SoftMCL: Soft Momentum Contrastive Learning for Fine-grained Sentiment-aware Pre-training

Jin Wang, Xuejie Zhang
Yunnan University

Contact: wangjin@ynu.edu.cn
xjzhang@ynu.edu.cn

Liang-Chih Yu
Yuan Ze University

Contact: lcyu@saturn.yzu.edu.tw

LREC-COLING 2024

Outline

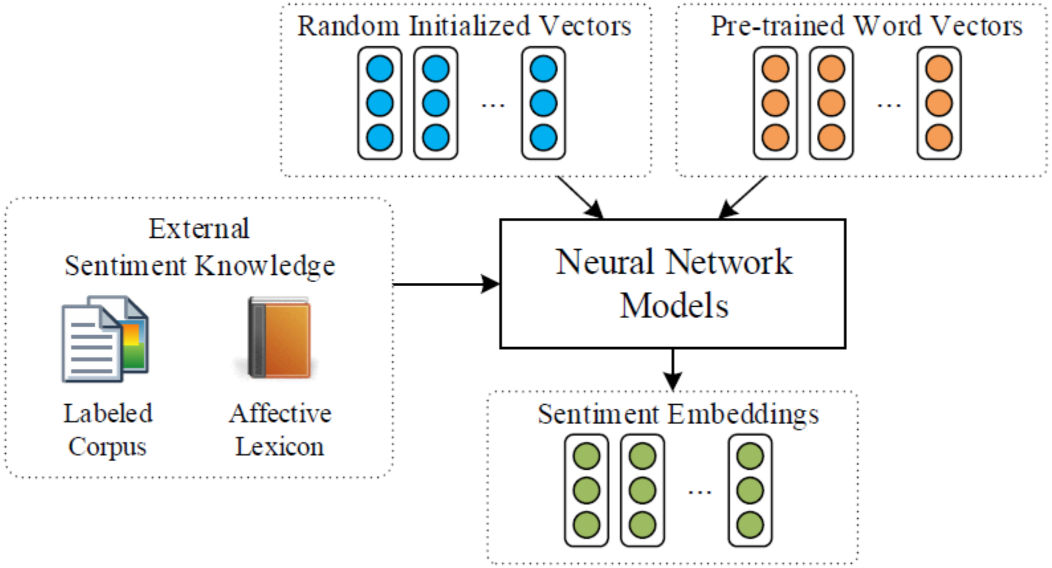
- **Introduction**
  - Sentiment-aware Pre-training
  - Sentiment refinement
- **Methods**
  - Sentiment Similarity
  - Sentiment-aware Contrastive Learning
  - Momentum Queue
- **Evaluation**
  - Both classification and regression tasks
- **Conclusions**

Introduction

- The core idea of pre-training is to generate similar contextual representations for words or sentences with similar contexts.
- Fail to distinguish the affective impact of a particular context to **a specific word since certain words may have special semantics and sentiments in specific contexts** (Agrawal et al., 2018).
- **For example:**
- **The battery life is *long*.** **Positive**
- **It takes a *long* time to focus.** **Negative**

