小口研究室　研究紹介　(2017年度)
(お茶の水女子大学理学部情報科学科)

大規模災害時におけるトラヒック制御システムの実アプリケーションによる評価
（研究担当：平久　紬）

研究背景

❖様々な種類の膨大なトラヒックが入り交じっている
❖全てのアプリケーションのトラヒックが同様に扱われている

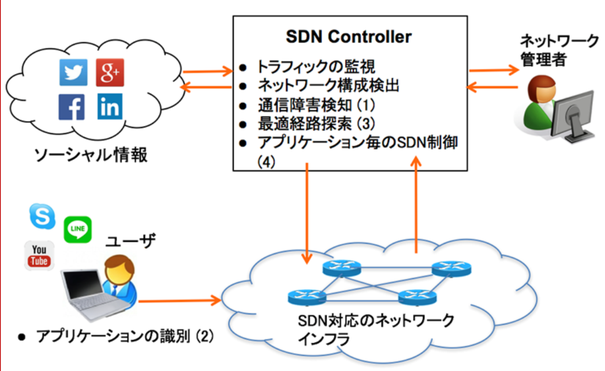
大規模災害時、ユーザが本当に必要とする
緊急情報に
なかなかアクセスできない

研究目的

大規模災害時におけるトラヒック制御システムの必要性
❖実社会の急激な状況変化を検出
❖トラフィック種別を判定し、アプリケーションごとにトラヒック制御

ソーシャル情報に基づく、
アプリケーション毎のトラヒック制御システム
の実装と評価

大規模災害時におけるトラヒック制御システム



(1) 通信障害検知
リアルタイムにツイートを監視し、通信障害がどこで起こっているかを検知する

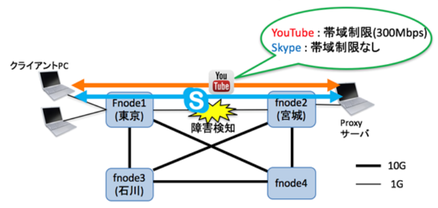
(2) アプリケーションの識別
流れるトラヒックからアプリケーションを識別する
アプリ識別子としてトレーラバイトを付加する

(3) 最適経路探索
経路切り替えを行う場合のみ行う
リンクの初期値: 1
障害ツイートの中にスイッチと対応させた地名を含むツイートが閾値以上あったら+1
ダイクストラ法を用い、トラヒックの経路を決定

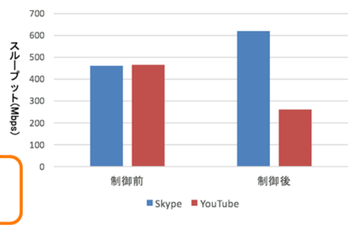
(4) アプリケーション毎のSDN制御
トレーラバイトを確認し、アプリケーションを識別しTable IDを代わりに付加
OpenFlowのREST-APIを用い、災害時にユーザが必要とするアプリケーションを優先するような制御を行う

アプリケーション毎の制御実験

❖ 東日本大震災時のTweetを使用
❖ YouTubeとSkypeを模擬したトラヒック流す
❖ 本システムを動作し、障害検知後に帯域制御



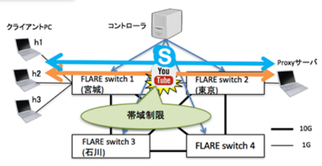
実験結果
- iperfで帯域制御前と後の各スループットを測定
- YouTubeのトラヒックのみ帯域が抑えられた
- Skypeのトラヒックの利用可能帯域が増加した



優先度が高いアプリを優先する制御が行われた

実アプリケーションによる実験

❖ 東日本大震災時のTweetを使用
❖ Skypeによるテレビ電話を行うと同時にYouTubeを流す
❖ 本システムを動作させ、障害検知後にYouTubeのトラヒックのみ帯域制限



実験結果
- 障害検知後、提案システムによりSkypeのパケットロス率が下がり、テレビ電話が行いやすくなった

	帯域制御前 Skypeの音声と画像の乱れ	帯域制御後 Skypeの音声と画像の乱れ
1Mbps	○	○
500Kbps	○	○
400Kbps	×	△
300Kbps	×	△or×

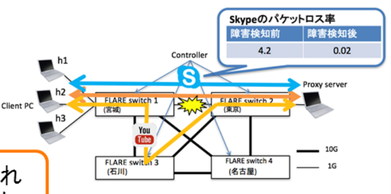
○ テレビ電話可能
△ 音声や画像の乱れはあるがテレビ電話可能
× テレビ電話不可能

	帯域制御前の Skypeのパケットロス率	帯域制御後の Skypeのパケットロス率
1Mbps	0	0
500Kbps	0.01	0
400Kbps	8.1	4.1
300Kbps	8.5	8.3

