CCDC: A Chinese-Centric Cross Domain Contrastive Learning Framework

Hao Yang[1(✉)], Shimin Tao[1], Minghan Wang[1], Min Zhang[1], Daimeng Wei[1], Shuai Zhao[2], Miaomiao Ma[1], and Ying Qin[1]

[1] 2012 Labs, Huawei Technologies Co., Ltd., Beijing, China
{yanghao30,taoshimin,wangminghan,zhangmin186,weidaimeng,mamiaomiao,qinying}@huawei.com

[2] Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Haidian District, Beijing, China
zhaoshuaiby@bupt.edu.cn

Abstract. Unsupervised/Supervised SimCSE [5] achieves the SOTA performance of sentence-level semantic representation based on contrastive learning and dropout data augmentation. In particular, supervised SimCSE mines positive pairs and hard-negative pairs through Natural Language Inference (NLI) entailment/contradiction labels, which significantly outperforms other unsupervised/supervised models. As NLI data is scarce, can we construct pseudo-NLI data to improve the semantic representation of multi-domain sentences? This paper proposes a Chinese-centric Cross Domain Contrastive learning framework (CCDC), which provides a "Hard/Soft NLI Data Builder" to annotate entailment/contradiction pairs through Business Rules and Neural Classifiers, especially out-domain but semantic-alike sentences as hard-negative samples. Experiments show that the CCDC framework can achieve both intra-domain and cross-domain enhancement. Moreover, with the Soft NLI Data Builder, the CCDC framework can achieve the best results of all domains with one model, improving 34% and 11% in terms of the Spearman correlation coefficient compared with the baseline (BERT-base) and strong baseline (unsupervised SimCSE). And through empirical analysis, this framework effectively reduces the anisotropy of the pre-trained models and shows semantic clustering over unsupervised SimCSE.

Keywords: Contrastive learning · SimCSE · Cross domain framework

1 Introduction

Learning universal sentence presentation is a fundamental problem in natural language processing and has been extensively studied in [12,16]. Based on Contrastive Learning [1,20], SimCSE [5] provides two simple but strong sentence semantic presentation models: (1) Unsupervised SimCSE, which extracts multi-view features [17] through dropout [15]. A sentence with itself is created as an

G. Memmi et al. (Eds.): KSEM 2022, LNAI 13369, pp. 225–236, 2022.
https://doi.org/10.1007/978-3-031-10986-7_18

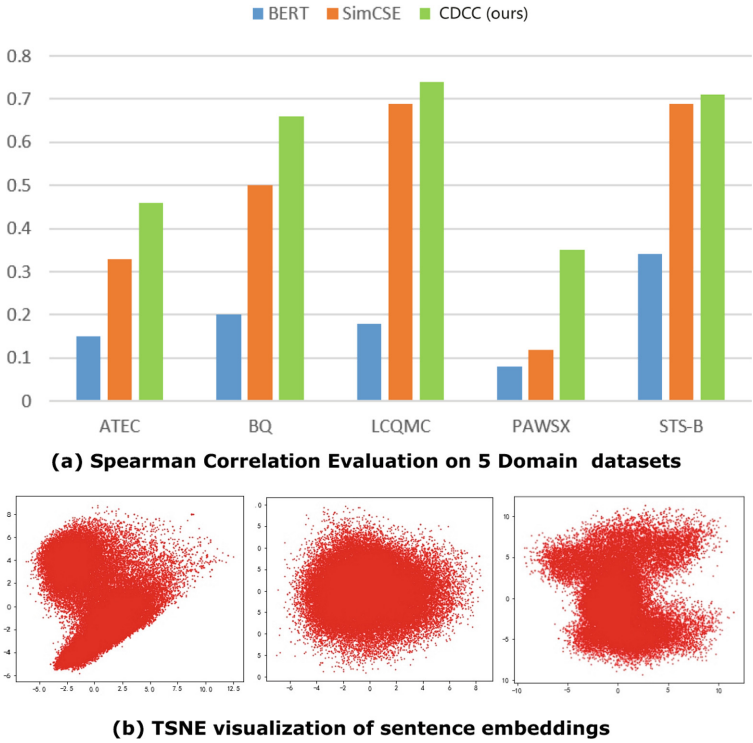

Fig. 1. Sentence-level representation and visual analysis of anisotropy in multiple domains under SBERT, unsupervised SimCSE and CCDC models