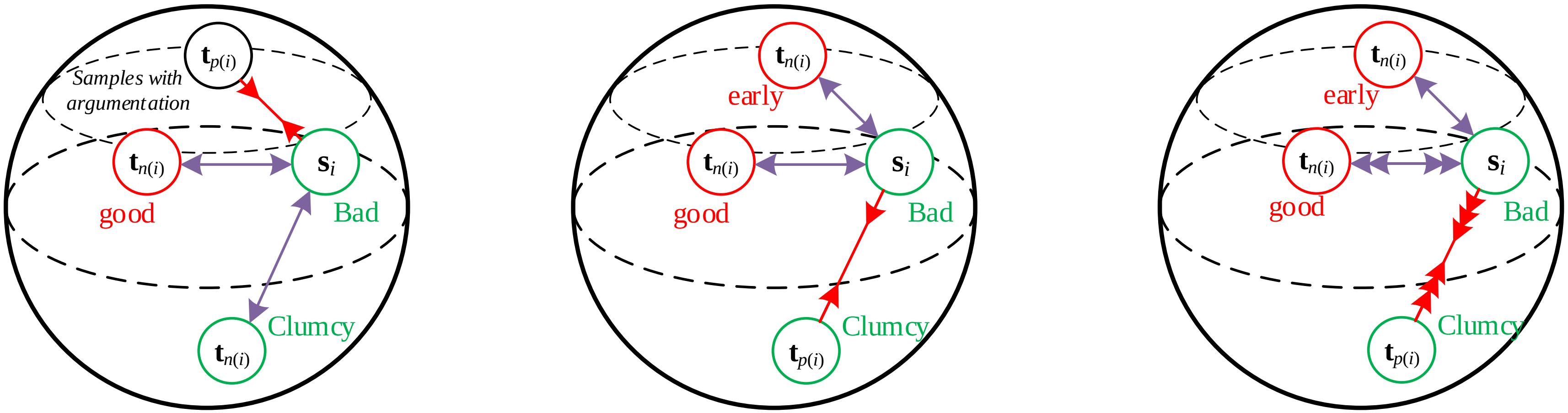
Introduction

- To assist PLMs in obtaining affective information
  - **Sentiment-aware Pre-training** (Abdalla et al., 2019; Fu et al., 2018)
  - **Sentiment Refinement** (Utsumi, 2019; Yu et al., 2018a,b)



Different Views

- Training a model to **<u>pull together an anchor and a positive sample</u>** in the latent space and **<u>push apart the anchor from many negative samples.</u>**



