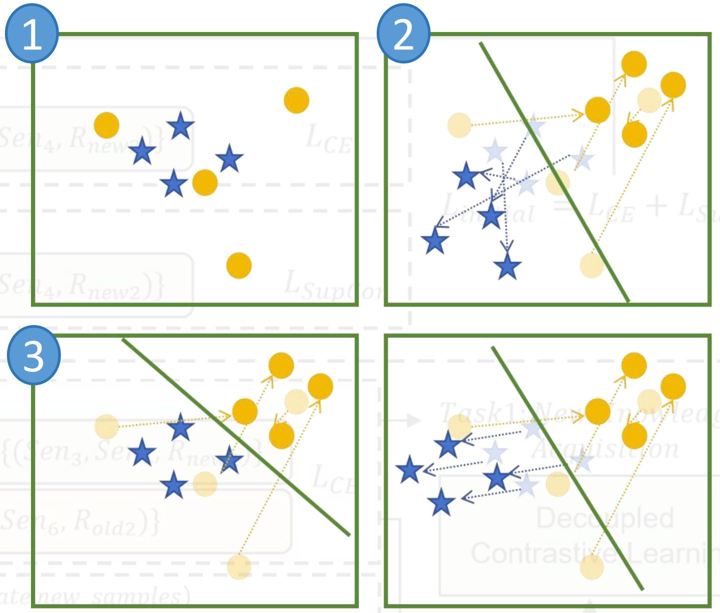
05 Methodology of DP-CRE

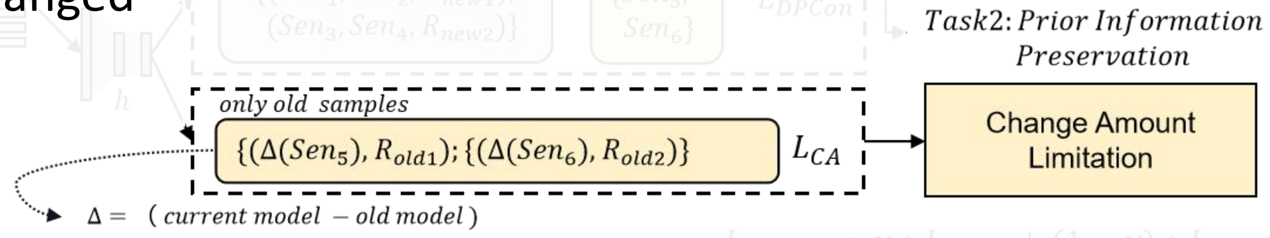
DP CRE: Change Amount Limitation to Preserve Prior Information

Repaly learning of T_k

Task2 : Prior Information Preservation

1. Untrained Samples in Feature Space
2. Contrastive Learning to All Data
3. Retaining Old Samples Unchanged



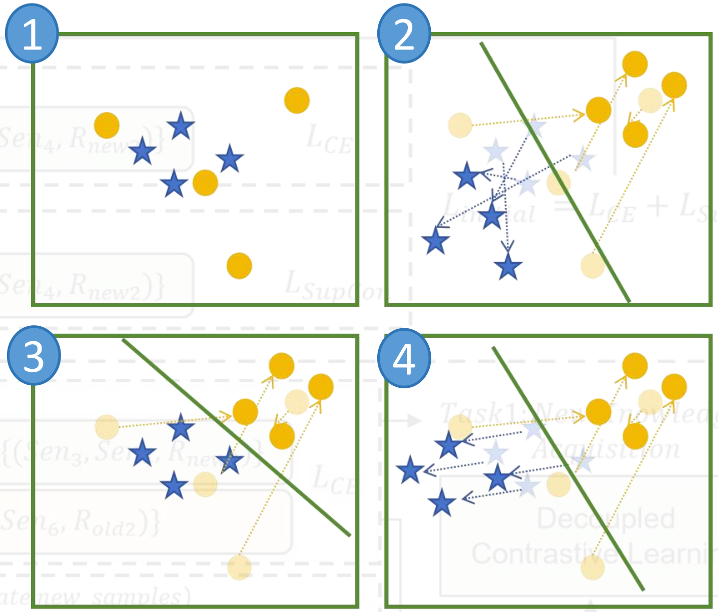


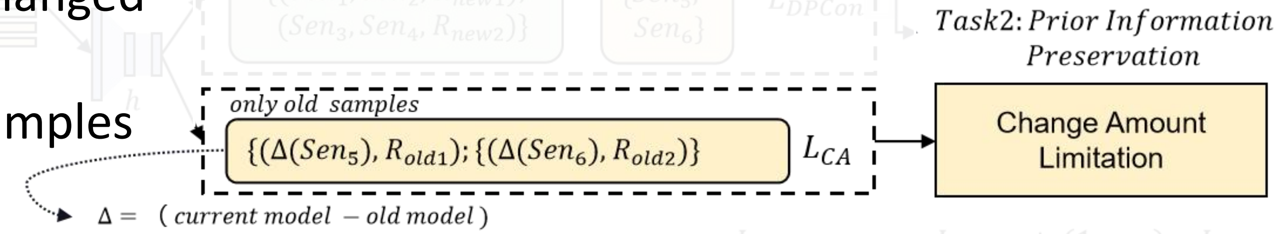
DP CRE: Change Amount Limitation to Preserve Prior Information

Repaly learning of T_k

Task2 : Prior Information Preservation

1. Untrained Samples in Feature Space
2. Contrastive Learning to All Data
3. Retaining Old Samples Unchanged
4. Decoupling Old and New Samples





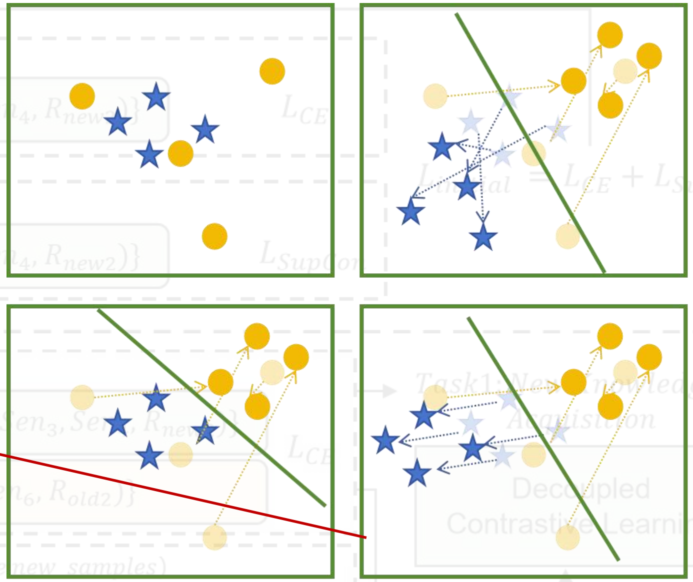
05 Methodology of DP-CRE

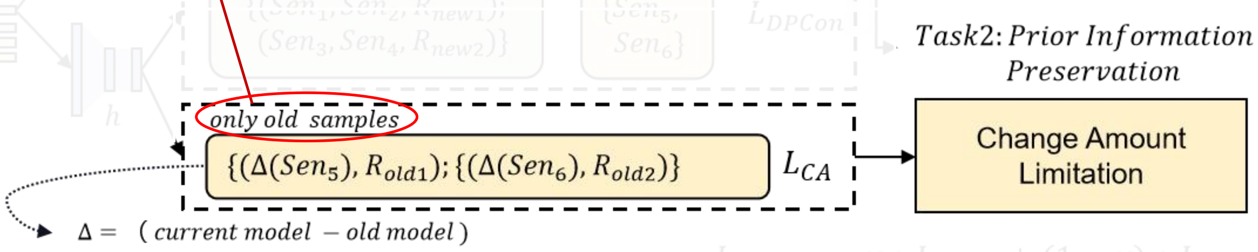
DP CRE: Change Amount Limitation to Preserve Prior Information

Repaly learning of T_k

Task2 : Prior Information Preservation

[Change Amount Limitation Loss]