anchor-positive pair, while an anchor-negative pair is formed with other sentences in the batch. InfoNCE [1] loss is used to shorten the positive value and push the negative value away, and model parameters are optimized. Based on the multi-view learning of dropout, unsupervised SimCSE outperforms other unsupervised/supervised models. (2) Supervised SimCSE further improves performance by using NLI data labels as data augmentation. The entailment sentence pair is pictured as the anchor-positive pair, and the contradiction sentence pair as the hard-negative pair. Results show that unsupervised SimCSE exceeded the previous SOTA model IS-BERT [20] by 7.9%, while supervised SimCSE exceeded unsupervised SimCSE by 6.9%. Supervised SimCSE is also 4.6% higher than the previous supervised SOTA models SBERT [14] and BERT-whitening [16].

Extending supervised SimCSE to multi-domain sentence representation scenarios [24] requires solving two problems. One is hard-negative mining for out-domain but semantic-alike samples. Another is how to generate pseudo-NLI data from popular Chinese Sentence corpora, like sentence pairs PAWS-X [25]/BQ [2] or regression sentence pairs STS-B [13], and so on.

To solve these two problems, this paper provides a Chinese-centric Cross Domain Contrastive learning framework (CCDC) that adds two features: (a)

Domain augmentation Constrative Learning and (b) Pseudo NLI Data Generator . Domain augmentation Constrastive Learning uses out-domain but semantic - alike sentences as hard-negatives, improving cross-domain performance. Pseudo-NLI data generators, which help create $\langle anchor, positive, negative \rangle$ triplets from classification/regression sentence pair datasets, include business rule-based Hard NLI Generators and neural classifier-based Soft NLI Generators.

In order to better understand the superiority of CCDC, three model embedding spaces are mapped: the original BERT base model, the unsupervised SimCSE model, and the CCDC model. It finds that the anisotropy properties are optimized by unsupervised SimCSE and CCDC, while the CCDC model shows the domain clustering tendency [22]. Additional singular value experiments are visualized, showing that the domain-enhanced Contrastive Learning objective "flats" the singular value distribution of the sentence embedding space, thereby improving consistency.

2 Related Work

2.1 Contrastive Learning

Contrastive Learning aims to learn effective representation by pulling semantically close neighbors together and pushing apart non-neighbors [7]. It assumes a set of paired examples $D = f\{(x_i, x_i^+)\}_{i=1}^m$, where x_i and x_i^+ are semantically related. Following the contrastive framework in [3], the training object is a cross-entropy loss with in-batch negatives: let h_i and h_i^+ denote the representations of x_i and x_i^+, for a mini-batch with N pairs, the training objective for (x_i, x_i^+) is:

$$l_i = -log \frac{e^{sim(h_i, h_i^+)/\tau}}{\sum_{j=1}^N e^{sim(h_i, h_j^+)/\tau}}, \tag{1}$$

where τ is a temperature hyperparameter and $sim(h1, h2)$ is the cosine similarity of $\frac{h_1^T h_2}{||h_1|| \cdot ||h_2||}$.

2.2 Unsupervised SimCSE

The idea of unsupervised SimCSE is extremely simple in a minimal form of data augmentation, where positive pairs are (x_i, x_i) , compared with the traditional (x_i, x_i^+). Unsupervised SimCSE takes exactly the same sentence as the positive pair, and its embeddings only differ in dropout masks.

$$l_i = -log \frac{e^{sim(h_i^{z_i}, h_i^{z_i'})/\tau}}{\sum_{j=1}^N e^{sim(h_i^{z_i}, h_j^{z_j'})/\tau}}, \tag{2}$$

where $h_i^{z_i}, h_i^{z_i'}$ are the same sentence x_i with different dropout presentations z_i, z_i'.