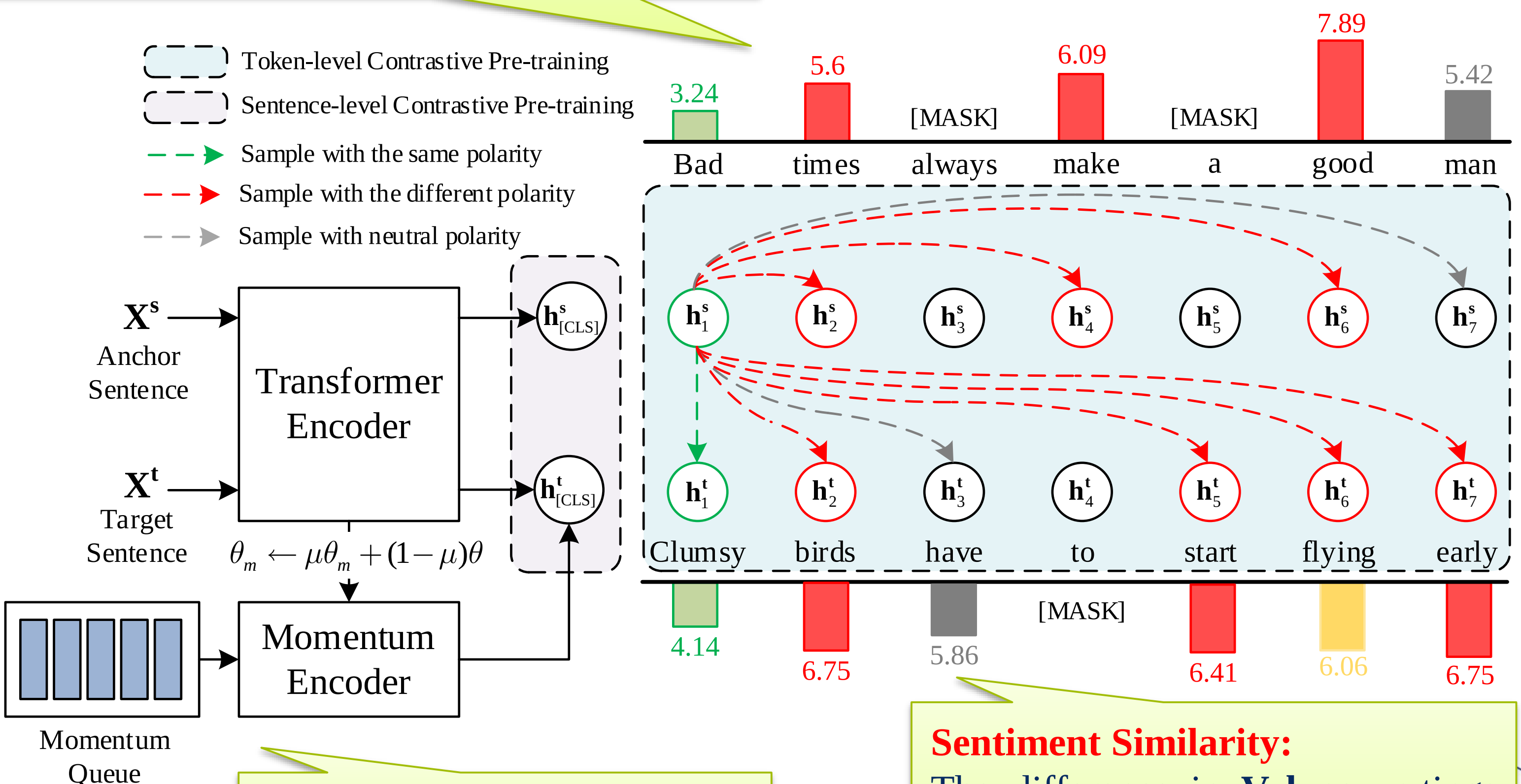
(a) Unsupervised Contrastive Learning

(b) Supervised Contrastive Learning with Hard Labels

(c) Supervised Contrastive Learning with Soft Labels

Soft Momentum Contrastive Learning

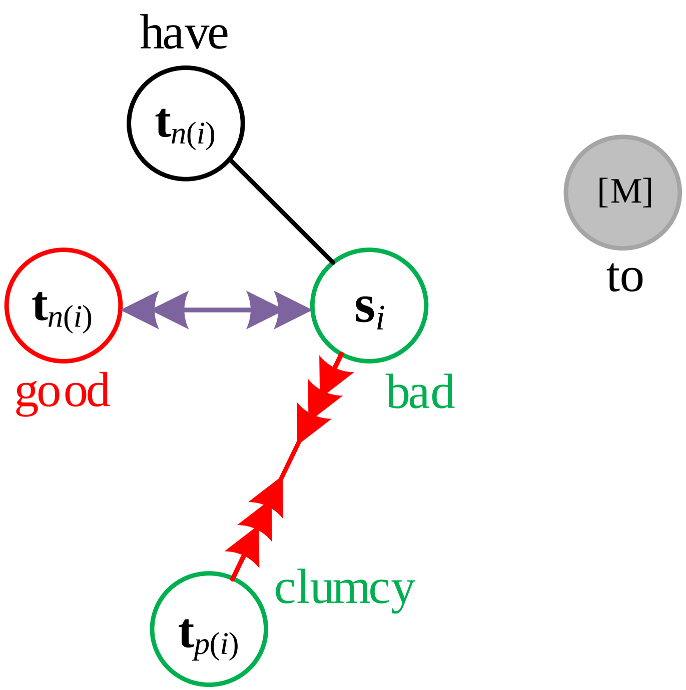
Sentiment-aware Contrastive Learning:
Token-level (VA Lexicon: **Extended-ANEW**)
Sentiment-level (VA Corpora: **EmoBank**)



Momentum Queue:
expand the contrastive samples.

Sentiment Similarity:
The difference in **Valence** ratings between samples thus measures the sentiment similarity

Sentiment Similarity



- **Valence-Arousal Ratings**

  bad (Valence: 3.24, **negative**)

  clumsy (Valence: 4.14, **negative**)

  good (Valence: 7.89, **positive**)

  have (Valence: 5.86, **neutral**)

  to (do not appear in lexicon, [MASK])

$$\Delta(\mathbf{h}_1, \mathbf{h}_2) = 1 - \frac{|y_1 - y_2|}{y_{\max} - y_{\min}}$$

Sentiment-aware Contrastive Learning

- **Hard Labels**

$$\mathcal{L}_{\mathrm{SCL}} = -\sum_{i\in I}\sum_{p\in P(i)}\frac{1}{|P(i)|}\log\frac{\exp(\mathbf{h}_i^{\top}\cdot\mathbf{h}_p/\tau)}{\sum_{k\in A(i)}\exp(\mathbf{h}_i^{\top}\cdot\mathbf{h}_k/\tau)}$$

Sentiment Similarity:
The **ratio of samples** which have the same sentiment with the target token or sentence.