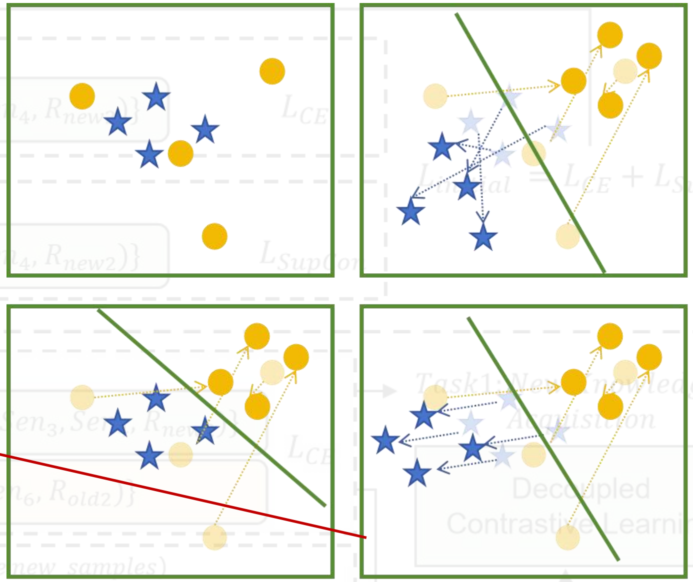


05 Methodology of DP-CRE

DP CRE: Change Amount Limitation to Preserve Prior Information

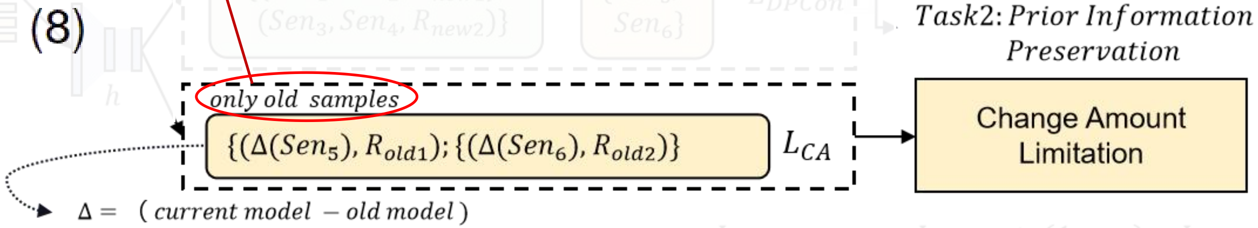
Repaly learning of T_k

Task2 : Prior Information Preservation



[Change Amount Limitation Loss]

$$L_{CA} = \sum_{i,j \in M} \frac{1}{|M|} \delta_{y_i = y_j} \times \| (h^k(z_{i,r}) - h^{k-1}(z_{i,r}^{k-1})) - (h^k(z_{j,r}) - h^{k-1}(z_{j,r}^{k-1})) \|_2, \quad (8)$$

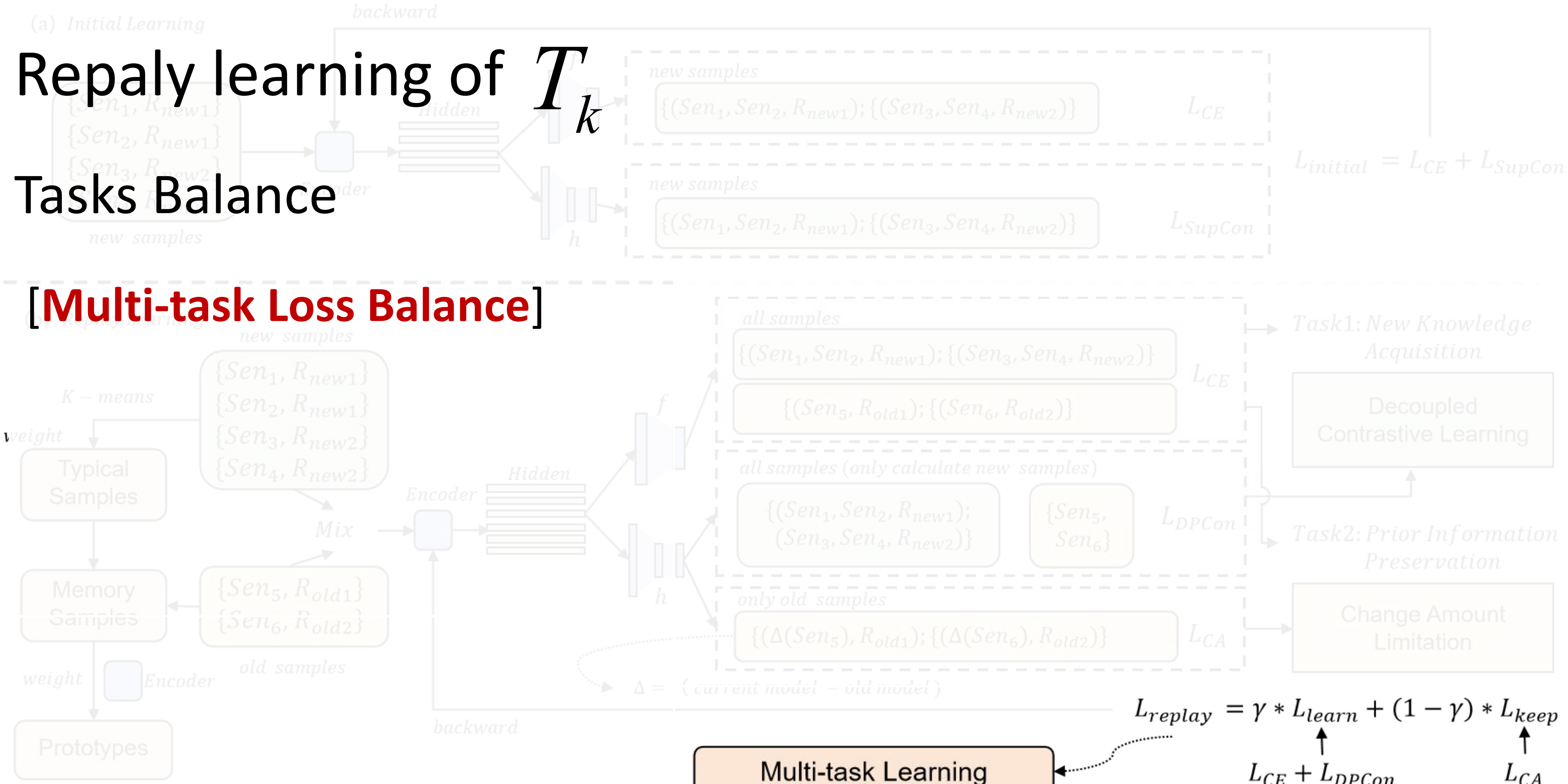


DP CRE: Multi-task Balance

Repaly learning of T_k

Tasks Balance

[**Multi-task Loss Balance**]



$$L_{replay} = \gamma * L_{learn} + (1-\gamma) * L_{keep}$$

L_{learn}: $L_{CE} + L_{DPCon}$

L_{keep}: L_{CA}

Multi-task Learning

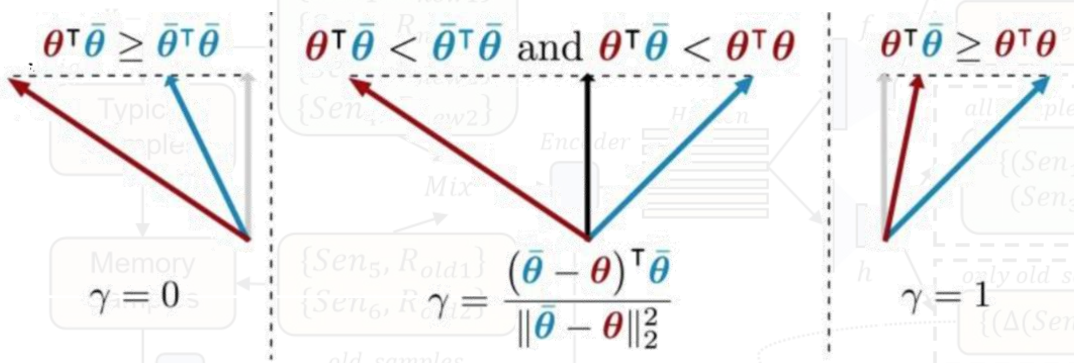
05 Methodology of DP-CRE

DP CRE: Multi-task Balance

Repaly learning of T_k

Tasks Balance

[**Multi-task Loss Balance**]



$$L_{replay} = \gamma * \underbrace{L_{learn}}_{L_{CE} + L_{DPCon}} + (1-\gamma) * \underbrace{L_{keep}}_{L_{CA}}$$

