ネットワーク運用におけるデータ活用に向けて

アーキテクチャ科学研究系　小林 諭　福田 健介

転移学習によるネットワークログのテンプレート自動生成

NII

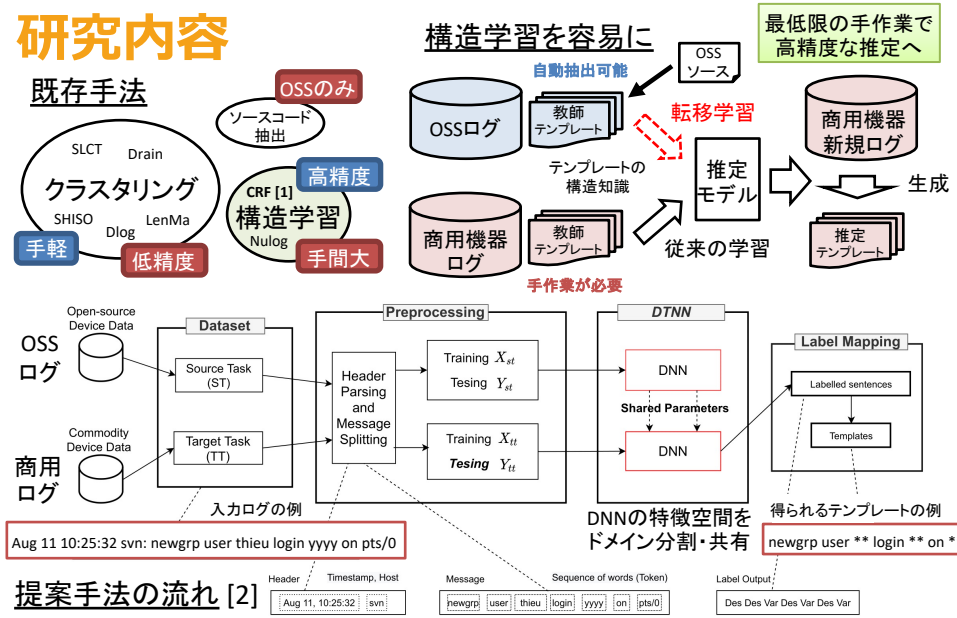
どんな研究？

システム障害の原因究明のため記録されるログデータは、SINETのような大規模ネットワークでは膨大な量です。これを機械的に解析するには、自由記述のログを出力フォーマットにより分類するテンプレート推定が必要です。この研究では特に、普段現れないエラーログなどのフォーマットも適切に推定可能なテンプレート推定技術を研究しています。

何がわかる？

ログテンプレート推定に転移学習を用いることで、教師データとして他のベンダ機器など推定対象と少し異なる群について得られたデータを活用することができます。手作業で作成する教師データが少数であっても他の群からの知識転移により高精度な推定が可能となり、商用の機器や非公開のソフトウェアについてもログの分類や解析が最低限の労力で実現できます。

研究内容



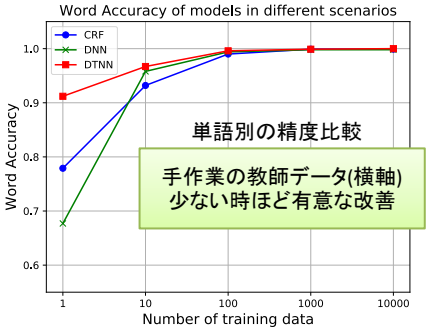
評価

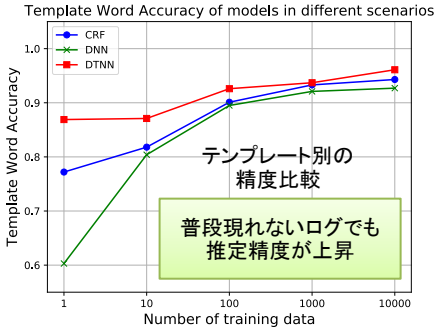
データセット

OSSログデータ: Vyatta 10万行
商用機器ログデータ: SINET4 7.5万行
(教師データは別日のログから手で作成)

手法比較

DTNN [2]: 提案手法 (知識転移あり)
DNN: 知識転移なし
CRF [1]: 知識転移なし、NNなし





今後の展開

- 処理の高速化 (既知のフォーマットに合致するログをスキップ)
- ログ解析フレームワーク amulog (https://github.com/cpflat/amulog)
  で利用可能に -> 因果解析・意味解析など発展的な解析での活用を容易に

[1] S. Kobayashi, K. Fukuda, H. Esaki, “Towards an NLP-based Log Template Generation Algorithm for System Log Analysis”, ACM CFI’14, p.4, 2014
[2] T. Nguyen, S. Kobayashi, K. Fukuda, “LogDTL: Network Log Template Generation with Deep Transfer Learning”, IFIP/IEEE AnNet’21, p.6, 2021

(本研究は総務省SCOPE #191603009の委託を受けたものです)