

分散ストリーム処理フレームワークを用いた 動作識別処理性能の調査

高崎 智香子[†]竹房 あつ子[‡]中田 秀基[§]小口 正人[†][†]お茶の水女子大学[‡]国立情報学研究所[§]産業技術総合研究所

1. はじめに

子供やお年寄りの見守りサービスや防犯を目的として家庭のセンサで取得した動画像をリアルタイムに機械学習で解析するには、データ量と解析計算量が課題となる。我々は、センサ側で姿勢推定ライブラリ OpenPose を用いて動画像データから抽出した特徴量のみを使用し、クラウド側で機械学習によって動画に含まれる動作を識別することで、プライバシーや処理遅延の問題に対処する分散処理手法を提案している [1]。このようなシステムでは、複数家庭のセンサから大量のデータがクラウドに継続して送られることが想定されるため、急激なデータの増加によるシステム負荷上昇への対応も必要である。

本研究では、クラウドにおいて、分散メッセージングシステム Apache Kafka(以降、Kafka と呼ぶ)と分散ストリーム処理フレームワーク Apache Flink(以降、Flink と呼ぶ)を用いて機械学習処理を行うシステムを構築し、分散ストリーム機械学習処理の性能特性を調査した。

2. センサ、クラウド間分散システムの概要

本研究では、図1のようなセンサ、クラウド間分散動画像解析システムを想定している。各一般家庭に設置されたセンサのカメラで動画像を取得し、センサ端末内で前処理を行った後、メッセージングシステムを用いてクラウドにデータを収集して分散ストリーム処理基盤上で分散機械学習を行う。クラウド側では動画像や静止画を用いず、センサ側で抽出したキーポイントデータのみを使用して解析を行う。センサ端末で OpenPose を用いてキーポイントの座標データを抽出し、Kafka Producer から Kafka Broker に転送する。クラウドでは Kafka Consumer を用いて Kafka Broker からデータを受け取り、Flink の分散ストリーム処理機能を用いて機械学習の推論を行うことで動画像に含まれる動作を識別する。その後、Kafka Producer を用いて Kafka Broker に解析結果を転送する。

3. 実験

Flink において (1) ニューラルネットワーク (NN) の推論を行う場合、(2)LSTM の推論を行う場合のスループ

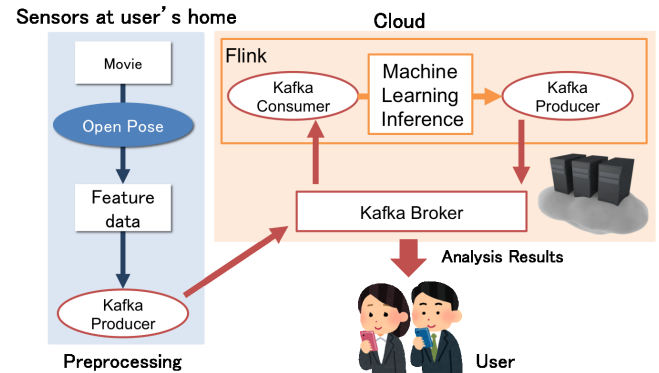


図1: 提案する動画像解析システム

トを測定し、スケーラビリティを調査した。

3.1 機械学習手法

(1) NN モデル、(2) LSTM モデルの2手法で動作識別モデルを予め TensorFlow を用いて作成して学習し、Flink プログラムで読み込んで推論で用いる。(1) NN は人の神経細胞を模したモデルであり、完全結合の NN(MLP) を用いた。(2) LSTM は、RNN の長期記憶ができないという欠点を解消し、データの長期依存を学習可能にした手法である。

3.2 使用データ

OpenPose を用いて動画像の各フレームから抽出したキーポイントの座標データを使用し、機械学習による動作識別モデルを作成する。データセットには、日常の動作100カテゴリの動画像を約1000ずつ集めた STAIR Actions[2]の動画像を利用する。STAIR Actions の各動画像から、0.3秒ごとの等間隔に10枚の静止画を取得した。その後、各静止画に対して OpenPose を用いて25のキーポイントの画像上の x, y 座標を取得して特徴量50のデータを取得し、データセットを作成した。各静止画の50のキーポイントを時系列順に並べて、合計500の特徴量を1入力データとして使用した。データ数は96807で、このうちの7割を学習に、3割をバリデーションに使用して動作識別モデルを作成した。

3.3 実験環境

同質の5つのノードを用いて実験を行う。Master ノードで Kafka Producer と Kafka Broker を動作させデータ転送を行う。加えて Flink の JobManager を動作させ、並列度に応じて4台の Worker ノードで動作している Flink TaskManager のスロットにタスクを分配する。各スロットでは、Kafka Consumer を動作させて Kafka Broker からデータを受け取り、キーポイントデータを用いた推論を

A Study on Performance of Action Recognition Processing using a Distributed Stream Processing Framework

Chikako Takasaki[†]

Atsuko Takefusa[‡]

Nakada Hidemoto[§]

Masato Oguchi[†]

[†]Ochanomizu University

[‡]National Institute of Informatics

[§]National Institute of Advanced Industrial Science and Technology (AIST)