Multi-task Learning

DP CRE: Multi-task Balance

Repaly learning of T_k

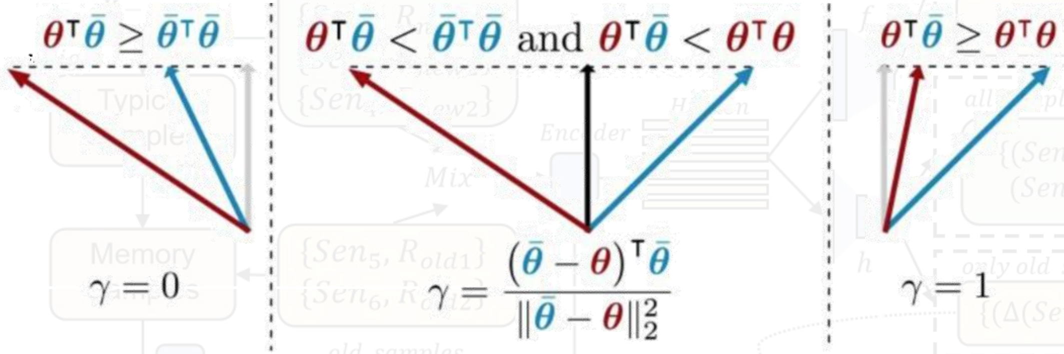
Tasks Balance

[**Multi-task Loss Balance**]

$$L_{keep} = k^{\lambda} \times L_{CA}$$

$$\gamma = \begin{cases} 1, & \theta_l^T \theta_k \geq \theta_l^T \theta_l \\ 0, & \theta_l^T \theta_k \geq \theta_k^T \theta_k \\ \frac{(\theta_k - \theta_l)^T \theta_k}{\|\theta_l - \theta_k\|_2^2}, & otherwise \end{cases} \tag{9}$$

$$L_{replay} = \gamma \times L_{learn} + (1-\gamma) \times L_{keep},$$



$$L_{replay} = \gamma * L_{learn} + (1-\gamma) * L_{keep}$$

$L_{learn} \leftarrow L_{CE} + L_{DPCon}$

$L_{keep} \leftarrow L_{CA}$

Multi-task Learning

06 Experimental Results of DP-CRE

DP-CRE: Main Performance

FewRel

Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
EA-EMR (Wang et al., 2019)	89.0	69.0	59.1	54.2	47.8	46.1	43.1	40.7	38.6	35.2
EMAR(BERT) (Han et al., 2020)	98.2	94.8	92.6	91.1	89.7	87.9	87.1	86.0	84.7	83.3
CML (Wu et al., 2021)	91.2	74.8	68.2	58.2	53.7	50.4	47.8	44.4	43.1	39.7
RP-CRE (Cui et al., 2021)	98.1	94.8	92.6	91.1	89.7	87.9	87.1	86.0	84.7	83.3
CR-ECL (Hu et al., 2022)	97.8	94.9	92.7	90.9	89.4	87.5	85.7	84.6	83.6	82.7
ACA (Wang et al., 2022b)	98.4	95.1	93.0	91.5	90.5	88.9	87.9	86.7	85.8	84.4
CRL (Zhao et al., 2022)	98.0	94.3	92.4	90.5	89.5	87.8	87.0	85.6	84.3	83.0
CEAR (Zhao et al., 2023)	98.3	**95.6**	93.5	92.0	90.8	89.3	88.0	86.8	85.6	84.0
Ours	**98.5**	95.4	**93.7**	**92.1**	**90.9**	**89.4**	**88.5**	**87.4**	**86.3**	**85.1**

TACRED

Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
EA-EMR (Wang et al., 2019)	47.5	40.1	38.3	29.9	24.0	27.3	26.9	25.8	22.9	19.8
EMAR(BERT) (Han et al., 2020)	98.0	93.0	89.7	84.7	82.7	81.5	79.0	77.5	77.6	77.1
CML (Wu et al., 2021)	57.2	51.4	41.3	39.3	35.9	28.9	27.3	26.9	24.8	23.4
RP-CRE (Cui et al., 2021)	96.6	91.4	88.8	84.8	82.8	81.0	77.9	77.4	76.5	75.7
CR-ECL (Hu et al., 2022)	97.3	92.5	88.2	85.6	83.7	83.3	81.8	80.1	77.7	76.8
ACA (Wang et al., 2022b)	**98.2**	93.8	89.9	85.9	84.2	82.7	80.5	78.4	78.6	77.5
CRL (Zhao et al., 2022)	98.0	**93.9**	90.8	86.0	84.9	82.9	80.1	79.2	79.4	78.5
CEAR (Zhao et al., 2023)	97.9	93.7	90.7	86.6	84.7	**84.3**	81.9	80.4	80.2	79.3
Ours	97.8	93.8	**91.5**	**87.5**	**85.7**	84.2	**82.9**	**81.3**	**81.5**	**80.7**

Experimental Results of DP-CRE

DP-CRE: Main Performance

FewRel										
Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
EA-EMR (Wang et al., 2019)	89.0	69.0	59.1	54.2	47.8	46.1	43.1	40.7	38.6	35.2
EMAR(BERT) (Han et al., 2020)	98.2	94.8	92.6	91.1	89.7	87.9	87.1	86.0	84.7	83.3
CML (Wu et al., 2021)	91.2	74.8	68.2	58.2	53.7	50.4	47.8	44.4	43.1	39.7
RP-CRE (Cui et al., 2021)	98.1	94.8	92.6	91.1	89.7	87.9	87.1	86.0	84.7	83.3
CR-ECL (Hu et al., 2022)	97.8	94.9	92.7	90.9	89.4	87.5	85.7	84.6	83.6	82.7
ACA (Wang et al., 2022b)	<u>98.4</u>	95.1	93.0	91.5	90.5	88.9	87.9	86.7	<u>85.8</u>	<u>84.4</u>
CRL (Zhao et al., 2022)	98.0	94.3	92.4	90.5	89.5	87.8	87.0	85.6	84.3	83.0
CEAR (Zhao et al., 2023)	98.3	**95.6**	<u>93.5</u>	<u>92.0</u>	<u>90.8</u>	<u>89.3</u>	<u>88.0</u>	<u>86.8</u>	85.6	84.0
1 **Ours**	**98.5**	<u>95.4</u>	**93.7**	**92.1**	**90.9**	**89.4**	**88.5**	**87.4**	**86.3**	**85.1**

TACRED										
Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
EA-EMR (Wang et al., 2019)	47.5	40.1	38.3	29.9	24.0	27.3	26.9	25.8	22.9	19.8
EMAR(BERT) (Han et al., 2020)	98.0	93.0	89.7	84.7	82.7	81.5	79.0	77.5	77.6	77.1
CML (Wu et al., 2021)	57.2	51.4	41.3	39.3	35.9	28.9	27.3	26.9	24.8	23.4
RP-CRE (Cui et al., 2021)	96.6	91.4	88.8	84.8	82.8	81.0	77.9	77.4	76.5	75.7
CR-ECL (Hu et al., 2022)	97.3	92.5	88.2	85.6	83.7	83.3	81.8	80.1	77.7	76.8
ACA (Wang et al., 2022b)	**98.2**	<u>93.8</u>	89.9	85.9	84.2	82.7	80.5	78.4	78.6	77.5
CRL (Zhao et al., 2022)	<u>98.0</u>	**93.9**	<u>90.8</u>	86.0	84.9	82.9	80.1	79.2	79.4	78.5
CEAR (Zhao et al., 2023)	97.9	93.7	90.7	86.6	84.7	**84.3**	81.9	80.4	80.2	79.3
1 **Ours**	97.8	<u>93.8</u>	**91.5**	**87.5**	**85.7**	<u>84.2</u>	**82.9**	**81.3**	**81.5**	**80.7**