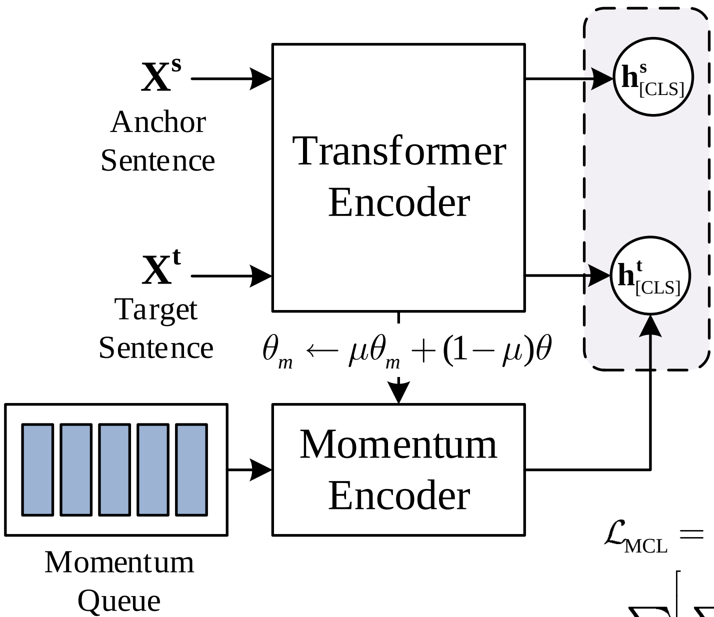
Semantic Similarity:
The similarity distribution of hidden representations learned by the encoder in the latent space.

- **Soft Labels**

$$\mathcal{L}_{\mathrm{SentiCL}} = -\sum_{i\in I}\sum_{j\in B(i)}\frac{\Delta(\mathbf{h}_i,\mathbf{h}_j)}{\sum_{l\in B(i)}\Delta(\mathbf{h}_i,\mathbf{h}_l)}\log\frac{\exp(\mathbf{h}_i^{\top}\cdot\mathbf{h}_j/\tau)}{\sum_{k\in B(i)}\exp(\mathbf{h}_i^{\top}\cdot\mathbf{h}_k/\tau)}$$

Sentiment Similarity:
The difference in **Valence** ratings between samples thus measures the sentiment similarity

Momentum Queue



- **The number of contrastive samples**

Previous method can contrast each one with other $B(i) - 1$ samples.

Momentum Queue stores $Q(i)$ samples. Thus, we obtain $B(i) + Q(i)$ samples for contrastive learning.

Samples from the same batch

$$\mathcal{L}_{\text{MCL}} = -\sum_{i\in I}\left[\sum_{j\in B(i)} \frac{\Delta(\mathbf{h}_i,\mathbf{h}_j)}{\sum_{l\in\{B(i),Q(i)\}}\Delta(\mathbf{h}_i,\mathbf{h}_l)} \log \frac{\exp(\mathbf{h}_i^\top\cdot\mathbf{h}_j/\tau)}{\sum_{k\in\{B(i),Q(i)\}}\exp(\mathbf{h}_i^\top\cdot\mathbf{h}_k/\tau)} + \sum_{m\in Q(i)} \frac{\Delta(\mathbf{h}_i,\mathbf{h}_m)}{\sum_{l\in\{B(i),Q(i)\}}\Delta(\mathbf{h}_i,\mathbf{h}_l)} \log \frac{\exp(\mathbf{h}_i^\top\cdot\mathbf{h}_m/\tau)}{\sum_{k\in\{B(i),Q(i)\}}\exp(\mathbf{h}_i^\top\cdot\mathbf{h}_k/\tau)}\right]$$