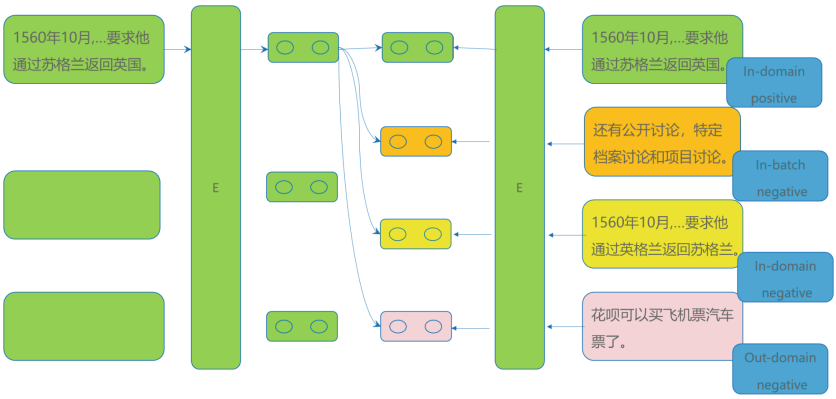

Fig. 2. CCDC framework

2.3 Supervised SimCSE

For supervised SimCSE, an easy hard-negative mining strategy is added that extends (x_i, x_i^+) to (x_i, x_i^+, x_i^-) . Prior work [6] has demonstrated that supervised Natural Language Inference (NLI) datasets [21] are effective for learning sentence embeddings, by predicting the relationship between two sentences and dividing it into three categories: entailment, neutral, or contradiction. Contradiction pairs are taken as hard-negative samples and give model significant negative signals.

$$l_i = -log \frac{e^{sim(h_i, h_i^+)/\tau}}{\sum_{j=1}^{N} e^{sim(h_i, h_j^+)/\tau} + e^{sim(h_i, h_j^-)/\tau}}, \tag{3}$$

where h_i, h_j^+ are positive pairs labelled as entailment, while h_i, h_j^- are hard-negative pairs labelled as contradiction.

2.4 Sentence Contrastive Learning with PLMs

The recent success of comparative learning is heavily dependent on the pre-trained language models (PLMs), like BERT [18], Roberta [11], and Albert [9]. Unsupervised/Supervised SimCSE [5], PairSupCon [26], IS-BERT [20], BERT-Position [19], BERT-whitening [16] are based on BERT, Roberta, Albert, and so on, for the pre-training model to improve the training efficiency and result.

3 CCDC Framework

3.1 Cross-Domain Sentences as Hard-Negative Samples

In order to enhance the sentence representation effect of multi-domain contrastive learning [23,24], the CCDC framework is designed as follows based on the supervised SimCSE framework. The pseudo-NLI data format is similar to

Table 1. CCDC samples

	Chinese sentence	Corresponding English sentence
Anchor	1560年10月，...要求他通过苏格兰返回英国。	In October 1560, ... asked him to return to United Kingdom through Scotland.
In-domain-positive	1560年10月，...要求他通过苏格兰返回英国。	In October 1560, ... asked him to return to United Kingdom through Scotland.
In-batch-negative	还有公开讨论，特定档案讨论和项目讨论。	There are also open discussions, file-specific discussions and project discussions.
In-domain-negative	1560年10月，...要求他通过英格兰返回苏格兰。	In October 1560, ... Asked him to return to Scotland through United Kingdom.
Out-domain-negative	花呗可以买飞机票汽车票了	You can buy a plane ticket, a bus ticket.

the SimCSE NLI format in that the former uses (DT:sen0, DT:sen1, DT:sen2) similar to the (anchor, positive, negative) triplet, where DT is short for domain tag. Anchor-negative pairs include negative examples of in-batch, in-domain, and out-domain, as highlighted in yellow, orange, and red in Fig. 2. The in-domain negative example and out-domain negative example can be considered as hard-negative samples, as can be seen in Table 1.

3.2 Hard NLI Data Builder

To construct pseudo-NLI data, the $(x, x+, x-)$ triplet needs to be generated based on three traditional sentence semantic problems, including classification, regression, and NLI. The Hard NLI Data Builder based on domain rules is implemented as follows. If (x_i, x_j) is a positive sample of semantic similarity (classification problem) or the similarity is greater than the average (regression problem), a negative sample of a sentence x_k is randomly selected to form an NLI triplet (x_i, x_j, x_k); if (x_i, x_j) is a negative sample or is less than or equal to the average value, the anchor is repeated to form an NLI triplet (x_i, x_i, x_j). The Hard NLI Data Builder process is as in Fig. 3.

As can be seen, classification/regression data is classified into entailment (positive) and contradiction (negative) categories. The created NLI data can be used to train supervised SimCSE.

3.3 Soft NLI Data Builder

In addition to the rule-based Hard NLI Data Builder, a Soft NLI Data Builder can be built based on a neural network classification model. The Entailment/Contradiction classifier can be encoded like a Siamese network, where two towers of BERT and pooling have shared weights, and output sentence embeddings as a feature. A Softmax Binary Classifier is a simple MLP network based on the feature triplet of $(f(x), f(x'), |f(x) - f(x')|)$, as can be seen in Fig. 3.