1 DP-CRE outperforms previous CRE work.

06 Experimental Results of DP-CRE

DP-CRE: Main Performance

FewRel										
Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
EA-EMR (Wang et al., 2019)	89.0	69.0	59.1	54.2	47.8	46.1	43.1	40.7	38.6	35.2
EMAR(BERT) (Han et al., 2020)	98.2	94.8	92.6	91.1	89.7	87.9	87.1	86.0	84.7	83.3
CML (Wu et al., 2021)	91.2	74.8	68.2	58.2	53.7	50.4	47.8	44.4	43.1	39.7
RP-CRE (Cui et al., 2021)	98.1	94.8	92.6	91.1	89.7	87.9	87.1	86.0	84.7	83.3
CR-ECL (Hu et al., 2022)	97.8	94.9	92.7	90.9	89.4	87.5	85.7	84.6	83.6	82.7
ACA (Wang et al., 2022b)	98.4	95.1	93.0	91.5	90.5	88.9	87.9	86.7	85.8	84.4
CRL (Zhao et al., 2022)	98.0	94.3	92.4	90.5	89.5	87.8	87.0	85.6	84.3	83.0
CEAR (Zhao et al., 2023)	98.3	**95.6**	93.5	92.0	90.8	89.3	88.0	86.8	85.6	84.0
Ours	**98.5**	95.4	**93.7**	**92.1**	**90.9**	**89.4**	**88.5**	**87.4**	**86.3**	**85.1**

TACRED										
Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
EA-EMR (Wang et al., 2019)	47.5	40.1	38.3	29.9	24.0	27.3	26.9	25.8	22.9	19.8
EMAR(BERT) (Han et al., 2020)	98.0	93.0	89.7	84.7	82.7	81.5	79.0	77.5	77.6	77.1
CML (Wu et al., 2021)	57.2	51.4	41.3	39.3	35.9	28.9	27.3	26.9	24.8	23.4
RP-CRE (Cui et al., 2021)	96.6	91.4	88.8	84.8	82.8	81.0	77.9	77.4	76.5	75.7
CR-ECL (Hu et al., 2022)	97.3	92.5	88.2	85.6	83.7	83.3	81.8	80.1	77.7	76.8
ACA (Wang et al., 2022b)	**98.2**	93.8	89.9	85.9	84.2	82.7	80.5	78.4	78.6	77.5
CRL (Zhao et al., 2022)	98.0	**93.9**	90.8	86.0	84.9	82.9	80.1	79.2	79.4	78.5
CEAR (Zhao et al., 2023)	97.9	93.7	90.7	86.6	84.7	**84.3**	81.9	80.4	80.2	79.3
Ours	97.8	93.8	**91.5**	**87.5**	**85.7**	84.2	**82.9**	**81.3**	**81.5**	**80.7**

1 DP-CRE outperforms previous CRE work.

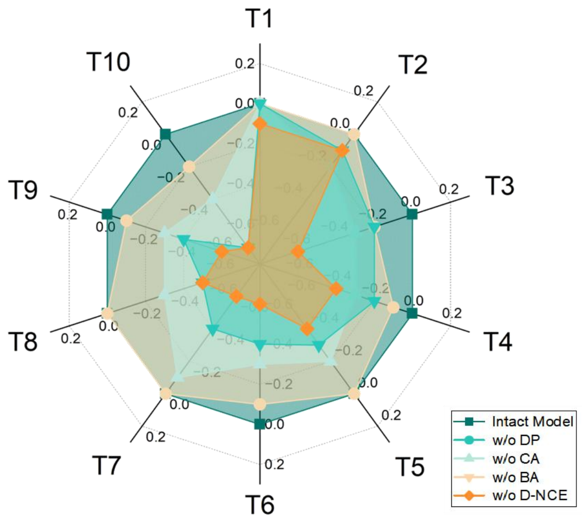
2 More significant enhancement in **the later CRE tasks**.
DP-CRE accumulate advantages when facing **denser** feature space and more **imbalanced** tasks number.

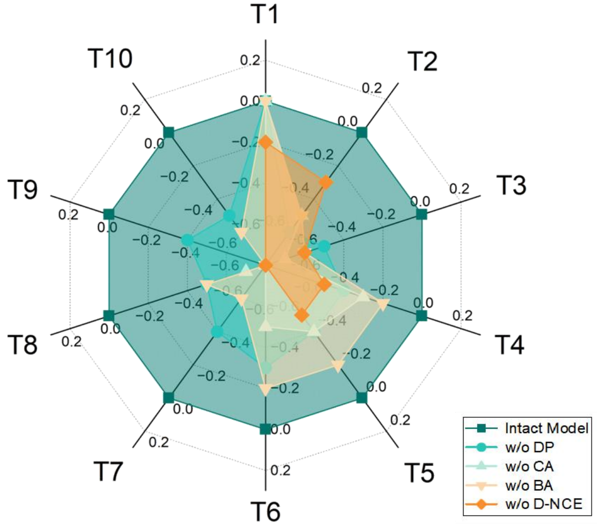
06 Experimental Results of DP-CRE

DP-CRE: Modules Ablation Study

Initial Learning(**IN**) , Decoupled Contrastive Learning(**DP**), Change Amount Limitation(**CA**), Multi-task Balance(**BA**) , Double-NCM Prediction(**D-NCM**)

	FewRel	TACRED
Intact Model	**85.1**	**80.7**
w/o IN	83.7	75.4
w/o DP	84.4	80.2
w/o CA	84.7	79.2
w/o BA	84.9	80.1
w/o D-NCM	84.4	79.7



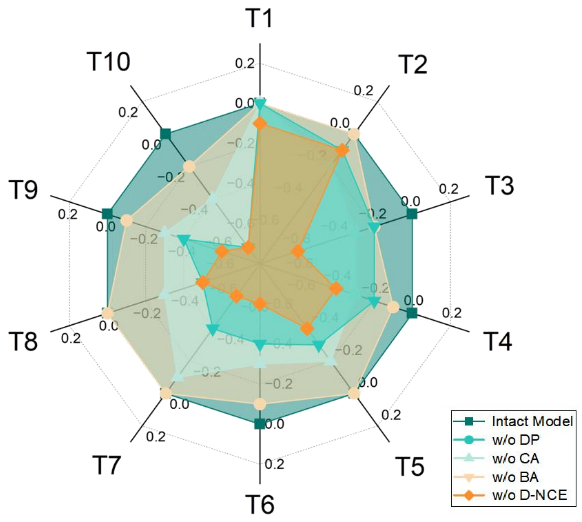


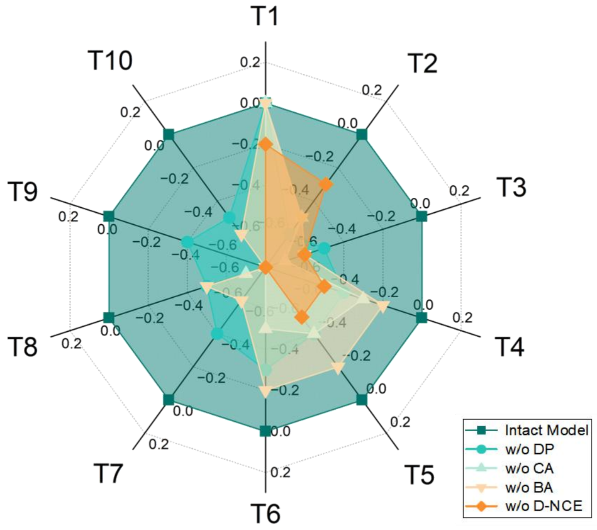
06 Experimental Results of DP-CRE

DP-CRE: Modules Ablation Study

Initial Learning(**IN**) , Decoupled Contrastive Learning(**DP**), Change Amount Limitation(**CA**), Multi-task Balance(**BA**) , Double-NCM Prediction(**D-NCM**)

	FewRel	TACRED
Intact Model	**85.1**	**80.7**
w/o IN	83.7	75.4
w/o DP	84.4	80.2
w/o CA	84.7	79.2
w/o BA	84.9	80.1
w/o D-NCM	84.4	79.7

