Cross-domain Sentiment Classification based on Adaptive Center Contrastive Learning

CS224N Natural Language Processing with Deep Learning Project

Xinqi Fan xinqifan@stanford.edu

Background & Problem

Sentiment classification
- To recognize affects from natural language
- Sentiment classes:
  - Positive
  - Negative
  - ~~Neutral~~

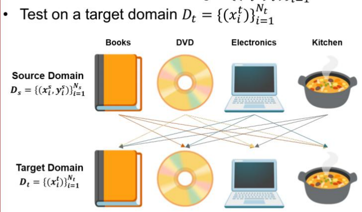
Domain adaptation (DA)
- Use different words in different context
- Significant performance decrease
- Main solution: domain invariant features

Problem
- To learn semantically meaningful representations and minimize intra-class variance: center contrastive learning (CCL)
- To reduce the domain representation discrepancy: adaptive contrastive learning (ACL) strategy

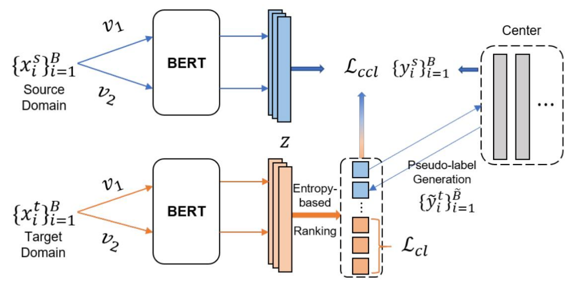
Dataset & Task

Amazon review dataset contains products reviews and their sentiment labels in 4 domains: books (B), DVD (D), electronics (E), and kitchens (K).

DA task:
- Train on a source domain $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$
- Test on a target domain $D_t = \{(x_i^t)\}_{i=1}^{N_t}$



Method



- **Data augmentation:**
  - Back translation: translate to another language, and then translate back
  - Offline (online data augmentation is time-consuming)
- **Baseline**:
  - Pretrained bidirectional encoder representations from transformers (BERT)
  - BERT base model – uncased from Hugging Face
  - Two projection layers' output as the representation z
  - Contrastive learning (CL) loss:

$$\mathcal{L}_{\mathrm{cl}} = -\sum_{i \in I} \log \frac{\exp\left(z_i \cdot z_{j(i)}/\tau\right)}{\sum_{a \in A(i)} \exp\left(z_i \cdot z_a/\tau\right)}$$

- **Center contrastive learning (CCL)**:
  - To learn semantically meaningful representations, we used supervised CL
  - To minimize intra-class variance, we modified the supervised CL by introducing two learnable center embeddings
  - CCL loss:

$$\mathcal{L}_{\mathrm{ccl}} = \sum_{i \in I} \frac{-1}{|P(c_i)|} \sum_{p \in P(c_i)} \log \frac{\exp\left(c_i \cdot z_p/\tau\right)}{\sum_{b \in B} \exp\left(c_i \cdot z_b/\tau\right)}$$

- **Adaptive contrastive learning (ACL):**
  - To reduce the domain representation discrepancy
  - Obtain entropy for all target domain samples, and select high confidence ones

$$E_i = \sum_c -z_i^c log(z_i^c)$$

  - Generate pseudo-labels by computing the similarity with centers

$$\tilde{y}_i^t = \mathrm{argmax}_k \{c_1 \cdot z_p, \ldots, c_k \cdot z_p, \ldots, c_K \cdot z_p\}$$

  - Compute CCL for samples with pseudo-labels

Experiment & Analysis

- **Quantitative results for data augmentation**
  - Back translation has the best result
  - Keep same semantic meaning

Data augmentation	Accuracy (%)
Back trans.	92.41
Random aug.	91.74
Contextual aug.	92.35

- **Quantitative ablation study**
  - ACCL is effective
  - K->E: 92.27% to 92.41%

CCL	ACL	Accuracy (%)
✗	✗	92.27
✓	✗	92.30
✓	✓	92.41

- **Comparison with other Methods**
  - Preform better when target domain is E or K
  - Best average result

S → T	Accuracy (%)								
	DANN	PBLM	HATN	ACAN	IATN	BERT	DAAT	CLIM	ACCL
D → B	81.70	82.50	86.30	82.35	87.00	89.40	90.86	91.25	**91.53**
E → B	78.55	71.40	81.00	79.75	81.80	86.50	88.91	89.08	**89.24**
K → B	79.25	74.20	83.30	80.80	84.70	87.55	87.98	**88.52**	88.41
B → D	82.30	84.20	86.10	83.45	86.80	88.96	89.70	**90.21**	89.97
E → D	79.70	75.00	84.00	81.75	84.10	87.95	90.13	**91.53**	91.42
K → D	80.45	79.80	84.50	82.10	84.10	87.30	88.81	**89.90**	89.37
B → E	77.60	77.60	85.70	81.20	86.50	86.15	89.57	91.92	**92.74**
D → E	79.70	79.60	85.60	82.80	86.90	86.55	89.30	90.15	**91.03**
K → E	86.65	87.10	87.00	86.60	87.60	90.45	91.72	92.27	**92.41**
B → K	76.10	82.50	85.20	83.05	85.90	89.05	90.75	**93.20**	92.67
D → K	77.35	83.20	86.20	78.06	85.80	87.53	90.50	91.89	**92.38**
E → K	83.95	87.80	87.90	83.35	88.70	91.60	93.18	93.21	**93.50**
Average	80.29	80.40	85.10	82.15	85.90	88.25	90.12	91.09	**91.22**

- **Qualitative analysis**
  - Explicit sentiment words: easy
  - No explicit sentiment words: hard, but we can do it correctly

Review	w/o ACCL	w/ ACCL	Ground truth
It certainly did. Excellent color qualit	+	+	+
I purchased this for my wife who wanted a case for her	-	+	+

Conclusion & Future Work

- **Conclusion**
  - CCL is useful for learning semantically meaningful features
  - Shared centers learned by both source and target domains are useful for the DA task
- **Future work**
  - Repeat experiments to reduce the stochasticity
  - Better center initialization other than random1

Reference:

[1] Tian Li, Xiang Chen, and et al. Cross-domain sentiment classification with contrastive learning and mutual information maximization. In IEEE International Conference on Acoustics, Speech and Signal Processing, pages 8203–8207. IEEE, 2021.

[2] Chunning Du, Haifeng Sun, Jingyu Wang, and et al. Adversarial and domain-aware bert for cross-domain sentiment analysis. In Proceedings of the 58th Annual Meeting of the Association of Computational Linguistics, pages 4019–4028, 2020.

[3] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. Advances in Neural Information Processing Systems, 33:18661–18673, 2020.