Samples from the momentum queue

Experimental Settings

- Datasets
  - Phrase-level Intensity Prediction (Regression) (Kiritchenko et al., 2016)
  - Sentence-level Classification
    - Stanford Sentiment Treebank (**SST-2/5**) (Socher et al., 2013)
    - Movie Review (**MR**) (Pang and Lee, 2005)
    - Customer Review (**CR**) (Conneau and Kiela, 2018)
    - **IMDB** (Maas et al., 2011)
    - **Yelp-2/5** (Zhang et al., 2013)
  - Sentence-level Regression
    - **Emobank** (Buechel and Hahn, 2017; Buechel and Hahn, 2016)
    - **Facebook** (Preotiuc-Pietro et al., 2016)
  - Aspect-level Sentiment Analysis (Pontikiet al., 2014)
    - **Lap14**
    - **Rest14**

Experimental Settings

- Baselines
  - General-purposed pre-trained models:
    BERT, XLNet, RoBERTa and **DeBERTa**
  - Sentiment-aware pre-trained models:
    BERT-PT, **SentiBERT**, **SentiLARE**, **SENTIX** and **SCAPT**
- Evaluation Metrics
  - Accuracy (*Acc*， Classification task)
  - Mean absolute error (*MAE*, Regression task)
  - Kendall's correlation coefficient (*k*, Regression task)
  - Pearson's correlation coefficient (*r*, Regression task)

Experimental Results (Phrase)

Model	Word		Phrase	
	$k\uparrow$	$\rho\uparrow$	$k\uparrow$	$\rho\uparrow$
General pre-trained models				
BERT	0.638	0.841	0.624	0.829
XLNet	0.644	0.852	0.637	0.833
RoBERTa	0.656	0.868	0.642	0.849
DeBERTa	0.667	0.866	0.644	0.852
Sentiment-aware pre-trained models				
BERT-PT	0.704	0.880	0.688	0.877
SentiBERT	0.699	0.886	0.692	0.882
SentiLARE	0.712	0.894	0.702	0.891
Proposed model with ablation study				
SoftMCL	**0.778**	**0.928**	**0.775**	**0.922**
w/o WP	0.698	0.878	0.692	0.874
w/o SP	0.702	0.889	0.685	0.865
w/o MoCL	0.724	0.895	0.713	0.904

Experimental Results (Classification)

Model	SST-2	SST-5	MR	CR	IMDB	Yelp-2	Yelp-5
	Acc ↑	Acc ↑	Acc ↑	Acc ↑	Acc ↑	Acc ↑	Acc ↑
General pretrained models							
BERT	90.82	53.37	87.52	88.22	93.87	97.74	70.16
XLNet	92.04	56.33	89.45	88.33	96.21	97.41	70.23
RoBERTa	92.28	54.89	89.41	88.47	94.68	97.98	70.12
DeBERTa	94.82	56.89	89.52	88.49	94.92	97.88	70.22
Sentiment-aware pretrained models							
BERT-PT	91.42	53.24	87.30	86.52	93.99	97.77	69.90
SentiBERT	92.18	56.87	88.59	88.24	94.04	97.66	69.94
SentiLARE	94.28	58.59	90.82	89.52	95.71	98.22	71.57
SENTIX	93.30	55.57	-	88.72	94.78	97.83	-
Proposed model with ablation study							
SoftMCL	**96.54**	**59.82**	**92.56**	**90.02**	**96.85**	**98.29**	**71.76**
w/o WP	94.01	58.25	90.14	87.67	94.31	95.73	69.88
w/o SP	93.65	58.04	89.79	87.32	93.96	95.35	69.62
w/o MoCL	91.33	56.61	87.55	85.15	91.62	92.99	67.89