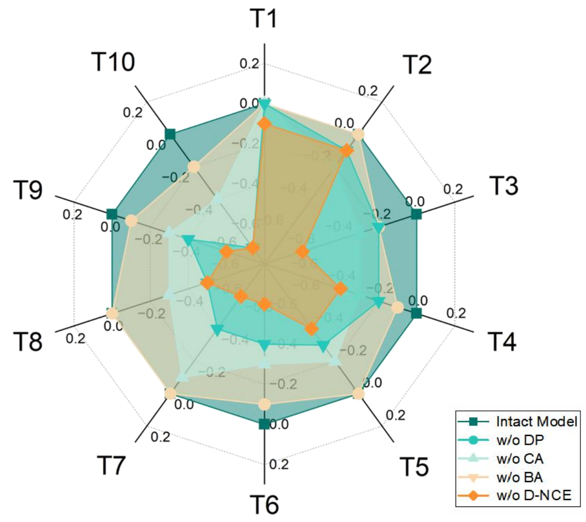


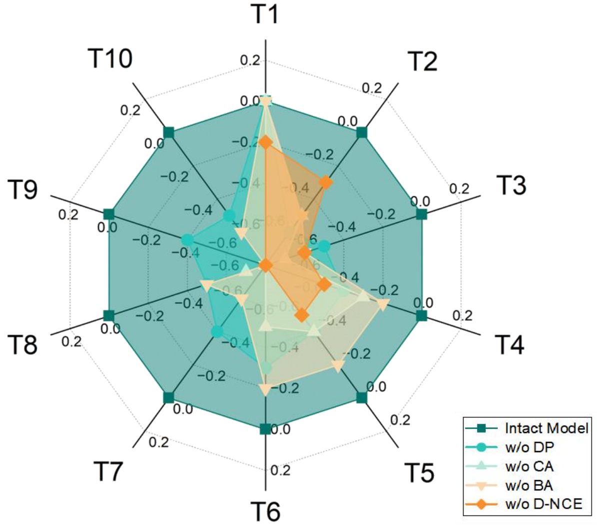


DP-CRE: Modules Ablation Study

Initial Learning(**IN**) , Decoupled Contrastive Learning(**DP**), Change Amount Limitation(**CA**), Multi-task Balance(**BA**) , Double-NCM Prediction(**D-NCM**)

	FewRel	TACRED
Intact Model	**85.1**	**80.7**
w/o IN	83.7	75.4
w/o DP	84.4	80.2
w/o CA	84.7	79.2
w/o BA	84.9	80.1
w/o D-NCM	84.4	79.7





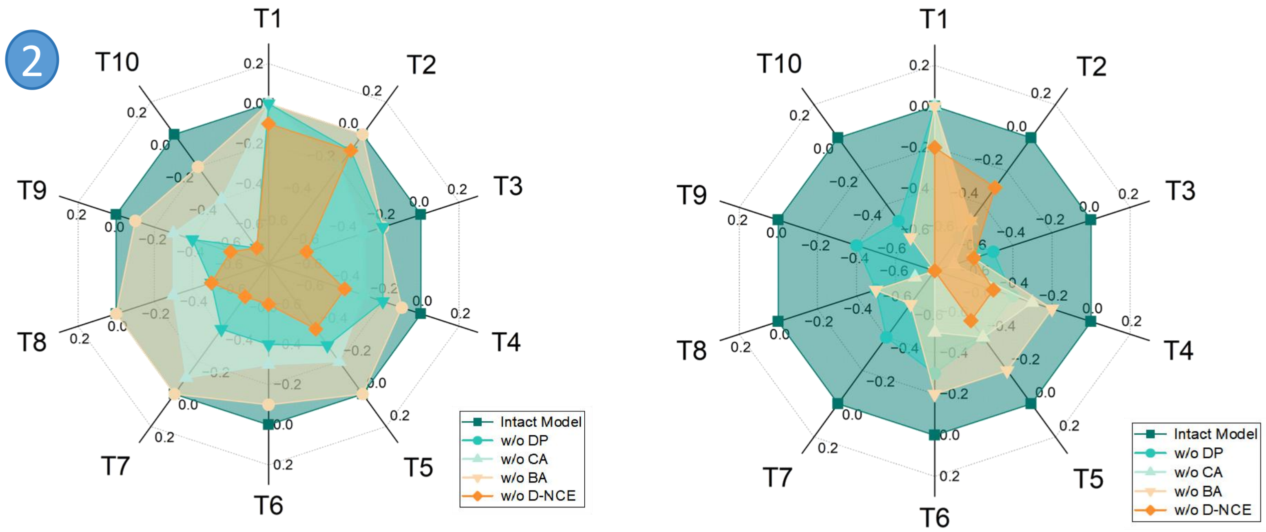
1 TACRED dataset consists of **a larger number of conflicting relation types**, CA-Limit more significant in handling frequent embedding changes.

DP-CRE: Modules Ablation Study

Initial Learning(**IN**) , Decoupled Contrastive Learning(**DP**), Change Amount Limitation(**CA**), Multi-task Balance(**BA**) , Double-NCM Prediction(**D-NCM**)

	FewRel	TACRED
Intact Model	**85.1**	**80.7**
w/o IN	83.7	75.4
w/o DP	84.4	80.2
w/o CA	84.7	79.2
w/o BA	84.9	80.1
w/o D-NCM	84.4	79.7

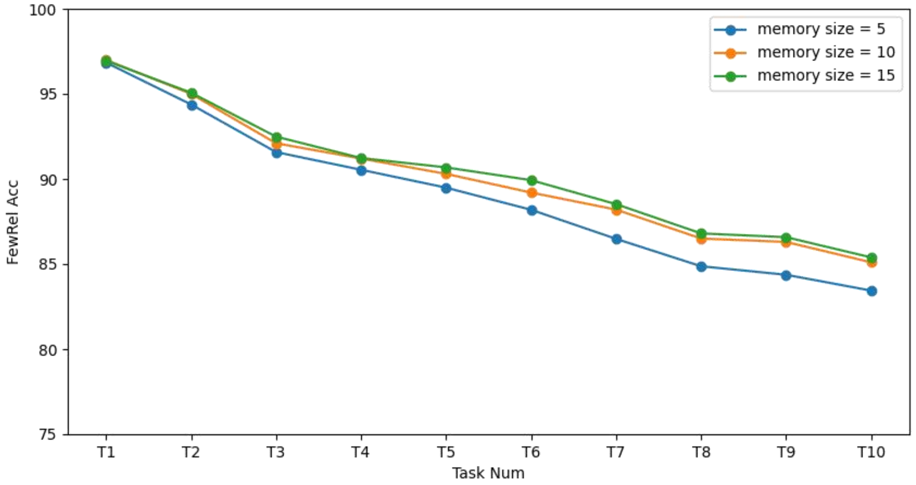
2

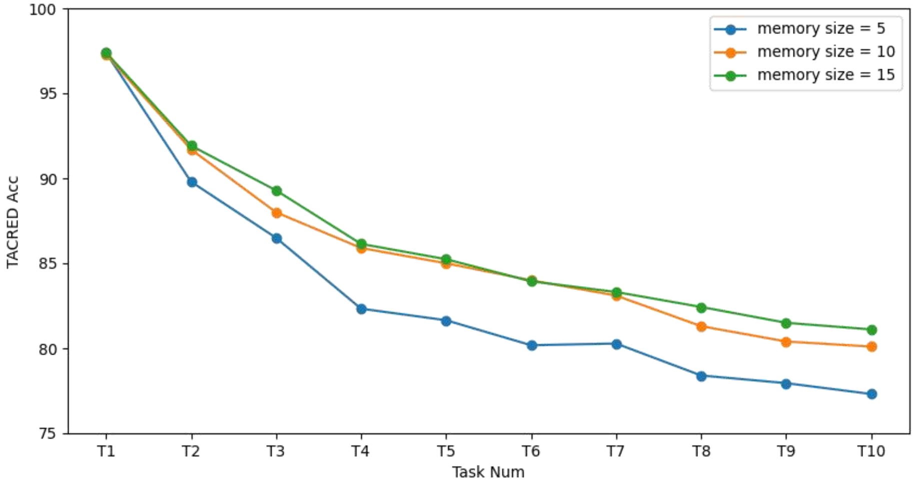


1 TACRED dataset consists of **a larger number of conflicting relation types**, CA-Limit more significant in handling frequent embedding changes.

