

AIの社会実装と最近のAI技術の
トレンド
事前課題

2021年1月13日

株式会社スキママッチング

岩渕 真人

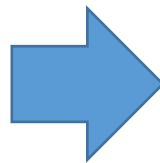
masato.iwabuchi@gmail.com

事前課題-1

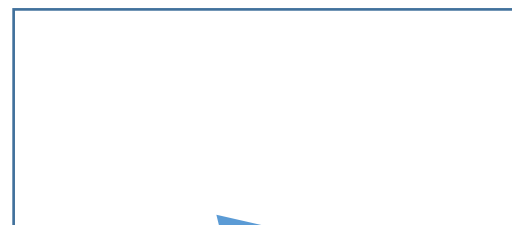
● 背景

- CNN(畳み込みニューラルネットワーク：詳細は講義で解説、この課題は中身を知らなくても回答できる) は2012の年 ILSVRC2012 以降急激に発展し、図形認識に対して、人以上の検出精度を出せるようになった。
- CNNの学習（モデル作成）では、一般に教師あり学習が用いられる

教師データ



学習（モデル作成）



未知のデータをモデルに入力



推論

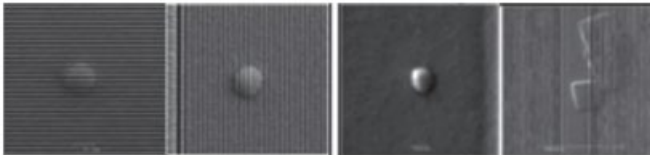
猫である確率 90%
犬である確率 10%

事前課題-1

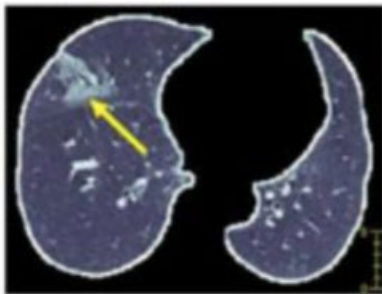
- 目的

- AIの社会実装の例として、CNNを製品の欠陥検査や医療の画像診断に応用したい

半導体製品の欠陥例



医療画像診断



事前課題-1

● 課題

- ところで、図形認識してラベルを付ける場合、一般の人の誤り率は4-5%程度、プロの検査員やドクターでも0.2-1%程度の誤りがあるというデータがある
- 従って、CNNの学習に使う教師データには1%の誤りが混在している、ということを想定するのが現実的
- このような教師データで学習したCNNの誤り率を1%以下にするのは困難
- 一方、外観検査で誤り率が1%というのは、100個検査すると1個不良を見逃す、ということの意味する。医療診断では100症例で1症例、誤診することになる。これは社会実装としてモデルをリリースし、使ってもらう場合には、受け入れられない水準である。
- 目標としては、Recall \geq 0.9999（つまり、誤り率：不良の見逃し率は0.01%以下）、Precision \geq 0.9（つまり、正常製品を不良としてしまう率が10%以下）ぐらいにしないといけない
- 教師データには1%の誤りが混在するものとして、学習モデルをリリースし、実用化するにはどうしたらよいか？

事前課題-2

- ある店で商品の毎日の売り上げを予測するモデルを作ろうとしている。
- 過去3年分（2017、2018、2019年）の日々の売り上げデータ、要因を及ぼすと思われるデータ（人の来場数、天候など）があり、モデル作成にはこのデータを利用できる。
- ところが、4年に一度のイベント（オリンピック）や、過去なかった新型コロナの影響は考慮されていない。
- 来年の売り上げ予測モデルを作り、リリースするにはどうしたらよいか？