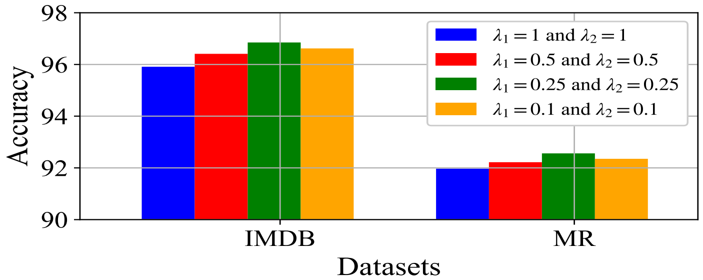
Experimental Results (Regression)

Model	Emobank		Facebook	
	MAE↓	r↑	MAE↓	r↑
General pretrained models				
BERT	0.581	0.521	0.719	0.635
XLNet	0.523	0.589	0.706	0.654
RoBERTa	0.518	0.592	0.694	0.642
DeBERTa	0.514	0.591	0.711	0.631
Sentiment-aware pretrained models				
BERT-PT	0.506	0.610	0.708	0.631
SentiBERT	0.505	0.612	0.688	0.677
SentiLARE	0.498	0.615	0.669	0.671
Proposed model with ablation study				
SoftMCL	**0.462**	**0.639**	**0.642**	**0.685**
w/o WP	0.475	0.629	0.663	0.661
w/o SP	0.495	0.620	0.734	0.634
w/o MoCL	0.492	0.622	0.710	0.662

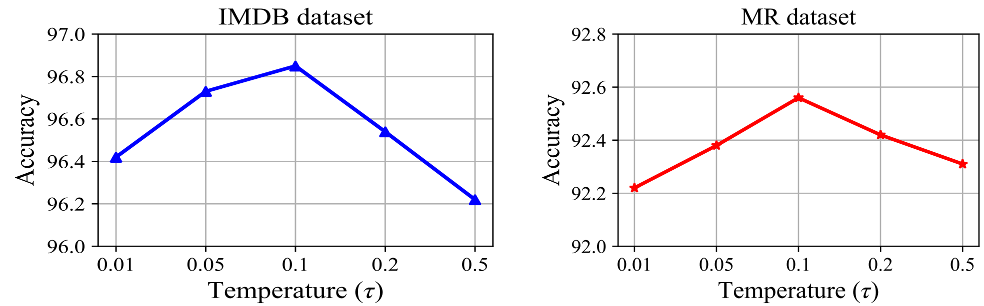
Experimental Results (Aspect Classification)

Model	Aspect-level sentiment classification			
	Lap14		Res14	
	Acc	F1	Acc	F1
General pretrained models				
BERT	78.53	73.11	83.77	76.06
XLNet	80.00	75.88	84.93	76.70
RoBERTa	81.03	77.16	86.07	79.21
DeBERTa	81.14	77.22	86.12	79.43
Sentiment-aware pretrained models				
BERT-PT	78.46	73.82	85.58	77.99
SentiBERT	76.87	71.74	83.71	75.42
SentiLARE	82.16	78.70	88.32	81.63
SENTIX	80.56	-	87.32	-
SCAPT	82.76	79.15	89.11	83.91
Proposed model with ablation study				
SoftMCL	**83.82**	**81.08**	**89.23**	**83.95**
w/o WP	81.63	78.96	86.91	81.75
w/o SP	81.31	78.66	86.57	81.43
w/o MoCL	79.29	76.7	84.42	79.41

