國立台北大學資訊工程學系專題報告

基於對比學習的入侵偵測模型

Intrusion Detection Model Based on Contrastive Learning

專題組員:曾慶哲、王致越、陳育昇、鄭宇謙

專題編號: PRJ-NTPUCSIE-112-007

執行期間:112 年 7 月 至 113 年 6 月

1. 摘要

基於深度學習模型在分類上取得的巨大成果，利用深度學習模型建立的入侵偵測模型已然成為了因應網路攻擊的重要方法。然而這樣的入侵偵測模型往往需要大量的有標籤資料，才得以訓練出具有一定成效的結果，但這些有標籤資料的取得不易，反而是無標籤的資料較易取得。

自監督式學習能夠非常有效地利用無標籤資料，而其中自監督式學習[1]的分支，對比學習則更符合我們建立入侵偵測模型的需求。且因為多分類能處理多種可能輸出類別，提供更細緻的預測，適用於複雜的入侵偵測場景，且可以為本次研究提供更高的挑戰性，因此本研究嘗試利用對比學習的特性結合多分類，來觀察在此領域中，是否也能在僅使用了部分有標籤資料的情況下，使用大量的無標籤資料來提高分類的準確率。

2. 簡介

2.1 研製背景

隨著網路時代的進步，一個成熟的伺服器每天都會遭受數以萬計的攻擊流量，而正常的流量更是攻擊流量的好幾倍。蒐集這些流量相對容易，然而要將這些流量進行標註則需要花費更多的成本，如何充分地使用這些無標籤資料將會是個優秀的解方。

對比學習是一種通過計算對比損失 (Contrastive Loss) 來調整模型參數的方法。這種方法的主要目的是通過最大化不同分類之間的對比，來提高模型在自監督學習中的效果。在影像分類問題上，對比學習已經展示了優異的效果，而且即使在不用構建超大型模型的情況下，也能取得顯著的成果。由於對比學習的設計，並未特別要求資料必須有影像資料的特性，因此有足夠泛用性潛力，使用在入侵偵測模型的優化，增加正常與攻擊行為，在訓練模型參數中的區別。

2.2 動機

有標籤數據的收集和標註過程通常費時費力或成本高昂，相比之下，無標籤數據通常更容易獲取，且數量龐大。因此，如何有效利用這些豐富的無標籤數據來提升模型性能成為了一個重要的研究課題。

而若利用 SMOTE[2]，VAE[3]，GAN[4]等過採樣技術在對訓練資料量進行擴增時，在有標籤數據不足的情況下，可能會因為數據多樣性不足而導致擴增樣本品質較低的問題。而對比學習作為一種有效的自監督學習方法，能夠充分利用真實無標籤數據來解決前段所提到樣本品質不足的問題。

因此，基於對比學習的研究在解決有標籤數據不足及樣本品質較低的問題上具有相當的潛力，我們決定以此作為本次專題目標。

2.3 相關研究文獻探討

2.3.1 Momentum Contrast[5]

Momentum Contrast（MoCo）的特色為其使用 momentum encoder，momentum encoder 使用了普遍的兩種方法，EndToEnd 跟 MemoryBank 的優點，並改善了其缺點。EndToEnd 無法支撐過大的 batch size，而 MemoryBank 雖能解決此問題，但卻會導致一致性不佳的問題。momentum encoder 不使用反向傳播而是以動量的方式更新。

$$\theta_{\mathrm{k}} \leftarrow m\theta_{\mathrm{k}} + (1-m)\theta_{\mathrm{q}}.$$

(圖一) momentum encoder 計算公式

θ_q 是目標圖像 Encoder 的 參數，而 θ_k 是正負樣本 Encoder 的參數，動量式的更新可以解決一致性的問題，也因為不須使用反向傳播，所以可用較大量的負樣本。