2 **Δ accuracy (%)** .

DP-CRE: Influence of Memory Size



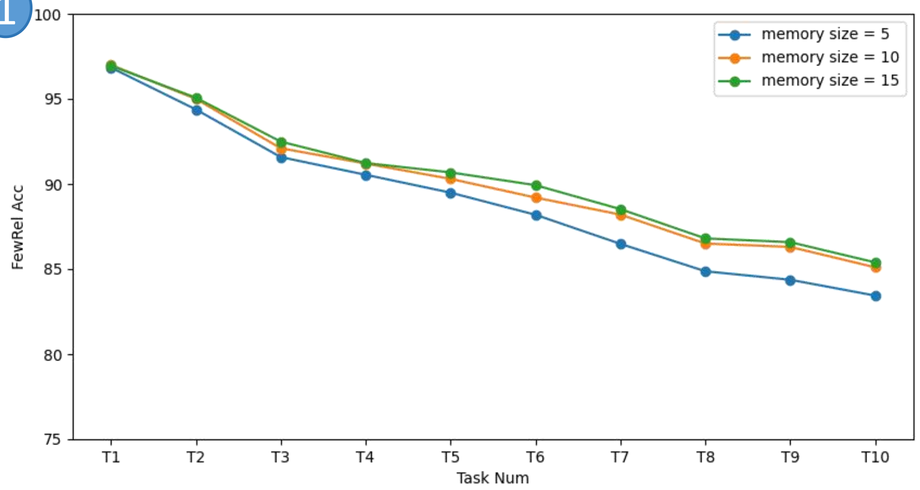


FewRel			
Memory Size	5	10	15
ACA (Wang et al., 2022b)	82.8	84.4	85.1
CRL (Zhao et al., 2022)	80.3	83.0	84.0
CEAR (Zhao et al., 2023)	82.6	84.0	84.9
Ours	**83.4**	**85.1**	**86.1**
TACRED			
Memory Size	5	10	15
ACA (Wang et al., 2022b)	76.2	77.5	78.7
CRL (Zhao et al., 2022)	75.0	78.5	79.7
CEAR (Zhao et al., 2023)	76.7	79.3	80.4
Ours	**77.3**	**80.7**	**81.3**

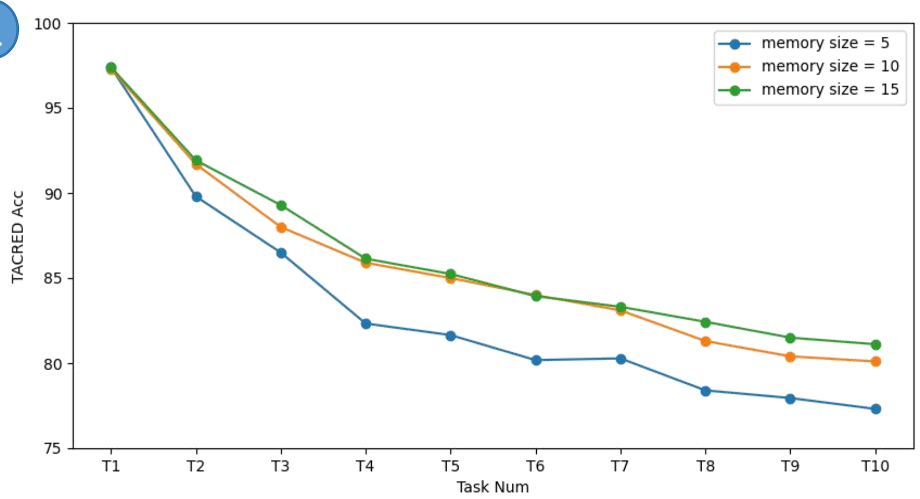
06 Experimental Results of DP-CRE

DP-CRE: Influence of Memory Size

1



1

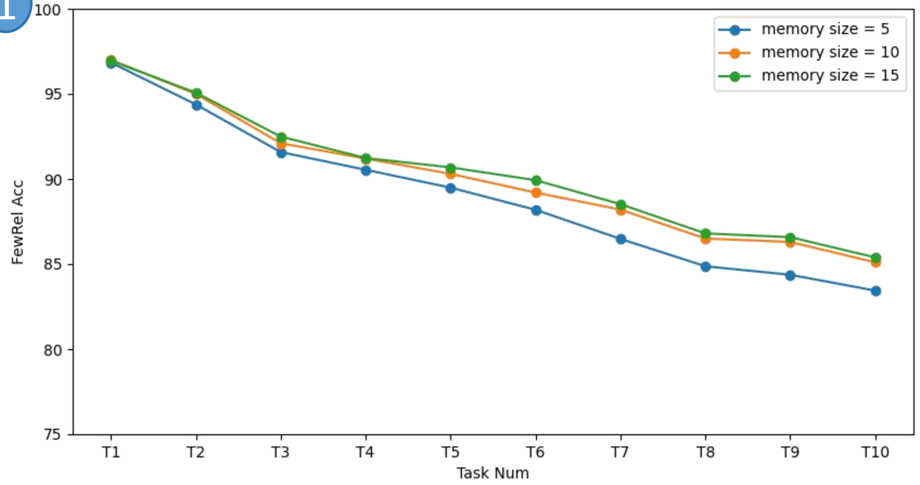


FewRel			
Memory Size	5	10	15
ACA (Wang et al., 2022b)	82.8	84.4	85.1
CRL (Zhao et al., 2022)	80.3	83.0	84.0
CEAR (Zhao et al., 2023)	82.6	84.0	84.9
Ours	**83.4**	**85.1**	**86.1**
TACRED			
Memory Size	5	10	15
ACA (Wang et al., 2022b)	76.2	77.5	78.7
CRL (Zhao et al., 2022)	75.0	78.5	79.7
CEAR (Zhao et al., 2023)	76.7	79.3	80.4
Ours	**77.3**	**80.7**	**81.3**

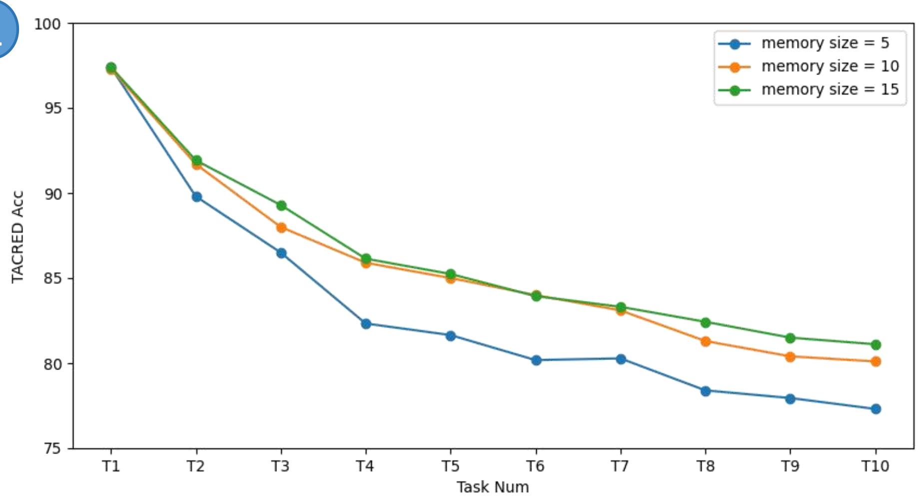
1 **Additional memory samples providing more information.**

DP-CRE: Influence of Memory Size

1



1



FewRel			
Memory Size	5	10	15
ACA (Wang et al., 2022b)	82.8	84.4	85.1
CRL (Zhao et al., 2022)	80.3	83.0	84.0
CEAR (Zhao et al., 2023)	82.6	84.0	84.9
Ours	**83.4**	**85.1**	**86.1**
TACRED			
Memory Size	5	10	15
ACA (Wang et al., 2022b)	76.2	77.5	78.7
CRL (Zhao et al., 2022)	75.0	78.5	79.7
CEAR (Zhao et al., 2023)	76.7	79.3	80.4
Ours	**77.3**	**80.7**	**81.3**

