SUPPLEMENTARY MATERIALS: UNCOVERING THE EFFECTIVENESS OF CALIBRATION ON OPEN INTENT CLASSIFICATION

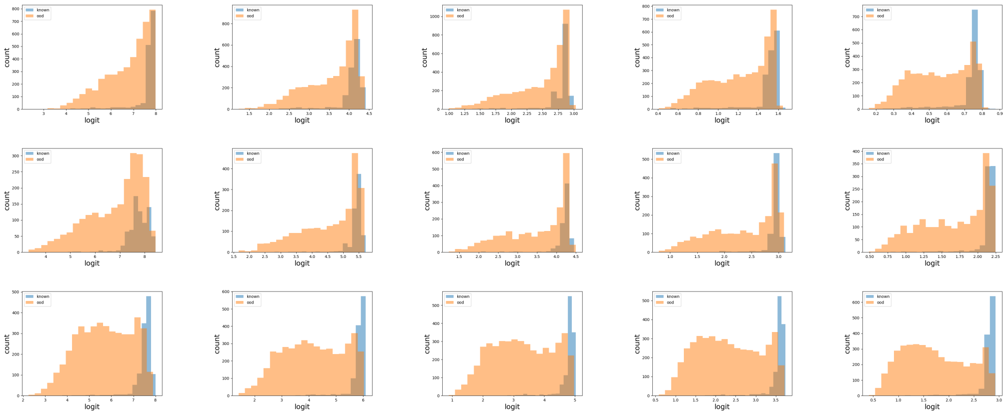
Anonymous authors
Paper under double-blind review

A EFFECTIVENESS OF CALIBRATION AT REDUCING BIAS OF CROSS-ENTROPY LOSS

To provide more detailed elaborations along with Section 4, we provide every experiment results on three datasets (STACKOVERFLOW, BAKINING, OOS) under various KLR and calibration strength levels. We provide additional visualizations of the maximum logit vector with KLR rate(i.e., 25%, 50%, 75%) in Figure 1, Figure 2, and Figure 3, as explained in Section 4.

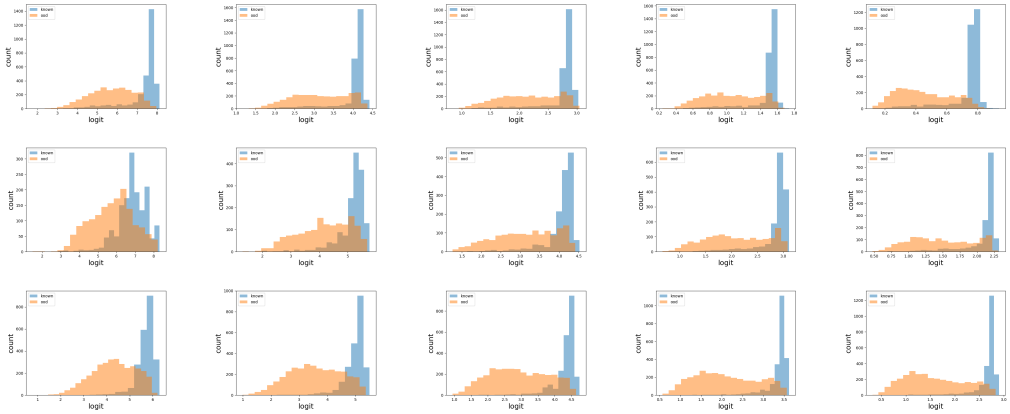
B DETAILED ILLUSTRATIONS ON REPRESENTATION SIMILARITY

This section describes analyses on representation similarities, established by CKA, in our experiments as described in Section 6 of main manuscript. Note that every experiments are conducted on fixed random seeds with the same KLR rate of 25%. Figure 4 and Figure 5 show the layer-wise representation similarity on STACKOVERFLOW in C-LC and C-ADB. Figure 6 and Figure 7 show the layer-wise representation similarity on BANKING in C-LC and C-ADB. Figure 8 and Figure 9 show the layer-wise representation similarity on OOS in C-LC and C-ADB.