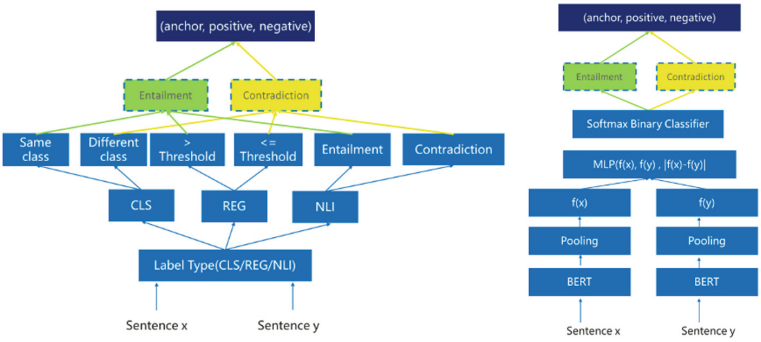

Fig. 3. Diagrams of Hard NLI data builder (left) and Soft NLI data builder (right)

4 Experiment

4.1 Data Preparation

The multi-domain semantic representation of sentences is like a different view enhancement of general semantic representation. Like [16], in order to verify the multi-view enhancement effect of the CCDC framework, the most famous Chinese sentence question pair datasets, Ant Financial Artificial Competition(ATEC) [4], BQ [2], LCQMC [10], PAWS-X from Google [25], and STS-B [2] are used. The detailed datasets are shown in Table 2.

Table 2. Chinese sentence pair datasets in 5 domains

Dataset	Type	Train	Test
ATEC	Semantic identification	62 k	20 k
BQ	Semantic identification	100 k	10 k
LCQMC	Semantic identification	238 k	12 k
PAWS-X	Binary identification	49 k	2 k
STS-B	Semantic similarity	5 k	1.3 k

4.2 Training Details

SBERT [14] is choosen as the training framework, and tiny, small, base, and large pre-trained models are selected for comparison. The training device is NVIDIA-V100, which has a 32G GPU. The batch size is 128 for tiny/small/base PLMs, 96 for huge PLM due to insufficient graphics memory, and the temperature is 0.05 using an Adam [8] optimizer. The learning rate is set as 5e-5 for tiny/small/base models and 1e-5 for large models, and warm-up steps account for 10% of the total training steps. Just like supervised SimCSE, our model is trained with 3 epochs.

4.3 CCDC with One-Domain Training and In-Domain/Out-Domain Testing

Like in [16], BERT-base is chosen as the baseline and unsupervised SimCSE as the strong baseline.

ATEC is used as training/testing data for in-domain experiments. The ATEC Spearman coefficient reached 46%, a performance improvement of 31% over the baseline and of 13% over the strong baseline. The other four in-domain experiments also achieved a 25–56% and 2–21% performance improvement over the baseline and strong baseline respectively, as can be seen in Table 3.

In the In Domain-Enhanced Confusion Matrix, all items are semi-positive, and most of them are positive. In-domain CCDC training can improve cross-domain performance, as can be seen in Table 4.

Table 3. CCDC with one-domain training and in-domain testing

	ATEC	BQ	LCQMC	PAWS-X	STS-B
BERT	0.15	0.2	0.18	0.08	0.34
SimCSE	0.33	0.5	0.69	0.12	0.69
ATEC	0.46 (+0.13)				
BQ		0.65 (+0.15)			
LCQMC			0.74 (+0.05)		
PAWS-X				0.33 (+0.21)	
STS-B					0.71 (+0.02)

Table 4. CCDC results with one-domain training and out-domain testing

	ATEC	BQ	LCQMC	PAWS-X	STS-B	Average
BERT	0.15	0.2	0.18	0.08	0.34	0.19
SimCSE	0.33	0.5	0.69	0.12	0.69	0.46
ATEC	0.46 (+0.31)	0.56 (+0.36)	0.68 (+0.50)	0.09 (+0.01)	0.66 (+0.32)	0.49
BQ	0.38 (+0.23)	0.65 (+0.45)	0.69 (+0.51)	0.10 (+0.02)	0.65 (+0.31)	0.49
LCQMC	0.34 (+0.19)	0.45 (+0.25)	0.74 (+0.56)	0.08 (+0.00)	0.69 (+0.35)	0.46
PAWS-X	0.24 (+0.09)	0.40 (+0.20)	0.55 (+0.37)	0.33 (+0.25)	0.57 (+0.23)	0.42
STS-B	0.23 (+0.08)	0.34 (+0.14)	0.63 (+0.45)	0.09 (+0.01)	0.71 (+0.37)	0.40