2

1 **Additional memory samples providing more information.**

2 Change amount for **each memory sample individually** make more memory samples information used.

DP-CRE: Task Balance Experiment

FewRel										
Model	Task	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
ACA	old	1.50	2.50	2.86	3.29	3.85	4.35	5.09	5.09	5.48
	new	1.33	2.03	3.06	3.13	4.69	4.06	5.31	6.34	5.53
CEAR	old	1.41	2.08	2.64	3.11	3.49	4.23	4.70	5.48	6.16
	new	1.08	1.80	**2.16**	**2.81**	3.66	3.41	4.94	5.59	5.13
Ours	old	**1.22**	**1.67**	**2.29**	**2.91**	**3.27**	**3.73**	**3.52**	**4.48**	**4.70**
	new	**0.96**	**1.63**	2.26	2.91	**3.42**	**3.09**	**4.78**	**5.34**	**4.59**

TACRED										
Model	Task	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
ACA	old	1.83	2.92	3.15	3.39	3.96	4.43	4.89	5.13	5.47
	new	1.30	2.33	3.05	3.10	4.75	4.18	5.28	5.85	5.85
CEAR	old	1.33	2.15	2.71	3.21	3.58	4.30	4.58	5.43	5.98
	new	**0.90**	2.00	2.18	**2.78**	**3.73**	3.25	5.13	5.05	5.45
Ours	old	**1.08**	**1.99**	**2.54**	**3.04**	**3.42**	**3.82**	**4.35**	**4.61**	**4.78**
	new	1.07	**1.80**	**2.08**	2.95	**3.73**	**3.13**	**4.90**	**5.00**	**5.10**

06 Experimental Results of DP-CRE

DP-CRE: Task Balance Experiment

FewRel

Model	Task	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
ACA	old	1.50	2.50	2.86	3.29	3.85	4.35	5.09	5.09	5.48
	new	1.33	2.03	3.06	3.13	4.69	4.06	5.31	6.34	5.53
CEAR	old	1.41	2.08	2.64	3.11	3.49	4.23	4.70	5.48	6.16
	new	1.08	1.80	**2.16**	**2.81**	3.66	3.41	4.94	5.59	5.13
Ours	old	**1.22**	**1.67**	**2.29**	**2.91**	**3.27**	**3.73**	**3.52**	**4.48**	**4.70**
	new	**0.96**	**1.63**	2.26	2.91	**3.42**	**3.09**	**4.78**	**5.34**	**4.59**

TACRED

Model	Task	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
ACA	old	1.83	2.92	3.15	3.39	3.96	4.43	4.89	5.13	5.47
	new	1.30	2.33	3.05	3.10	4.75	4.18	5.28	5.85	5.85
CEAR	old	1.33	2.15	2.71	3.21	3.58	4.30	4.58	5.43	5.98
	new	**0.90**	2.00	2.18	**2.78**	**3.73**	3.25	5.13	5.05	5.45
Ours	old	**1.08**	**1.99**	**2.54**	**3.04**	**3.42**	**3.82**	**4.35**	**4.61**	**4.78**
	new	1.07	**1.80**	**2.08**	2.95	**3.73**	**3.13**	**4.90**	**5.00**	**5.10**

1 New and old tasks calculated **separately**.

06 Experimental Results of DP-CRE

DP-CRE: Task Balance Experiment

FewRel										
Model	Task	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
ACA	old	1.50	2.50	2.86	3.29	3.85	4.35	5.09	5.09	5.48
	new	1.33	2.03	3.06	3.13	4.69	4.06	5.31	6.34	5.53
CEAR	old	1.41	2.08	2.64	3.11	3.49	4.23	4.70	5.48	6.16
	new	1.08	1.80	**2.16**	**2.81**	3.66	3.41	4.94	5.59	5.13
Ours	old	**1.22**	**1.67**	**2.29**	**2.91**	**3.27**	**3.73**	**3.52**	**4.48**	**4.70**
	new	**0.96**	**1.63**	2.26	2.91	**3.42**	**3.09**	**4.78**	**5.34**	**4.59**

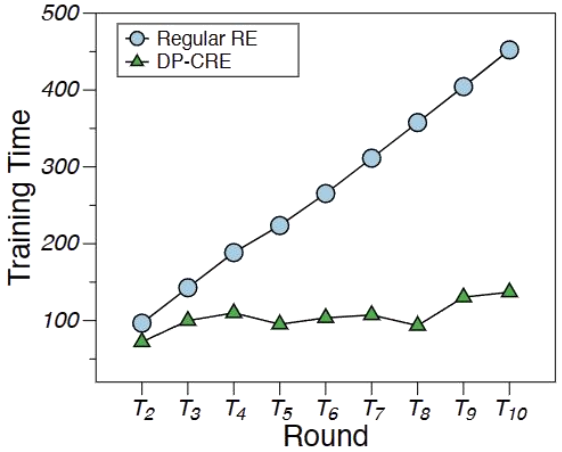
TACRED										
Model	Task	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
ACA	old	1.83	2.92	3.15	3.39	3.96	4.43	4.89	5.13	5.47
	new	1.30	2.33	3.05	3.10	4.75	4.18	5.28	5.85	5.85
CEAR	old	1.33	2.15	2.71	3.21	3.58	4.30	4.58	5.43	5.98
	new	**0.90**	2.00	2.18	**2.78**	**3.73**	3.25	5.13	5.05	5.45
Ours	old	**1.08**	**1.99**	**2.54**	**3.04**	**3.42**	**3.82**	**4.35**	**4.61**	**4.78**
	new	1.07	**1.80**	**2.08**	2.95	**3.73**	**3.13**	**4.90**	**5.00**	**5.10**

1 New and old tasks calculated **separately**.

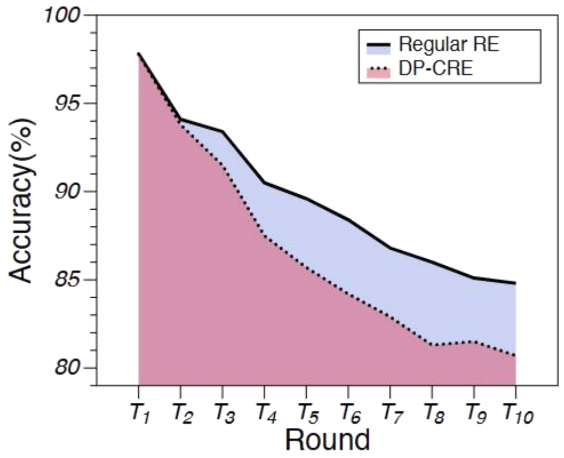
2 ΔF1 of the CRE model and the regular RE model.
Prevents any **over-bias towards either side** in case of conflicts, thereby ensuring a balanced model.

