

Hierarchical Clustering Approach Based on Swarm Intelligence for Real-time Flow Classification

リアルタイムフロー分類のための群知能に基づく階層的クラスタリング手法

大阪大学 大学院情報科学研究科
情報ネットワーク学専攻 村田研究室
須恵 匠

2016/2/15



リアルタイムフロー分類への要求

ネットワークアプリケーションの種類が増加
ネットワークへの性能要求が多様化

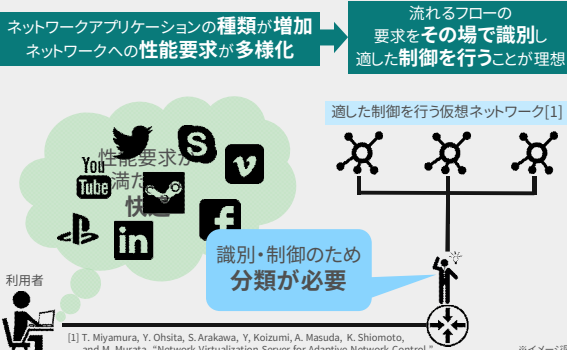
流れるフローの要求をその場で識別し
適した制御を行うことが理想

適した制御を行う仮想ネットワーク[1]

性能要求が満たない
利用者

識別・制御のため分類が必要

[1] T. Miyamura, Y. Ohsita, S. Arakawa, Y. Koizumi, A. Masuda, K. Shiimoto, and M. Murata, "Network Virtualization Server for Adaptive Network Control."



研究の背景

[2] A. Dainotti, A. Pescapé, and K. Claffy, "Issues and Future Directions in Traffic Classification."


従来の特徴量に基づくトラフィック分類手法 [2]

- フローの特徴量を観測
- 観測した特徴量をクラスタリング
 - 分類器を訓練

問題点

- 事前にフローの特徴量を確定することが前提
 - フロー完了時に確定=正確な特徴量
 - リアルタイムフロー分類に適用不可
 - フロー開始直後に確定=不正確な特徴量
 - 正確なフロー分類が不可

リアルタイムフロー分類へ適用可能な特徴量の正確さを自律的に判断するクラスタリング手法が必要



研究の目的と方針

目的

リアルタイムフロー分類を想定した正確で高速なオンラインクラスタリング手法の提案と評価

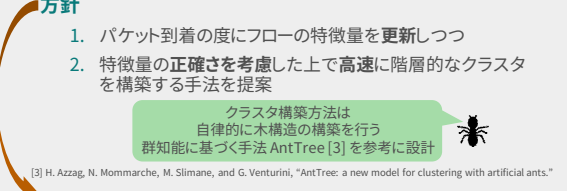
方針

- バケット到着の度にフローの特徴量を更新しつつ
- 特徴量の正確さを考慮した上で高速に階層的なクラスタを構築する手法を提案

クラスタ構築方法は自律的に木構造の構築を行う群知能に基づく手法 AntTree [3] を参考に設計

[3] H. Azzag, N. Mommarche, M. Slimane, and G. Venturini, "AntTree: a new model for clustering with artificial ants."

正確さと速度を自律的に両立しリアルタイムフロー分類へ適用可能なクラスタリングを実現



方針のイメージ

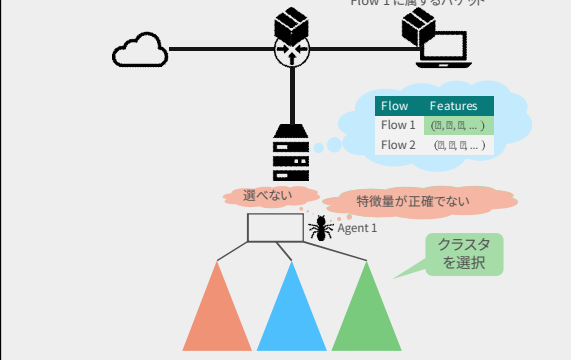
Flow 1 に属するバケット

Flow Features
Flow 1 (0, 0, 0, ...)
Flow 2 (0, 0, 0, ...)