優先度が高いアプリを優先する制御が行われたことが実アプリを用いた実験により示された

❖ 東日本大震災時のTweetを使用
❖ 本システムを動作させ、障害検知後にYouTubeのトラヒックのみ経路切り替え

実験結果
- 障害検知後、提案システムによりSkypeのパケットロス率が下がり、テレビ電話が行いやすくなった



ビッグデータ分散処理基盤におけるパラメータ制御の一検討
（研究担当:加藤 香澄）

研究背景

ライフログの取得・蓄積と利用の拡大
- カメラやセンサの発達
- クラウドコンピューティングの普及

一般家庭にサーバやストレージを設置し、解析処理を行うのは困難

動画像データが集約し、解析処理によってクラウドに大きな負荷

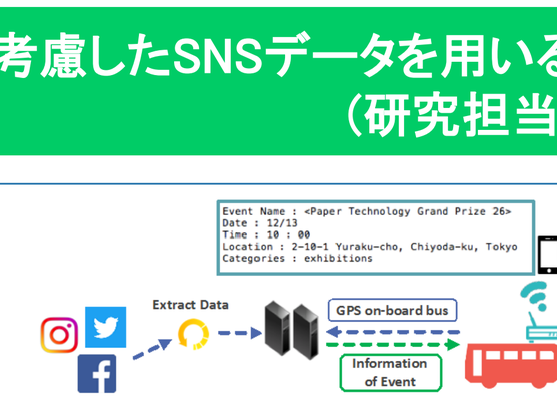
Apache Sparkを用いて
Chainerによる解析処理を並列高速化
- パラメータ制御により処理を効率化

Spark
- 分散コンピューティングプラットフォーム
- 高速かつ汎用的

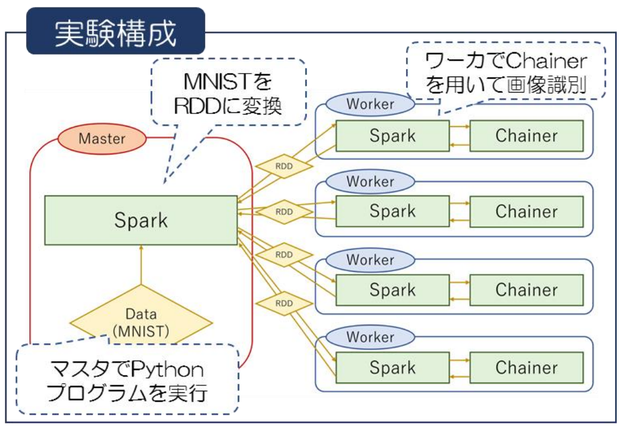
Chainer
- ディープラーニングフレームワーク
- 柔軟性、直感的、高機能

研究概要

各家庭から収集した動画像データをSparkのストリーミング機能を用いて複数のワーカに分散させ、動画像データの機械学習処理を並列化



実験構成



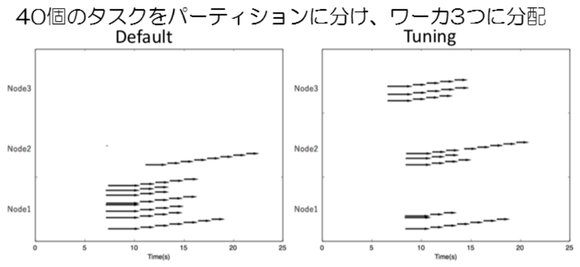
実験結果

Sparkによる並列機械学習処理部分の実験を実施

調整するパラメータ
1. Locality wait time
   タスクの割り当て前の待機時間
   デフォルトでは3secところを0secに変更
2. パーティションの作成方法
   以下3つの手法について比較
   ① メソッドpartitionBy()を利用
   ② メソッドrepartition()を利用
   ③ メソッドpartitionBy()においてパーティショナをラウンドロビン的に変更