2.3.2 Barlow Twins[6]

Barlow Twins 在概念上簡單且易於實現，並且在學習上是有用的表示，而不是微不足道的解。與其他方法相比，比如 MoCo 使用 queue 的方式去模擬很大的 batch size，它在樣本上不太注重，也無需使用負樣本進行訓練，不需要大批次，因此不受限於批量大小，可以在小批量資料上進行訓練，也不需要任何非對稱機制，如預測網絡，動量編碼器等。Barlow Twins 在低數據狀態下的半監督分類方面優於當時 ImageNet 上以前的方法，並且在具有線性分類器的 ImageNet 分類以及分類和對象檢測的轉移任務方面與當時的最新技術相媲美。

2.3.3 半監督學習[7]

半監督學習是一種利用有標籤和無標籤數據來構建模型的學習方式。半監督學習方法能通過使用額外的無標籤數據來提高學習性能。提供一種探索無標籤數據中潛在模式的方法，減少對大量標籤的需求。半監督學習的主要方式是以有標籤資料為基底，對無標籤數據生成偽標籤來擴展訓練數據集並提升模型性能。

2.3.4 UNSW-NB15 資料集[8]

UNSW-NB15 資料集是由澳洲新南威爾斯大學坎培拉網路範圍實驗室的 IXIA PerfectStorm 工具所建立。擷取 2015 年 1 月 22 日(16 小時)及 2015 年 2 月 17 日（15 小時） 間的網路流量，形成了具有 49 個特徵和 2,540,044 筆紀錄之資料集。其中第 48 及 49 個特徵為標籤，第 48 個特徵有 9 種網路攻擊行為之標籤類別，第 49 特徵有 2 種標籤，0 代表正常，1 代表網路攻擊。UNSW-NB15 具多樣性，使之成為一個完整的實驗資料集，以幫助研究人員驗證處理複雜多樣的網路安全問題。

2.4 目標

在本次研究中我們分別利用了兩種對比學習模型 Momentum Contrast（MoCo)和 Barlow Twins 來建立兩個入侵偵測模型，期待能夠利用對比學習來有效地從無標籤數據中學習有用的特徵，從而減少對大量標籤數據的依賴，降低數據標註的成本和時間。並為網路安全的領域提供一個能降低構建和維護入侵偵測系統的成本的方案與參考。

3. 專題進行方式

3.1 資料前處理

I. 進行資料讀取與合併。

II. 根據特徵的類型(名詞、整數、

二進位和浮點數)分組，以便於後續的資料轉換和處理。

III. 對於名詞特徵，去除前後空白；對於數字特徵，將無效值設置為 NaN，並進一步替換為 0。

IV. 對名詞特徵進行 One-Hot 編碼，隨後應用 PCA 進行降維。

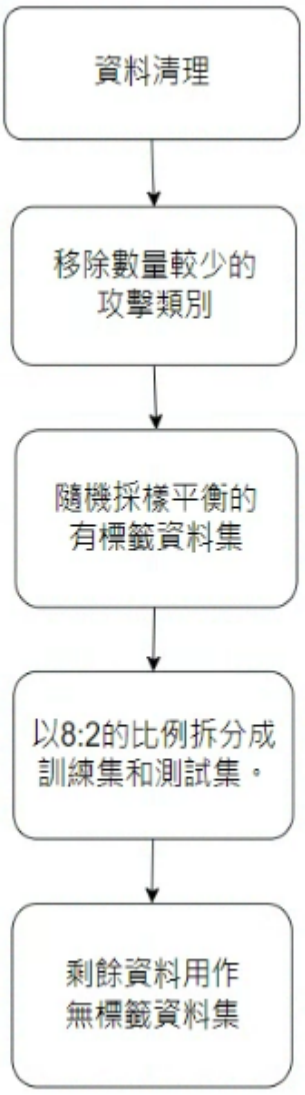
V. 因為特定的攻擊類別資料在原始資料集內的數量過少，故移除掉這些攻擊類別 ('Analysis' , 'Backdoor', 'Shellcode' , 'Worms')。