選ばない
特徴量が正確でない

Agent 1

クラスタを選択



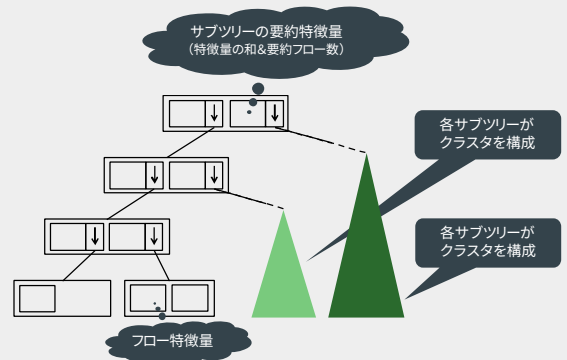
提案手法のデータ構造

サブツリーの要約特徴量 (特徴量の和&要約フロー数)

各サブツリーがクラスタを構成

各サブツリーがクラスタを構成

フロー特徴量



フローのクラスタリング 7

フローと対応するエージェントが
基準値 θ と閾値 δ を比較しながら木を移動し所属クラスタを探索

- $\theta = (\theta_{up} - \theta_{down}) \times \eta$ η : パラメータ、 θ_{up} : バケット数、 θ_{down} : 更新前後の特徴量差分
 - η バケット観測時と現在の特徴量の差分の推定値
- $\delta = \frac{\theta_{up}}{\theta_{down}}$ θ_{up} : 最類似クラスタとの距離
 θ_{down} : 2番目に似たクラスタとの距離
 - 最近傍クラスタとの類似度の確信度

エージェントの動作

- θ_{up} 、 θ_{down} がともに閾値 δ 、 θ_{down} よりも小さい場合
最類似クラスタに移動
- それ以外の場合閾値を緩和し
次の特徴量更新まで待機

評価環境 8

トラフィックデータ

- 研究室と外部ネットワークの中継点においてキャプチャした約1時間分のトラフィック

特徴量

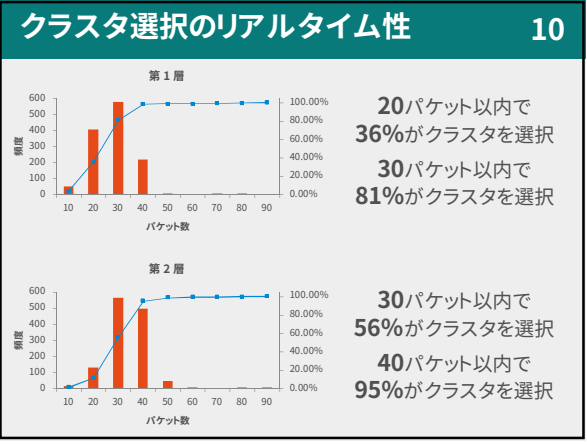
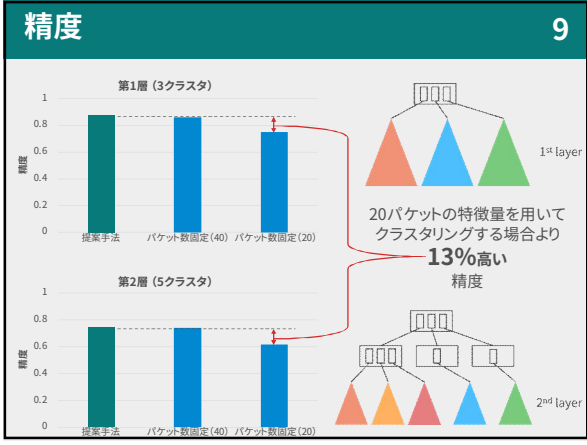
- パケットサイズの平均(上り方向・下り方向)

評価指標

- Overall Accuracy = $\frac{\sum_{i=1}^n \delta_i}{n}$ δ_i : オフラインでノードに属したフロー
 θ_i : オンラインでノードに属したフロー
 θ : 全フロー数
- フロー完了時の特徴量をオフラインでクラスタリングした結果を正解とした場合の精度を評価
- 第1層、第2層に到達した時点のバケット数
- クラスタ選択のリアルタイム性を評価

比較対象

- 固定数バケットからの特徴量でクラスタリングを行う場合



まとめ・今後の課題 11

まとめ

- リアルタイムフロー分類を想定した正確で高速なオンラインクラスタリング手法の提案と評価
 - 20バケットの特徴量でクラスタリングを行う場合と比較して **13%高い精度**を達成
 - 30**バケット以内で**81%**が第1層のクラスタを選択
 - 40**バケット以内で**95%**が第2層のクラスタを選択

今後の課題

- パラメータの調整による性能向上の検討
- 応用への適用の検討

Credits 12

- Icon made by <http://www.freepik.com>
- Freepik from <http://www.flaticon.com>
- Flaticon is licensed under <http://creativecommons.org/licenses/by/3.0/>