

群知能を用いたクラスタリング手法の トラフィック分類への適用と評価

須 恵 匠
大阪大学基礎工学部情報科学科
村田研究室

2014/2/21

特別研究報告発表会

1

研究背景

- トラフィックのアプリケーションごとの分類・識別がネットワーク管理において重要
 - 各アプリケーションの性能要求に応じたネットワーク制御
- ネットワークを利用するアプリケーションの多様化
 - WEB上で提供されるアプリケーションも多数
 - ポート番号ベースのアプリケーション識別は困難



現在ネットワーク内を流れているWEBアプリケーションのトラフィックを自動的に分類・識別する手法が必要

- 要件
 - WEBアプリケーションの性能要求に応じてトラフィックを分類可能
 - 新たに登場するWEBアプリケーションにも対応可能
 - 短い処理時間でWEBアプリケーションのトラフィックの分類が可能

2014/2/21

特別研究報告発表会

2

研究の目的とアプローチ

目的

- WEBアプリケーションのトラフィックを分類する手法

アプローチ

- トラフィックの統計情報をもとにクラスタリングを用いて分類
 - クラスタリングを行うことにより、似た統計的性質を持つトラフィックを集めたクラスタを構成することが可能
 - **構築されたクラスタの正確度・クラスタリングの速度は用いる統計情報とクラスタリング手法に依存**

検討内容

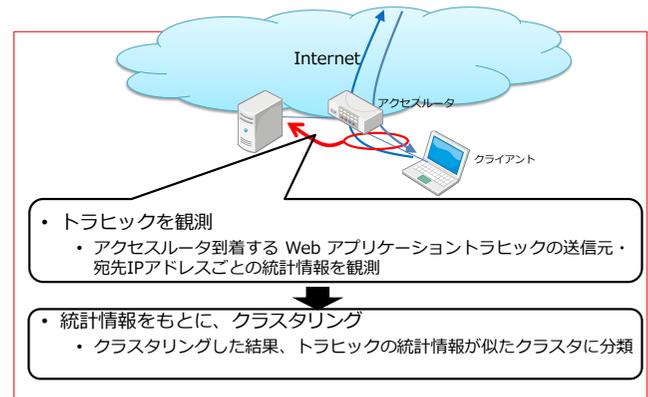
- WEB アプリケーショントラフィック分類に用いる統計情報の検討
- トラフィックの分類に適したクラスタリング手法の議論

2014/2/21

特別研究報告発表会

3

クラスタリングを用いたトラフィック分類の概要



2014/2/21

特別研究報告発表会

4

WEB アプリケーション分類に用いる統計情報

方針

- アプリケーションのネットワークへの性能要求が反映される統計情報を検討
 - インタラクティブ性：クライアントからのリクエストに応じて動作しているか(即応性の要求)
 - トラフィックの定常性：一定のトラフィックが連続で流れているか(帯域の要求)

用いた指標

- インタラクティブ性：
 - クライアント発ペイロード付パケット到着レートとサーバ発のペイロード付パケット到着レートの比
$$R = \log \frac{p^{DOWN}}{p^{UP}}$$
- トラフィックの定常性：
 - サーバ発のペイロード付パケット到着レートの変動係数
 - クライアント発のペイロード付パケット到着レートの変動係数
$$CV^{DOWN} = \frac{Std(p^{DOWN})}{Avg(p^{DOWN})}$$

$$CV^{UP} = \frac{Std(p^{UP})}{Avg(p^{UP})}$$

p^{DOWN} : 下りパケット到着レート
 p^{UP} : 上りパケット到着レート
Std: 標準偏差
Avg: 平均値

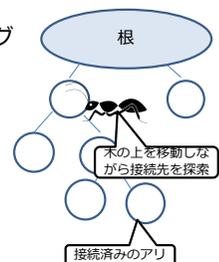
2014/2/21

特別研究報告発表会

5

クラスタリングアルゴリズム

- 群知能を用いた適応的なクラスタリング
- AntTree [1]
 - アリがつながって木構造を作る行動に着想を得た手法
 - 各アリが自律的に移動しながら、近隣のアリと自身を比較して木構造を構築
 - 最終的に根につながっている各部分木が一つのクラスタを構成
- 群知能以外のクラスタリング
- K-means+ +法 [2]
- 群平均法
 - 類似したクラスタを再帰的に結合してクラスタを生成する手法