VI. 對每個類別進行處理。抽取 10000 筆作為訓練集，並將剩餘的數據用作無標籤資料集。否則就將該類別的所有數據都用作訓練。

VII. 將各個類別的訓練數據合併。

VIII. 減少無標籤資料集中的正常資料數量至 220000，避免過度的不平衡。

IX. 將訓練集按 8:2 的比例拆分為訓練集和測試集。因此，我們的訓練集與測試集將會為已平衡的資料。



(圖二)前處理流程圖

3.2 將資料轉換成圖片

因為我們採用 CNN 的架構來進行實驗，將資料轉換為圖片以便後續使用。

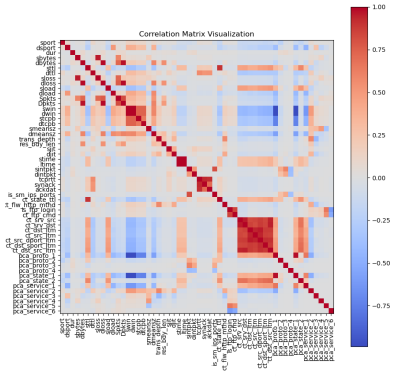
I. 建構相關係數矩陣(圖三)

II. 選擇相關係數總和最大的五個特徵

III. 將這五個特徵加入圖片中心與其四周的像素點

IV. 圖片其他像素排入與相鄰點相關係數總和最大的剩餘特徵

V. 保存轉換後的圖像。

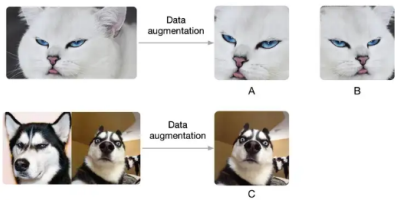


(圖三) 相關係數矩陣

3.3 利用對比學習進行實驗

I. 讀取無標籤資料。

II.對資料作 data augmentation 來建立正負樣本對。data augmentation 能藉由旋轉，局部放大，變色，去雜訊，高斯模糊等方式將原來的一筆數據，擴增成 N 筆數據，目的是提升模型泛化性。

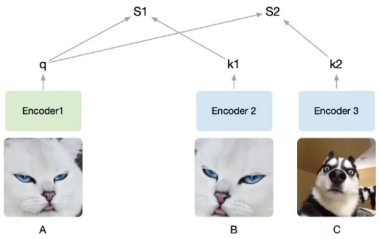


(圖四) Data Augmentation[9]

(圖 A,圖 B 互為正樣本，

而圖 C 為負樣本)

III. 利用 ResNet18[10]當作模型主幹進行特徵提取並根據 Contrastive loss 來更新參數。



(圖五) Pretext Task[9]

(圖片各自通過 encoder 得到 representation q 跟 K1 和 K2。而我們的目標就是盡可能提高 S1，並降低 S2。)

$$\mathcal{L}_{\mathrm{N}} = -\mathop{\mathbb{E}}_{X}\left[\log \frac{f_k(x_{t+k}, c_t)}{\sum_{x_j \in X} f_k(x_j, c_t)}\right]$$

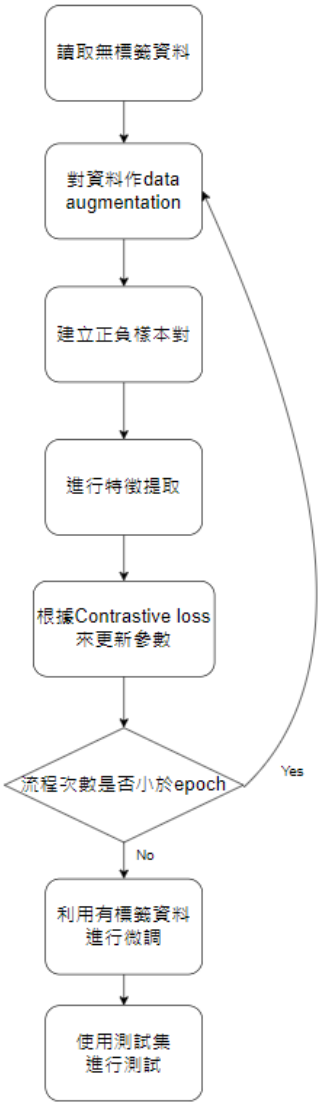
(圖六)MoCo Contrastive Loss

$$\mathcal{L}_{\mathcal{BT}} \triangleq \underbrace{\sum_i (1 - \mathcal{C}_{ii})^2}_{\text{invariance term}} + \lambda \underbrace{\sum_i \sum_{j \neq i} \mathcal{C}_{ij}{}^2}_{\text{redundancy reduction term}}$$