Table 5. CCDC results with all-domain training and the Hard/Soft NLI data builder

	ATEC	BQ	LCQMC	PAWS-X	STS-B	Average
BERT	0.15	0.2	0.18	0.08	0.34	0.19
SimCSE	0.33	0.5	0.69	0.12	0.69	0.46
ATEC	0.46	0.56	0.68	0.09	0.66	0.49
BQ	0.38	0.65	0.69	0.10	0.65	0.49
LCQMC	0.34	0.45	0.74	0.08	0.69	0.46
PAWS-X	0.24	0.40	0.55	0.33	0.57	0.42
STS-B	0.23	0.34	0.63	0.09	0.71	0.40
CCDC w/All-Hard-NLI-Builder	**0.46**	**0.66**	**0.74**	0.33	0.67	0.57
CCDC w/All-Soft-NLI-Builder	**0.46**	**0.66**	**0.74**	**0.35**	**0.71**	**0.58**

4.4 CCDC with the Hard/Soft NLI Data Builder

With the Hard NLI Data Builder, a multi-domain CCDC model is trained. Domains ATEC, BQ, LCQMC, and PAWS-X all achieved the best performance of 46%, 66%, 74%, and 33% respectively, with BQ being especially better than in-domain SOTA. Only STS-B falls short by 4%, maybe because the training data volume is insufficient. The average performance achieved 57%.

To eliminate the impact of data imbalances, a multi-domain model is trained with the Soft NLI Data Builder [24]. The CCDC model achieved the best performance in all domains, and even PAWS-X outperformed the in-domain SOTA by 2%. The average performance of all domains achieved 58%, an improvement of 41% and 12% over the baseline and the strong baseline respectively in Table 5.

5 Analysis

Based on the research of [5] on neurolinguistic representation, the baseline of the traditional pre-training model has two problems. (1) Due to the anisotropy of the space, the space will eventually shrink to a narrow cone. (2) The eigenvalues corresponding to the eigenvectors will be attenuated greatly, leading to a huge gap between the head feature and the back feature.

Empirical visual analysis for Anisotropy. visual analysis has been performed on the baseline, strong baseline, and CCDC models. 5000 data entries were extracted from each test set, with a total of 50,000 sentences, and three sentence vectors are visualized. As shown in Fig. 1, the original PLM model has an obvious narrow cone phenomenon, and unsupervised SimCSE shows that all directions are evenly distributed, but there is no multi-domain feature. The CCDC avoids the narrow cone phenomenon and also shows some multi-domain characteristics. It has the most appropriate representation of each domain and has some degree of differentiation between different domains.

Singular Values Decay. In addition, in singular value analysis (Fig. 4), the large gap between the head singular value of the traditional PLM and other singular values is well narrowed in unsupervised SimCSE, while the CCDC model supports sentence representation of multiple domains while still maintaining the uniform characteristics of singular values. In terms of homogeneity of the singular value distribution, the CCDC is comparable to unsupervised SimCSE.

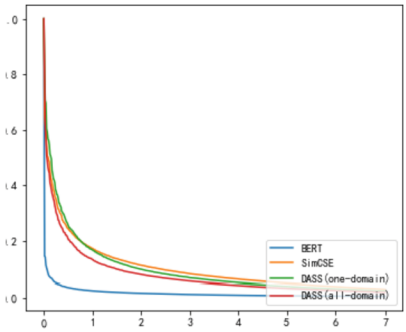

Fig. 4. Singular analysis on BERT, SimCSE, and CCDC (one-domain/all-domain)

6 Conclusion

The paper proposes the CCDC, a multi-domain enhanced contrastive learning sentence embedding framework, which uses a pseudo-NLI data generator to obtain a multi-domain sentence representation model that significantly outperforms the baseline model (BERT) and the strong baseline model (unsupervised SimCSE). Deep analysis shows that the CCDC framework solves the anisotropy and eigenvalue attenuation problems well.

Future research will focus on the knowledge mining in comparative learning. As shown in the preceding analysis, the CCDC performance is less than 50% in PASW-X. It should be able to perform hard-negative mining or multi-view mining based on knowledge graph or terminology knowledge [24] on hard cases such as "Scotland to England" and "England to Scotland" in PASW-X.