1. Locality wait time

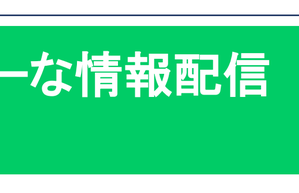
40個のタスクをパーティションに分け、ワーカ3つに分配



- 全体の処理時間はほぼ変化なし
- 調整によりパーティションがほぼ均等にワーカに分配

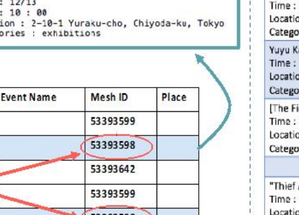
2. パーティションの作成方法

①メソッドpartitionBy()を利用



- ノード数増加により処理時間が1/3に
- パーティション数32ほどで計測結果が横ばいに

②メソッドrepartiton()を利用



- パーティション数の変化による処理時間のばらつきが大きい
- ノード増加により処理時間が2/5になっているところも

③メソッドpartitionBy()においてパーティショナを調整

- ノード数3以上で処理効率が向上しない
- 計測結果はパーティション数24で明確に横ばい

まとめと今後の課題

まとめ
- Sparkを用いてChainerの機械学習処理を並列化
- パラメータ調整による挙動の変化を確認

今後の課題
- ログ解析による分散環境のさらなる調査
- パラメータ制御による処理の効率化

場所と時間を考慮したSNSデータを用いる訪日外国人観光客へのタイムリーな情報配信
（研究担当：工藤　瑠璃子）

研究背景

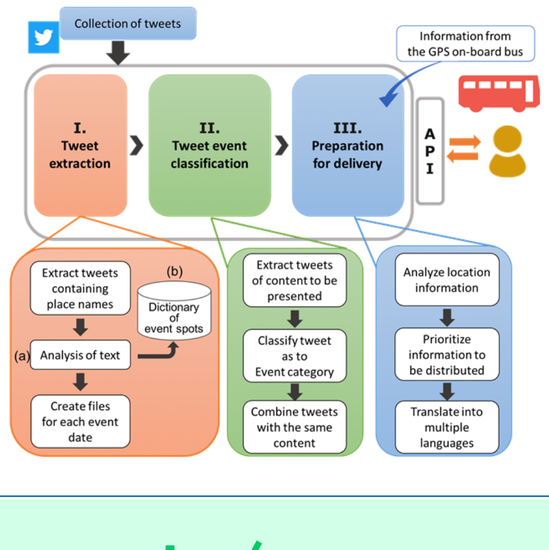
◆ 2020年の東京オリンピック開催を受けインバウンドが増加しているが、観光情報の発信は不足
◆ 恒常的に人気のあるスポットはガイドブックなどから取得可能だが、今まさに開催されているイベントは情報を取得しにくい
◆ 地理的・時間的な制約がある旅行者などが必要とする「その時」「その場」で役立つ情報配信は少ない

時間とともに移動していく旅行者などにリアルタイムに有益な情報を配信するシステム構築
★イベントや観光地の位置情報と時間情報

提案システム

ツイートの抽出
- Twitter APIのキーワード検索によりツイートを収集
- ◇地名をキーワードとする：23区・山手線の駅 etc.
- テキスト解析
- ◇正規表現→日付と時間 / URL / 住所
- ◇イベントスポット辞書→アミューズメント施設・美術館 etc.
- 日付ごとにファイル出力

ツイートのイベント分類
- 有用でないツイートを排除
- 観光客の嗜好に合わせて配信することを見据えてカテゴリ分類
- 3手法で精度を比較
  1. SVM + IPADIC
  2. SVM + ipadic-NEologd
  3. Random Forest + ipadic-NEologd



移動方向を考慮したスケジューリング

バスのルートに沿って、ユーザの移動方向先の情報のみを配信

D.六本木ヒルズ周辺のイベント情報

- Within 1 km -
Time : 10 : 00
Location : Mori Museum
Category : Exhibition

Yuyu Kawakami : Solo Event 2017 - Mysterious Grand Club TYPE - A -
Time : 11 : 30
Location : Roppongi
Categories : music events

[The First stage of the burglar actors]
Time : 11 : 55
Location : Roppongi
Categories : movies

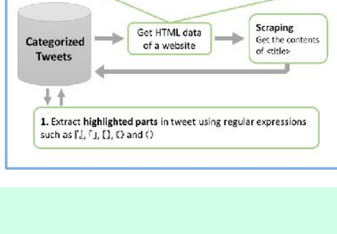
"Thief Actors"
Time : 15 : 20
Location : Roppongi Hills
Categories : movies

"I want to become an actor,"
Time : 17 : 15
Location : Roppongi
Categories : movies

- Within 2 km (north) -
-[Mone Kamishiraishi LIVETOUR 2017 "and".
"Time : 17 : 00
Location : Akasaka BLITZ
Categories : music events

土地勘のない旅行者に有効
「その場」で役立つ情報提供

外部情報を利用した情報補完



対象件数	Correct	Collection rate
979	784	0.80

手法1.のみでは0.48
→ 0.32 up

Method	対象件数	Correct	Collection rate	
1. パターンマッチ	604	473	0.78	ツイート
2-a. URLありスクレイピング	77	51	0.66	外部情報
2-b. Web検索スクレイピング	397	260	0.65	外部情報

外部の情報をとって補強することで、有用な情報が増える

今後の展望

- イベント分類、テキスト情報整理の精度向上
- システムの有用性の評価

http://ogl.is.ocha.ac.jp/

oguchi@is.ocha.ac.jp