表 1: 実験で使用した計算ノードの性能とソフトウェアバージョン

CPU	Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz 14Cores 28Threads × 2
OS	Alpine Linux 3.10.1
Memory	125Gbyte
Kafka Version	2.2.0
Flink Version	1.7.2
ZooKeeper Version (Master のみ)	3.4.14

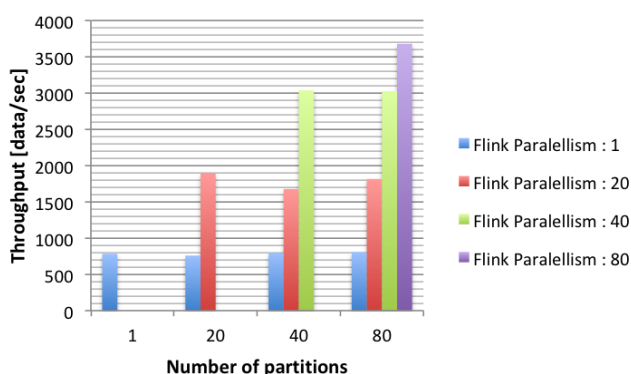


図 2: (1)Kafka Broker から Kafka Consumer がデータを受信し, NN の推論を行う場合のスループット

行って動作識別を行う。その後, Kafka Producer を各スロットで動作させ, 動作識別結果を Kafka Broker に転送する。実験に用いた計算機の性能とソフトウェアバージョンを表 1 に示す。Kafka の Producer および Consumer プログラムは Java で実装した。

3.4 実験結果

図 2, 図 3 に NN および LSTM の推論を行う場合のスループットを示す。また, 1 データの NN と LSTM の推論時間は, NN は平均 0.508ms, LSTM は平均 1.835ms であった。図 2 から, (1)NN の推論をおこなう場合は, Flink の並列度に比例してスループットが向上しており, Flink の並列処理機能を用いることで高効率にデータを処理できていることがわかる。図 3 から, (2)LSTM の推論を行う場合は, Flink の並列度を 20, 40 に増やすことでスループットが向上していることが分かる。一方, 並列度を 80 にしてもスループットがあまり向上していない。これは, Kafka Broker での処理がボトルネックになっていると考えられる。

4. 関連研究

Hara ら [3] は, 動画から行動を識別するため, 3D Residual Network(ResNet)[4] による性能改善を示した。しかし, 動作識別処理は計算量が膨大であるため, 一般家庭において深層学習を使用した解析を行うことは難しい。

本研究はエッジとクラウドで処理を分散させる事によって深層学習を用いた解析をリアルタイムに行う。また, エ

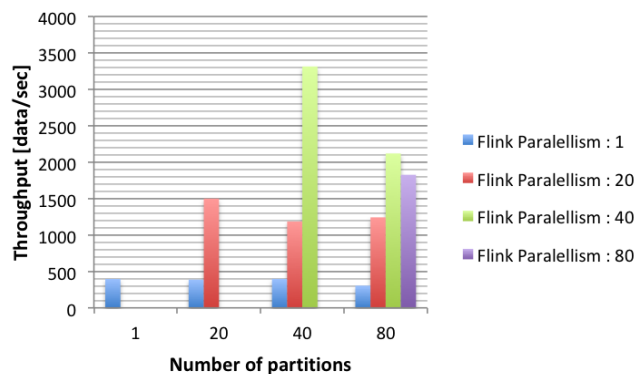


図 3: (2)Kafka Broker から Kafka Consumer がデータを受信し, LSTM の推論を行う場合のスループット

ジでの前処理により抽出した動画像に含まれる人間のキーポイントの座標値のみをクラウドでの解析に使用することで, 生の動画像データをクラウドに送信する通信料やプライバシーの問題に対処可能である。

5. まとめと今後の予定

本稿では, STAIR Actions データセットの動画像から取得した画像から OpenPose を用いて抽出したキーポイントデータを Kafka で収集し, クラウドで Flink を用いて機械学習による動作識別処理を行うシステムを構築し, 解析処理性能を調査した。実験から, Flink の並列処理機能を用いることで機械学習処理のスループットがスケールすることが確認できた。

今後は, 家庭に配備可能なエッジデバイスとクラウド環境でのスループットや遅延時間について調査し, リアルタイム解析の実現を目指す。

謝辞

この成果の一部は, JSPS 科研費 JP19H04089 及び, 2020 年度国立情報学研究所公募型共同研究 (20S0501) の助成を受けたものです。

参考文献

- [1] C. Takasaki, A. Takefusa, H. Nakada, and M. Oguchi. A study of action recognition using pose data toward distributed processing over edge and cloud. In *Proc. the 11th IEEE International Conference on Cloud Computing Technology and Science (CloudCom2019)*, pp. 111–118, 2019.
- [2] Yuya Yoshikawa, Jiaqing Lin, and Akikazu Takeuchi. Stair actions: A video dataset of everyday home actions. *arXiv preprint arXiv:1804.04326*, 2018.
- [3] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In *Proc. the IEEE conference on Computer Vision and Pattern Recognition*, pp. 6546–6555, 2018.
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.