深い最近傍の異常検出

Liron Bergman [*1] **Niv Cohen** [*1] **Yedid Hoshen** [1]

Abstract

ニアレストネイバーは、異常検出のための成功した長年の手法である。最近、自己教師付きディープメソッド（例えばRotNet）によって大きな進歩が達成された。しかし、自己教師付き特徴量は、一般的にImagenetで事前に訓練された特徴量を下回る。本研究では、最近の進歩が、Imagenetで事前に訓練された特徴空間上で動作する最近傍手法を本当に上回ることができるかどうかを調査する。単純な最近傍ベースのアプローチは、画像分布に関する仮定が少ない一方で、精度、数ショット汎化、学習時間、ノイズ頑健性において、自己教師あり手法を上回ることが実験的に示されている。

1. Introduction

世界と相互作用するエージェントは常に連続的なデータの流れにさらされている。エージェントは特定のデータを異常、つまり特に興味深い、あるいは予期せぬものとして分類することで利益を得ることができる。このような識別は、必要な観測にリソースを割り当てるのに役立つ。このメカニズムは、人間がチャンスを発見したり、危険を警告するために使用される。人工知能による異常検知は、詐欺検知、サイバー侵入検知、重要な産業機器の予知保全など、多くの重要な応用がある。

機械学習では、異常検出のタスクは、データポイントを正常または異常としてラベル付けできる分類器を学習することから成る。教師あり分類では、異常なデータはノイズとみなされるのに対して、正常なデータに対してはうまく機能しようとする。異常検知手法の目標は、変動が激しく予測が困難な極端なケースを特別に検知することである。このため、異常検知のタスクは困難である（そしてしばしば仕様が不十分である）。

異常検知の3つの主な設定は、教師あり、半教師あり、教師なしである。

ヘブライ大学コンピューターサイエンス・工学部、イスラエル。宛先 Yedid Hoshen <yedid.hoshen@mail.huji.ac.il>.

教師あり設定では、正常データと異常データに対してラベル付けされた訓練例が存在する。従って、他の分類タスクと基本的な違いはない。この設定はまた、多くの異常検出タスクにとっては制約が多すぎる。例えば、新しい病気の出現のように、興味のある異常の多くは過去に見たことがないからである。より興味深い半教師あり設定では、全ての学習画像は正常であり、異常は含まれない。正常-異常分類器を学習するタスクは1クラス分類となる。最も困難な設定は教師なし設定で、正常データと異常データの両方からなるラベル付けされていない訓練セットが存在する。典型的な仮定は、異常データの割合が正常データよりも有意に小さいことである。本稿では、半教師あり設定と教師なし設定の両方を扱う。典型的な異常検知手法は、距離、分布、または分類に基づく。ディープニューラルネットワークの出現は、それぞれのカテゴリーに大きな改善をもたらした。この2年間で、ディープな分類に基づく手法は、他の全ての手法を大幅に凌駕した。これは主に、正常なデータに対してあるタスクを実行するように訓練された分類器は、未見の正常なデータに対してはこのタスクをうまく実行するが、異常なデータに対しては、異なるデータ分布に対する汎化がうまくいかないために失敗するという原理に依存している。

最近の論文で，Gu ら（2019）は，生データ上の K 最近傍（kNN）アプローチが，表データ上の最先端手法と競合することを実証した．驚くべきことに、kNNは現在のほとんどの画像異常検出の論文では使われていないし、比較もされていない。本論文では、生の画像データに対するkNNの性能は良くないが、強力な市販の汎用特徴抽出器と組み合わせると、最新技術を凌駕することを示す。具体的には、Imagenetで事前学習されたResNet特徴抽出器を用いて、全ての（訓練とテストの）画像を埋め込む。各テスト画像の埋め込みとトレーニングセットの間のK最近傍（KNN）距離を計算し、単純な閾値ベースの基準を使用して、データが異常かどうかを判断します。

我々は、一般的に使用されるデータセットと、Imagenetとは全く異なるデータセットの両方で、このベースラインを広範囲に評価した。我々は、このベースラインが既存の手法と比較して大きな利点があることを発見した。

Deep Nearest Neighbor Anomaly Detection

Liron Bergman [*1] **Niv Cohen** [*1] **Yedid Hoshen** [1]

Abstract

Nearest neighbors is a successful and long-standing technique for anomaly detection. Significant progress has been recently achieved by self-supervised deep methods (e.g. RotNet). Self-supervised features however typically underperform Imagenet pre-trained features. In this work, we investigate whether the recent progress can indeed outperform nearest-neighbor methods operating on an Imagenet pretrained feature space. The simple nearest-neighbor based-approach is experimentally shown to outperform self-supervised methods in: accuracy, few shot generalization, training time and noise robustness while making fewer assumptions on image distributions.

1. Introduction

Agents interacting with the world are constantly exposed to a continuous stream of data. Agents can benefit from classifying particular data as anomalous i.e. particularly interesting or unexpected. Such discrimination is helpful in allocating resources to the observations that require it. This mechanism is used by humans to discover opportunities or alert of dangers. Anomaly detection by artificial intelligence has many important applications such as fraud detection, cyber intrusion detection and predictive maintenance of critical industrial equipment.

In machine learning, the task of anomaly detection consists of learning a classifier that can label a data point as normal or anomalous. In supervised classification, methods attempt to perform well on normal data whereas anomalous data is considered noise. The goal of an anomaly detection methods is to specifically detect extreme cases, which are highly variable and hard to predict. This makes the task of anomaly detection challenging (and often poorly specified).

The three main settings for anomaly detection are: supervised, semi-supervised and unsupervised. In the *supervised* setting, labelled training examples exist for normal and anomalous data. It is therefore not fundamentally different from other classification tasks. This setting is also too restrictive for many anomaly detection tasks as many anomalies of interest have never been seen before e.g. the emergence of new diseases. In the more interesting *semi-supervised* setting, all training images are normal with no included anomalies. The task of learning a normal-anomaly classifier is now one-class classification. The most difficult setting is *unsupervised* where an unlabelled training set of both normal and anomalous data exists. The typical assumption is that the proportion of anomalous data is significantly smaller than normal data. In this paper, we deal both with the semi-supervised and the unsupervised settings. Anomaly detection methods are typically based on distance, distribution or classification. The emergence of deep neural networks has brought significant improvements to each category. In the last two years, deep classification-based methods have significantly outperformed all other methods, mainly relying on the principle that classifiers that were trained to perform a certain task on normal data will perform this task well on unseen normal data, but will fail on anomalous data, due to poor generalization on a different data distribution.

In a recent paper, Gu et al. (2019) demonstrated that a K nearest-neighbours (kNN) approach on the raw data is competitive with the state-of-the-art methods on tabular data. Surprisingly, kNN is not used or compared against in most current image anomaly detection papers. In this paper, we show that although kNN on raw image data does not perform well, it outperforms the state of the art when combined with a strong off-the-shelf generic feature extractor. Specifically, we embed every (train and test) image using an Imagenet-pretrained ResNet feature extractor. We compute the K nearest neighbor (KNN) distance between the embedding of each test image and the training set, and use a simple threshold-based criterion to determine if a datum is anomalous.

We evaluate this baseline extensively, both on commonly used datasets as well as datasets that are quite different from Imagenet. We find that it has significant advantages over existing methods: i) higher than state-of-the-art accuracy ii) extremely low sample complexity iii) it can utilize

[*]Equal contribution [1]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel. Correspondence to: Yedid Hoshen <yedid.hoshen@mail.huji.ac.il>.