[1] H. Azzag, N. Monmarche, M. Slimane and G. Venturini, "AntTree: a new model for clustering with artificial ants"
[2] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding"

2014/2/21

特別研究報告発表会

6

評価シナリオ

- 評価観点
 - トラフィックのアプリケーションの識別の正確性
 - 新たなアプリケーションが登場した場合への対応
 - 分類の速さ
- 評価シナリオ
 - 3種類のアプリケーションのトラフィックが流れる環境での評価
 - 5種類のアプリケーションのトラフィックが流れる環境での評価
- 評価に用いたトラフィックデータ
 - 以下のアプリケーション利用時のトラフィックキャプチャデータ

Download	クラウドストレージからファイルをダウンロード (Dropbox, Bitcasa, Boxを利用)
Upload	クラウドストレージへファイルをアップロード (Dropbox, Bitcasa, Boxを利用)
Video Live Streaming	動画をリアルタイムで配信 (Ustreamの生中継)
Video on Demand	動画ファイルをダウンロードしながら再生 (YouTube, ニコニコ動画, Ustreamの録画配信)
Interactive	ユーザーの操作に応じた返答 (Google Map, Yahoo!地図)

2014/2/21

特別研究報告発表会

7

評価指標

クラスタリングの正確度

- 各アプリケーションのトラフィックについて適切なクラスタに分類された割合
 - 本評価では以下の条件を満たす分類を「適切である」と定義
 - 分類先クラスタ属性の正当性: 当該クラスタに分類されたトラフィックのうちクラスタ属性と同じアプリケーションのものが最も多い
 - 当該クラスタサイズの正当性: 同一アプリケーションのトラフィックがクラスタ内に 3 個以上含まれる

計算時間

- クラスタリングのプログラムの実行時間
 - トラフィックデータ数を 50 から 1000 まで変化させて計測

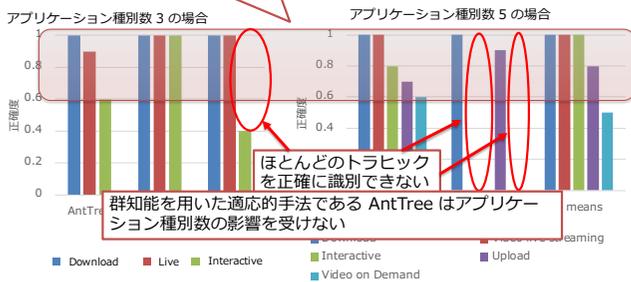
2014/2/21

特別研究報告発表会

8

正確性の評価結果

- AntTree はアプリケーションの種別数によらず全アプリケーションのトラフィックを比較的高い正確度で分類可能
- K-means では実際のアプリケーション種別数に応じて適切な K の設定が必要
 - ほとんどのトラフィックを正確に識別できない場合が存在



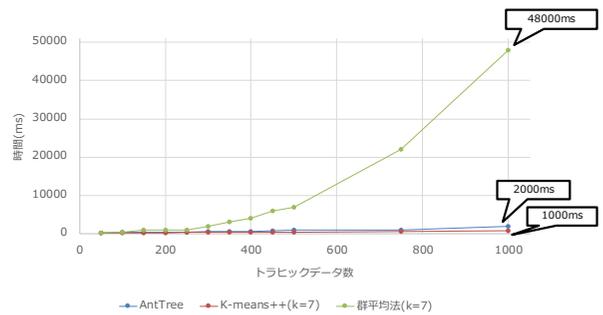
2014/2/21

特別研究報告発表会

9

計算時間の評価結果

- AntTreeは、K meansと同様、短い計算時間で分類可能



2014/2/21

特別研究報告発表会

10

まとめと今後の課題

- まとめ
 - WEB トラフィックの分類に、クラスタリング手法を適用
 - 分類に用いる統計情報の検討
 - インタラクティブ性と定常性
 - 分類に適した手法の議論
 - 群知能を用いた手は比較的高い正確度で種別によらずアプリケーションを分類可能
 - 1000 フローを 2 秒で分類
- 今後の課題
 - アプリケーション分類の正確性の向上
 - ネットワークを流れるトラフィックをリアルタイムに分類し制御を行う手法への適用

2014/2/21

特別研究報告発表会

11