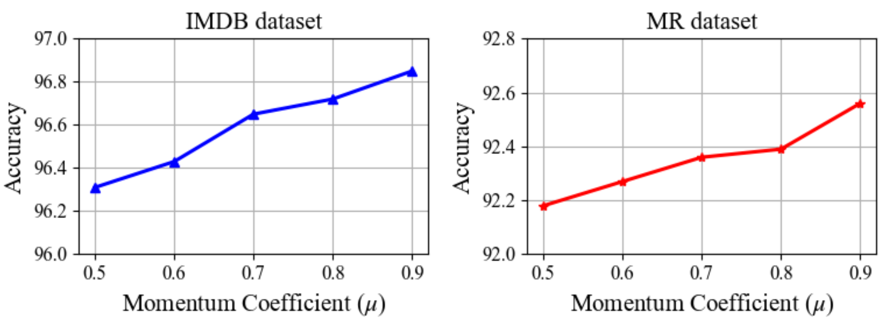
More Detailed Analysis



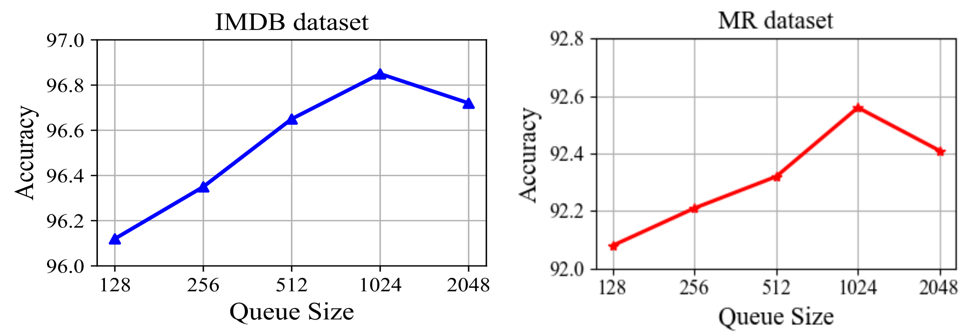
(a) The effect of different balance coefficient.



(b) The effect of different temperature.

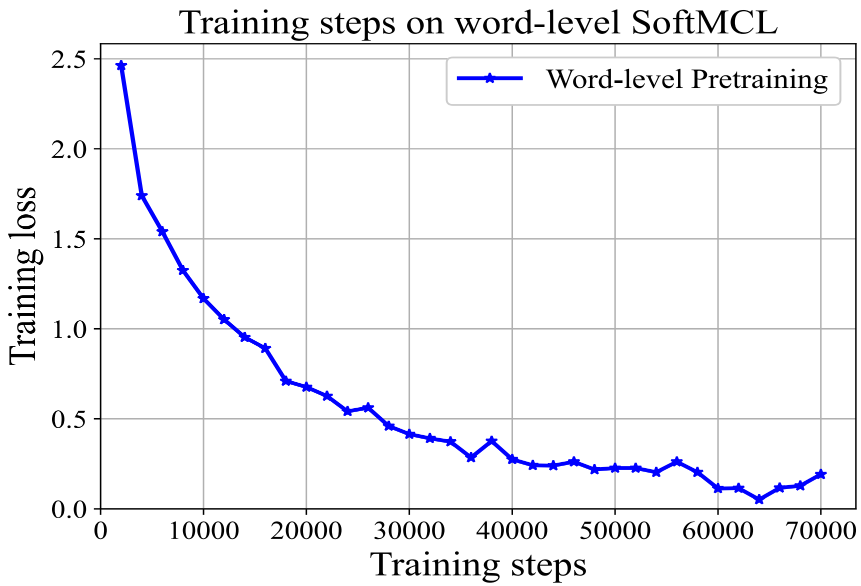


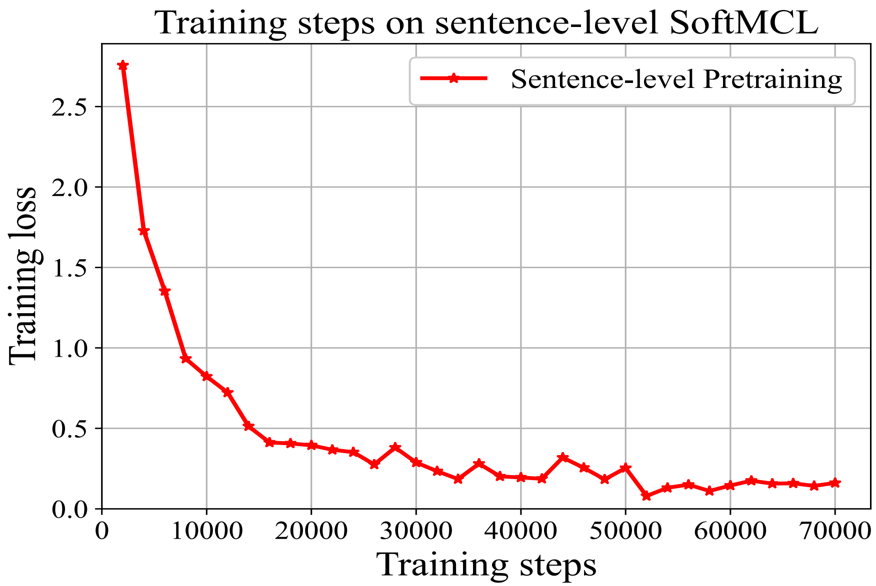
(c) The effect of different momentum coefficient.