06 Experimental Results of DP-CRE

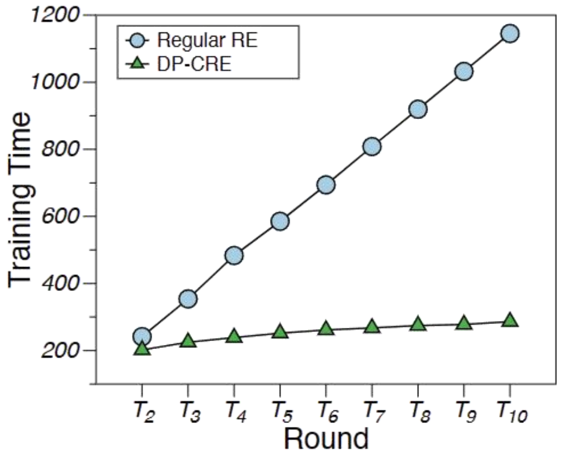
DP-CRE: Training Time Discussion



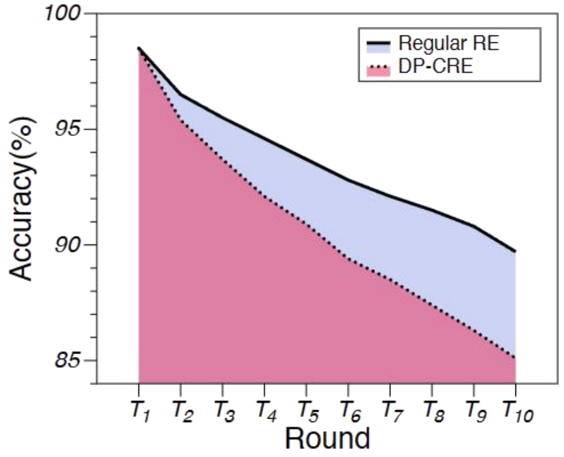
(a) TACRED



(a) TACRED



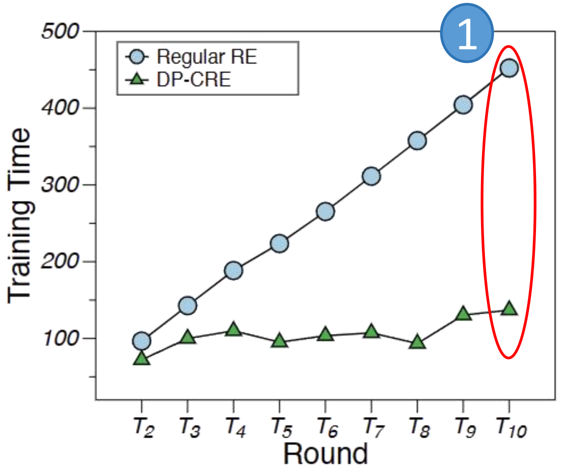
(b) FewRel



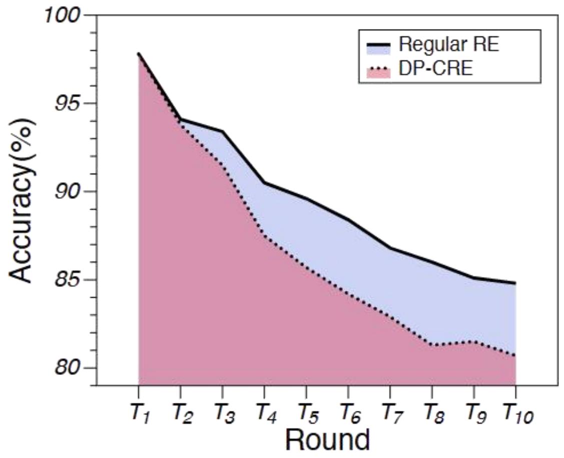
(b) FewRel

06 Experimental Results of DP-CRE

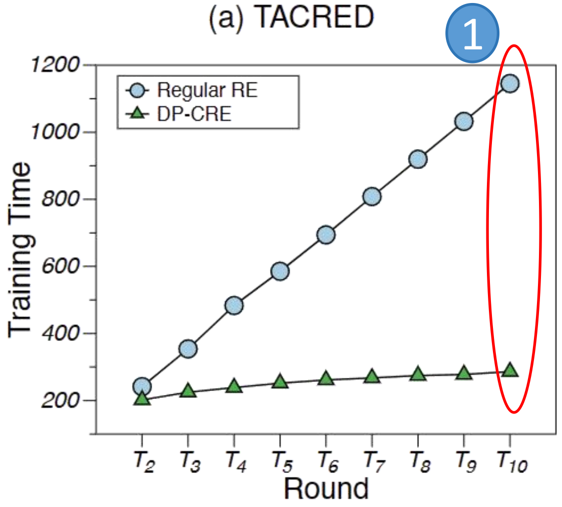
DP-CRE: Training Time Discussion



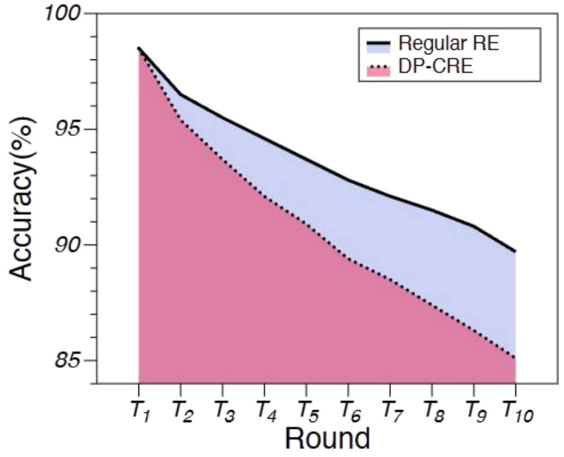
(a) TACRED



(a) TACRED



(b) FewRel



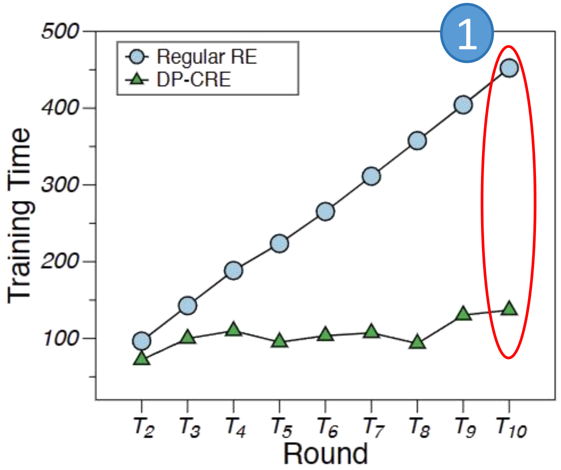
(b) FewRel

1

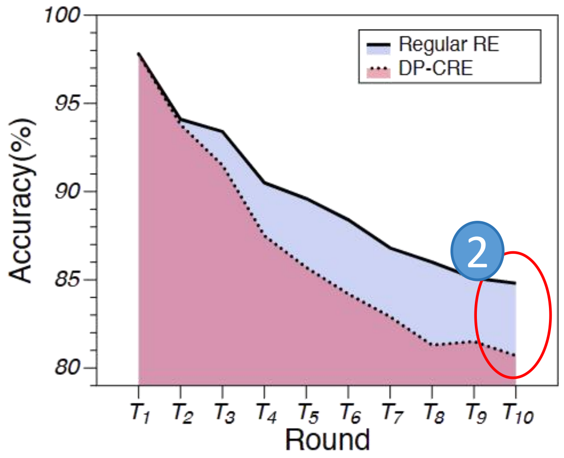
Reduce the **training time and the cost** .

06 Experimental Results of DP-CRE

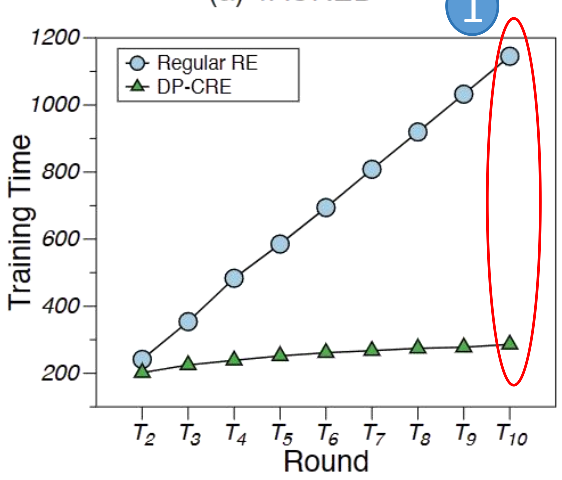
DP-CRE: Training Time Discussion



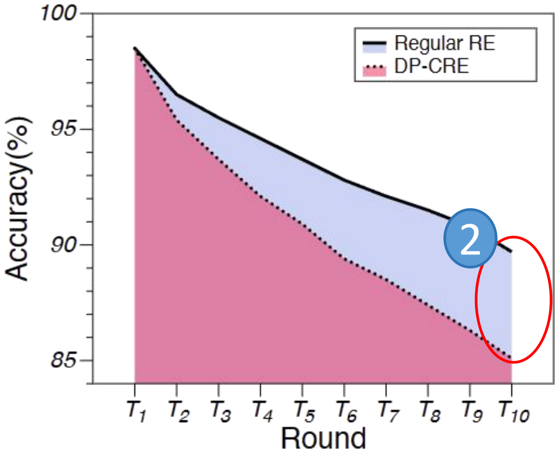
(a) TACRED



(a) TACRED



(b) FewRel



(b) FewRel

1 Reduce the **training time and the cost** .

2 A minor reduction in accuracy

DP-CRE

Balance prior information preservation **and new** knowledge acquisition.

Monitor the **changes** in embedding and maintain the **structural information of** memory samples.

Enhance the performance of SOTA CRE models.