7 Appendix

7.1 CCDC with Different PLM and Different Pooling Layer

For comparison of different model architectures, model sizes, and pooling types, [27] is used as a reference, which provides 3 types of ALBERT and 3 types of Roberta based pre-trained models for Chinese. And the pooling type could be mean/cls/max, respectively indicating average pooling, class token pooling, and max pooling. Table 6 lists 19 different PLM + Pooling layer results($6 * 3 + 1 = 19$).

As can be seen, ALBERT is not as good as BERT-base, even with large model ,due to parameter sharing. Roberta large achieved all-domain best performance, and mean pooling achieved most of the best performance in most domains, while ATEC, BQ, and LCQMC offered the best performance with the max pooling layer.

7.2 Case Analysis

The sentence-level similarity of the corresponding three groups of models was calculated, and it was found that: (1) BERT-base, regardless of semantically related or not, similarity score are all over 0.93. (2) For the unsupvised SimCSE model, similarity score is 0.90 vs 0.86 for semantic related or not. (3) For the CCDC model, similar score is 0.93 vs 0.81. The CCDC model has better discrimination than BERT and Unsupervised SimCSE, as can be seen in Table 7

Table 6. CCDC results with different PLMs and different pooling layer

	Batch-size	Pooler	ATEC	BQ	LCQMC	PAWSX	STS-B	Average
Bert-base-chinese	128	Mean	0.46	0.66	0.74	0.35	0.71	0.58
Albert-tiny	128	Mean	0.36	0.58	0.64	0.19	0.63	0.48
Albert-tiny	128	cls	0.36	0.58	0.64	0.2	0.63	0.48
Albert-tiny	128	Max	0.34	0.55	0.64	0.15	0.61	0.46
Albert-base	128	Mean	0.4	0.63	0.68	0.19	0.63	0.51
Albert-base	128	cls	0.4	0.62	0.68	0.18	0.63	0.5
Albert-base	128	Max	0.38	0.60	0.68	0.17	0.62	0.49
Albert-large	96	Mean	0.41	0.64	0.68	0.17	0.63	0.51
Albert-large	96	cls	0.40	0.62	0.66	0.17	0.62	0.49
Albert-large	96	Max	0.40	0.63	0.65	0.17	0.62	0.49
Roberta-tiny	128	Mean	0.38	0.58	0.66	0.14	0.62	0.48
Roberta-tiny	128	cls	0.37	0.58	0.64	0.14	0.60	0.47
Roberta-tiny	128	Max	0.36	0.58	0.65	0.14	0.61	0.47
Roberta-base	128	Mean	0.46	0.67	0.75	0.46	0.69	0.61
Roberta-base	128	cls	0.45	0.66	0.74	0.46	0.67	0.60
Roberta-base	128	Max	0.45	0.66	0.74	0.44	0.67	0.59
Roberta-large	96	Mean	0.47	0.67	0.74	**0.48**	**0.72**	0.62
Roberta-large	96	cls	0.47	**0.68**	0.76	0.39	0.7	0.6
Roberta-large	96	Max	**0.48**	**0.68**	**0.77**	0.45	0.7	0.62

Table 7. Case Analysis for BERT-base, Unsup-SimCSE, and CCDC, Label = 1 is semantically identical and vice versa, label = 0

Sentence-a	Sentence-b	Label	BERT	SimCSE	CCDC
1560年10月，他在巴黎秘密会见了英国大使Nicolas Throckmorton，要求他通过苏格兰返回英国。	1560年10月，他在巴黎秘密会见了英国大使尼古拉斯·斯罗克莫顿，并要求他通过英格兰返回苏格兰的护照。	0	0.9789	0.8683	0.8174
(Related Translation)In October 1560, he met secretly in Paris ... to return to England via Scotland.	In October 1560 he met secretly in Paris ... to return to Scotland via England.	–	–	–	–
1975年的NBA赛季 - 76赛季是全美篮球协会的第30个赛季。	1975-76赛季的全国篮球协会是NBA的第30个赛季。	1	0.9697	0.906	0.9335
(Related Translation)The 1975 NBA season-76 was the 30th season of the National Basketball Association.	The 1975-76 season of the National Basketball Association was the NBA's 30th season.	–	–	–	–
还有具体的讨论，公众形象辩论和项目讨论。	还有公开讨论，特定档案讨论和项目讨论。	0	0.9399	0.6389	0.70
(Related Translation)There are also specific discussions, public image debates, and project discussions.	There are also open discussions, file-specific discussions, and project discussions.	–	–	–	–