(圖七) Barlow Twins Contrastive Loss

IV.將全連階層改為一輸出為 6 的線性層架構，使其能夠預測資料集內的 6 個類別，並進行微調。

V. 使用測試集進行測試。



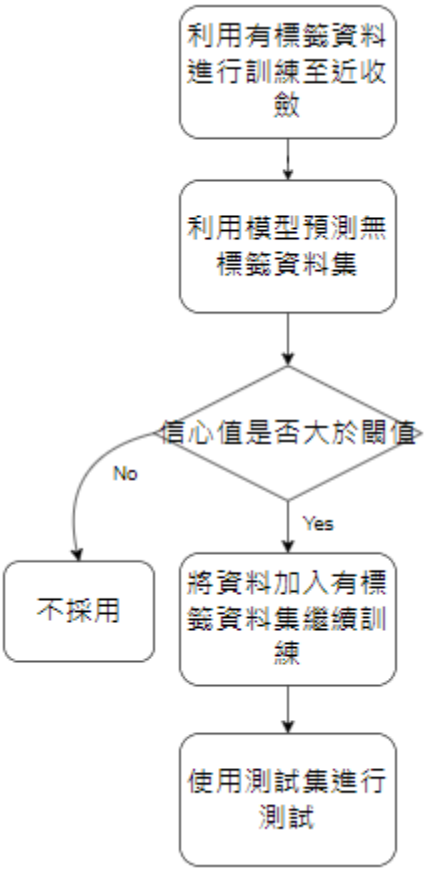
(圖八)對比學習流程圖

3.4 利用半監督學習進行實驗對照

I. 利用有標籤資料進行訓練至近收斂

II. 利用模型預測無標籤資料集

III. 若信心值大於閾值，將資料加入有標籤資料集繼續訓練



IV.使用測試集進行測試

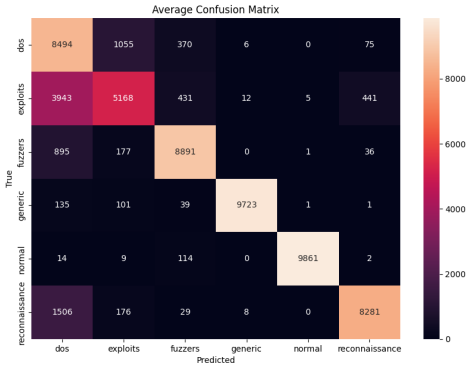
(圖九) 半監督學習流程圖

4. 主要成果與評估

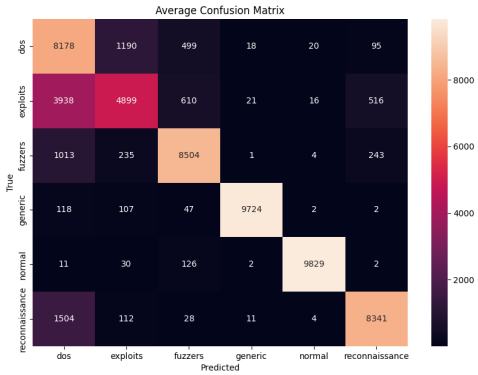
我們目前測試了多種圖像轉換（transform）方法的組合，在多次實驗和比較之後，我們發現調整對比度的轉換方法在我們的預訓練模型中顯示出了最好的成果，於是最後以對比度來作為 transform 的選擇，以下是採用 5-fold validation 並取平均的結果。

	Accuracy	Precision	Recall	F1 Score
MoCo	84.03%	86.23%	84.02%	84.17%
Barlow Twins	82.46%	84.37%	82.46%	82.52%
Semi supervised	83.93%	86.38%	83.92%	84.09%