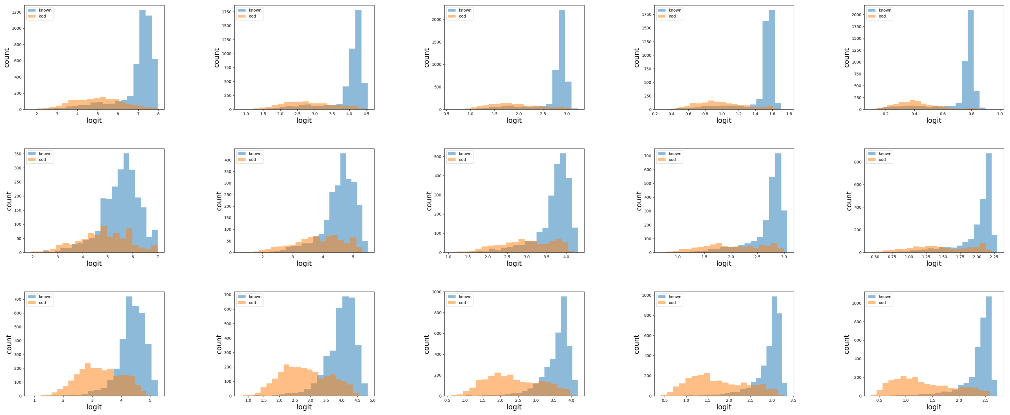
(a) 0 (b) 0.2 (c) 0.5 (d) 0.8 (e) 0.9

Figure 1: Distribution of maximum value of logit between known intent (blue) and unknown (orange) intent samples under KLR of 25%. From top to bottom, each row indicates the dataset type of STACKOVERFLOW, BANKING, and OOS. From left to right, each column indicates calibration strengths of 0, 0.2, 0.5, 0.8, and 0.9. Note that test samples in each dataset are utilized in the analyses.



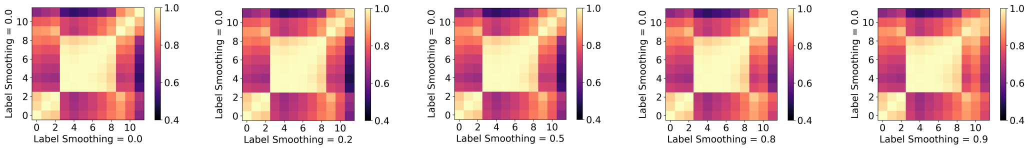
(a) 0 (b) 0.2 (c) 0.5 (d) 0.8 (e) 0.9

Figure 2: Distribution of maximum value of logit between known intent (blue) and unknown (orange) intent samples under KLR of 50%. From top to bottom, each row indicates the dataset type of STACKOVERFLOW, BANKING, and OOS. From left to right, each column indicates calibration strengths of 0, 0.2, 0.5, 0.8, and 0.9. Note that test samples in each dataset are utilized in the analyses.



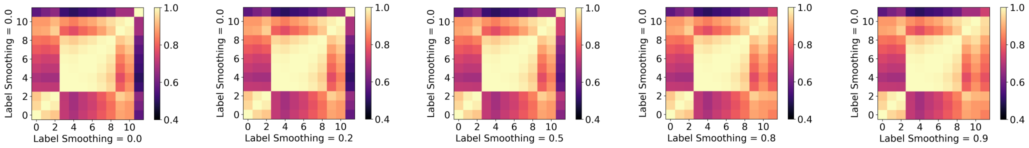
(a) 0 (b) 0.2 (c) 0.5 (d) 0.8 (e) 0.9

Figure 3: Distribution of maximum value of logit between known intent (blue) and unknown (orange) intent samples under KLR of 75%. From top to bottom, each row indicates the dataset type of STACKOVERFLOW, BANKING, and OOS. From left to right, each column indicates calibration strengths of 0, 0.2, 0.5, 0.8, and 0.9. Note that test samples in each dataset are utilized in the analyses.



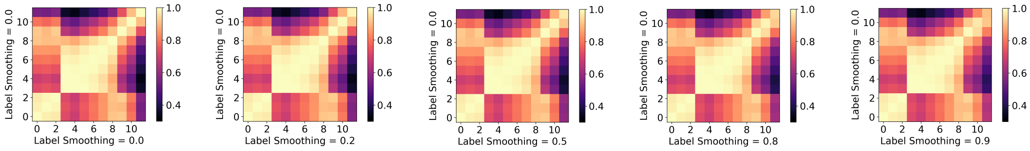
(a) 0 v. 0 (b) 0 v. 0.2 (c) 0 v. 0.5 (d) 0 v. 0.8 (e) 0 v. 0.9

Figure 4: Representation similarity between C-LC without calibration and C-LC models with various calibration strengths, where the model is trained with STACKOVERFLOW dataset under KLR rate of 25%. The results show that high-level representations, which are known to be contextual understanding, become different along with calibration strengths. Note that values in the subcaption imply calibration strengths.