(d) The effect of different queue size.

More Detailed Analysis





- The training loss curve of the proposed SoftMCL on word- and sentence-level pre-training. The SoftMCL model gradually converges within 40,000 steps for word-level and 10,000 steps for sentence-level.

Conclusions and Future Work

- This study proposed a soft momentum contrastive learning for sentiment-aware pre-training on both the word- and sentence-level to enhance the PLM's ability to learn affective information.
- Future Work
  - Incorporate the sentiment information into decoder-only architecture.
  - Conduct more detailed analysis to continue enhancing the proposed method.

Thank You for listening

References

- Ameeta Agrawal, Aijun An, and Manos Papagelis. 2018. Learning Emotion-enriched Word Representations. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING-2018)*, pages 950–961.
- Mohamed Abdalla, Magnus Sahlgren, and Graeme Hirst. 2019. Enriching Word Embeddings with a Regressor Instead of Labeled Corpora. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-2019)*, pages 6188–6195.
- Peng Fu, Zheng Lin, Fengcheng Yuan, Weiping Wang, and Dan Meng. 2018. Learning Sentiment-specific Word Embedding via Global Sentiment Representation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-2018)*, pages 4808–4815.
- Akira Utsumi. 2019. Refining Pretrained Word Embeddings Using Layer-wise Relevance Propagation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP-2018)*, pages 4840–4846.
- Liang-Chih Yu, Jin Wang, K. Robert Lai, and Xuejie Zhang. 2018a. Refining Word Embeddings for Sentiment Analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP-2017)*, pages 534–539.
- Liang-Chih Yu, Jin Wang, K.Robert Lai, and Xuejie Zhang. 2018b. Refining Word Embeddings Using Intensity Scores for Sentiment Analysis. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 26(3):671–681.

References

- Svetlana Kiritchenko, Saif M Mohammad, and Mohammad Salameh. 2016. SemEval-2016 Task 7: Determining sentiment intensity of English and Arabic phrases. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2016)*, pages 42–51.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2013)*, pages 1631–1642.
- Bo Pang and Lillian Lee. 2005. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL-2005)*, pages 115–124.
- Alexis Conneau and Douwe Kiela. 2018. SentEval: An Evaluation Toolkit for Universal Sentence Representations. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC-2018)*, pages 1699–1704.
- Andrew L Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning Word Vectors for Sentiment Analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011)*, pages 142–150.
- Xiang Zhang, Junbo Zhao, and Yann LeCun. 2013. Character-level Convolutional Networks for Text Classification. In *Proceedings of Advances in Neural Information Processing Systems (NIPS-2015)*, pages 3057–3061.

References

- Sven Buechel and Udo Hahn. 2016. Readers vs. writers vs. texts : Coping with different perspectives of text understanding in emotion annotation. In *Proceedings of the 11th Linguistic Annotation Workshop in EACL 2017*, pages 1–12.
- Sven Buechel and Udo Hahn. 2017. EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2017)*, pages 578–585.
- Daniel Preotiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes C. Eichstaedt, Margaret Kern, Lyle Ungar, and Elizabeth P. Shulman. 2016. Modeling valence and arousal in Facebook posts. In *Proceedings of Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA-2016)*, pages 9–15.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect-based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation, Proceedings (SemEval-2014)*, pages 27–35.