References

1. Aitchison, L.: Infonce is a variational autoencoder. arXiv preprint arXiv:2107.02495 (2021)
2. Chen, J., Chen, Q., Liu, X., Yang, H., Lu, D., Tang, B.: The BG corpus: a large-scale domain-specific Chinese corpus for sentence semantic equivalence identification. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 4946–4951 (2018)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A Simple Framework for Contrastive Learning of Visual Representations, pp. 1597–1607 (2020). http://proceedings.mlr.press/v119/chen20j.html
4. Financial, A.: Ant Financial Artificial Competition (2018)
5. Gao, T., Yao, X., Chen, D.: SimCSE: Simple Contrastive Learning of Sentence Embeddings. arXiv:2104.08821 [cs] (2021). zSCC: 0000087
6. Gillick, D., et al.: Learning dense representations for entity retrieval, pp. 528–537 (2019). https://www.aclweb.org/anthology/K19-1049
7. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17–22 June 2006, New York, pp. 1735–1742. IEEE Computer Society (2006). https://doi.org/10.1109/CVPR.2006.100
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2015)
9. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R.: Albert: a lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942 (2019)
10. Liu, X., et al.: Lcqmc: a large-scale Chinese question matching corpus. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 1952–1962 (2018)
11. Liu, Y., et al.: Roberta: a robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
12. Meng, Y., et al.: COCO-LM: correcting and contrasting text sequences for language model pretraining. arXiv preprint arXiv:2102.08473 (2021)

13. Reimers, N., Beyer, P., Gurevych, I.: Task-Oriented Intrinsic Evaluation of Semantic Textual Similarity, pp. 87–96 (2016). https://www.aclweb.org/anthology/C16-1009
14. Reimers, N., Gurevych, I.: Sentence-BERT: sentence embeddings using Siamese BERT-networks, pp. 3982–3992 (2019). https://doi.org/10.18653/v1/D19-1410
15. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
16. Su, J., Cao, J., Liu, W., Ou, Y.: Whitening sentence representations for better semantics and faster retrieval. arXiv preprint arXiv:2103.15316 (2021)
17. Sun, X., Sun, S., Yin, M., Yang, H.: Hybrid neural conditional random fields for multi-view sequence labeling. Knowl. Based Syst. **189**, 105151 (2020)
18. Vaswani, A., et al.: Attention is all you need. arXiv:1706.03762 [cs] (2017). zSCC: 0033821
19. Wang, B., et al.: On position embeddings in Bert. In: International Conference on Learning Representations (2020)
20. Wang, L., Huang, J., Huang, K., Hu, Z., Wang, G., Gu, Q.: Improving neural language generation with spectrum control (2020). https://openreview.net/forum?id=ByxY8CNtvr
21. Williams, A., Nangia, N., Bowman, S.: A broad-coverage challenge corpus for sentence understanding through inference, pp. 1112–1122 (2018). https://doi.org/10.18653/v1/N18-1101
22. Yang, H., Chen, J., Zhang, Y., Meng, X.: Optimized query terms creation based on meta-search and clustering. In: 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 2, pp. 38–42. IEEE (2008)
23. Yang, H., Deng, Y., Wang, M., Qin, Y., Sun, S.: Humor detection based on paragraph decomposition and BERT fine-tuning. In: AAAI Workshop 2020 (2020)
24. Yang, H., Xie, G., Qin, Y., Peng, S.: Domain specific NMT based on knowledge graph embedding and attention. In: 2019 21st International Conference on Advanced Communication Technology (ICACT), pp. 516–521. IEEE (2019)
25. Yang, Y., Zhang, Y., Tar, C., Baldridge, J.: PAWS-X: a cross-lingual adversarial dataset for paraphrase identification. arXiv preprint arXiv:1908.11828 (2019). zSCC: NoCitationData[s0]
26. Zhang, D., et al.: Pairwise Supervised Contrastive Learning of Sentence Representations, p. 13, zSCC: 0000001
27. Zhang, N., et al.: CBLUE: A Chinese Biomedical Language Understanding Evaluation Benchmark. arXiv preprint arXiv:2106.08087 (2021)