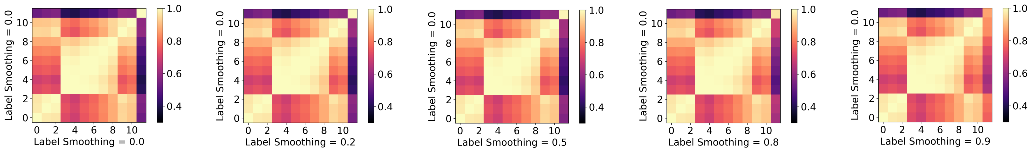
(a) 0 v. 0 (b) 0 v. 0.2 (c) 0 v. 0.5 (d) 0 v. 0.8 (e) 0 v. 0.9

Figure 5: Representation similarity between C-ADB without calibration and C-ADB models with various calibration strengths, where the model is trained with STACKOVERFLOW dataset under KLR rate of 25%. The results show that high-level representations, which are known to be contextual understanding, become different along with calibration strengths. Note that values in the subcaption imply calibration strengths.



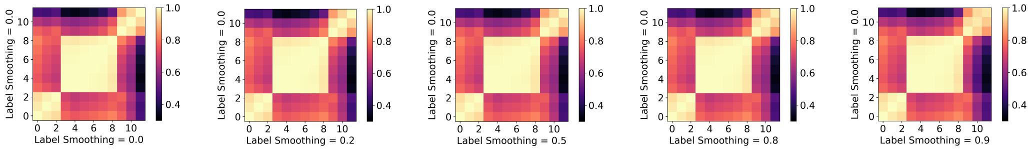
(a) 0 v. 0 (b) 0 v. 0.2 (c) 0 v. 0.5 (d) 0 v. 0.8 (e) 0 v. 0.9

Figure 6: Representation similarity between C-LC without calibration and C-LC models with various calibration strengths, where the model is trained with BANKING dataset under KLR rate of 25%. The results show that high-level representations, which are known to be contextual understanding, become different along with calibration strengths. Note that values in the subcaption imply calibration strengths.



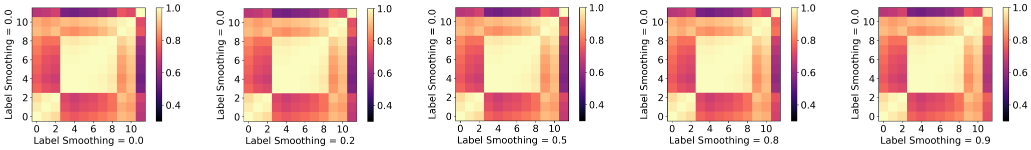
(a) 0 v. 0 (b) 0 v. 0.2 (c) 0 v. 0.5 (d) 0 v. 0.8 (e) 0 v. 0.9

Figure 7: Representation similarity between C-ADB without calibration and C-ADB models with various calibration strengths, where the model is trained with BANKING dataset under KLR rate of 25%. The results show that high-level representations, which are known to be contextual understanding, become different along with calibration strengths. Note that values in the subcaption imply calibration strengths.



(a) 0 v. 0 (b) 0 v. 0.2 (c) 0 v. 0.5 (d) 0 v. 0.8 (e) 0 v. 0.9

Figure 8: Representation similarity between C-LC without calibration and C-LC models with various calibration strengths, where the model is trained with OOS dataset under KLR rate of 25%. The results show that high-level representations, which are known to be contextual understanding, become different along with calibration strengths. Note that values in the subcaption imply calibration strengths.



(a) 0 v. 0 (b) 0 v. 0.2 (c) 0 v. 0.5 (d) 0 v. 0.8 (e) 0 v. 0.9

Figure 9: Representation similarity between C-ADB without calibration and C-ADB models with various calibration strengths, where the model is trained with OOS dataset under KLR rate of 25%. The results show that high-level representations, which are known to be contextual understanding, become different along with calibration strengths. Note that values in the subcaption imply calibration strengths.