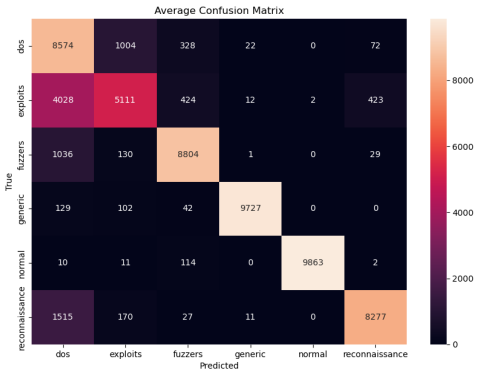
(表一) 表現結果



(圖十) MoCo 混淆矩陣



(圖十一) Barlow Twins 混淆矩陣



(圖十二) Semi-Supervised Learning 混淆矩陣

除了 MoCo 與 Barlow Twins 外，我們也使用了同為對比學習的 SimSiam[11]以及 SimClr[12]進行了實驗，然而此兩者都遇上了準確率相當低或是模型塌陷的情況，即使嘗試各種調適[13]後問題仍然無法獲得解決，所以並未採用其兩者數據。

由兩種對比學習算法與半監督結果的比較來看，兩者的結果數據並沒有取得明顯的差距，Barlow Twins 表現甚至不如半監督。對於 Barlow Twins 的情況，我們認為也許是算法天生不適用於此領域，而我們也認為若跳出圖像轉換的框架，開發出專屬於流量特徵的 transform 方式，並應用於對比學習之中的 data augmentation，也許能取得更加出色的結果。

5. 結語與展望

在本研究中，我們探討了使用對比學習中的 MoCo (Momentum Contrast) 和 Barlow Twins 兩種算法來對網路流量進行預測，判斷其屬於正常流量還是某種惡意攻擊。

對比學習作為自監督學習中的一種訓練方式，已經在圖像和語言等多個領域展現了其優越性。本研究嘗試將其應用於網路流量的入侵偵測中。雖然此兩種算法有著不錯的準確度，但並沒有與半監督學習的成績拉開差距，對此我們認為若找出一個更合適的 transform 方法來製造正樣本對的話，也許能較大程度地改善結果。

6. 銘謝

非常感謝教授在這段時間裡對專題研究提供的悉心指導和建議以及實驗室的資源和設備。無論是在研究方向的選擇還是算法改良的過程中，教授的建議都起到了至關重要的作用。

也感謝實驗室的學長姐們在研究過程中給予的幫助，使我們的專題研究得以如期順利完成。

7. 參考文獻

[1] Self-Supervised Learning: Generative or Contrastive
Xiao Liu; Fanjin Zhang; Zhenyu Hou; Li Mian; Zhaoyu Wang; Jing Zhang; Jie Tang

[2] SMOTE: Synthetic Minority Over-sampling Technique
N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer
arXiv:1106.1813

[3] VOS: a Method for Variational Oversampling of Imbalanced Data
Val Andrei Fajardo, David Findlay, Roshanak Houmanfar, Charu Jaiswal, Jiaxi Liang, Honglei Xie
arXiv:1809.02596

[4] Conditional Wasserstein GAN-based Oversampling of Tabular Data for Imbalanced Learning
Justin Engelmann, Stefan Lessmann
arXiv:2008.09202

[5] Barlow Twins: Self-Supervised Learning via Redundancy Reduction Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun,

Stéphane Deny
arXiv:2103.03230

[6] Momentum Contrast for Unsupervised Visual Representation Learning
Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, Ross Girshick
arXiv:1911.05722

[7] A Survey on Deep Semi-Supervised Learning
Xiangli Yang; Zixing Song; Irwin King; Zenglin Xu

[8] https://research.unsw.edu.au/projects/unsw-nb15-data

[9] https://blog.csdn.net/qq_38308388/article/details/129862346

[10] Deep Residual Learning for Image Recognition
Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun
arXiv:1512.03385

[11] Exploring Simple Siamese Representation Learning
Xinlei Chen, Kaiming He
arXiv:2011.10566

[12] A Simple Framework for Contrastive Learning of Visual Representations
Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton
arXiv:2002.05709

[13] How Does SimSiam Avoid Collapse Without Negative Samples? A Unified Understanding with Self-supervised Contrastive

Learning

Chaoning Zhang, Kang Zhang,

Chenshuang Zhang, Trung X.

Pham, Chang D. Yoo, In So

Kweon

arXiv:2203.16262