

放射線技師養成における機械学習の動作原理を学ぶ学習活動

Learning Activity to Understand Action Principle of Machine Learning in Radiologic Tech Training Program

原田 義富^{*1}, 後藤 善友^{*2}Yoshitomi HARADA^{*1}, Yoshitomo GOTO^{*2}^{*1}日本文理大学医療専門学校^{*1}Medical Collage, University of Nippon Bunri^{*2}別府大学短期大学部^{*2}Beppu University Junior College

Email: haradayt@nbu.ac.jp

あらまし: X線画像から機械学習で何らかの予測を行うことなどは応用上も重要な研究として注目されている。そのため、教育的な視点から、医療関係者が機械学習の基本原則や限界などを体験的に理解できれば非常に意義深い。そこで本研究では、診療放射線技師を目指す学生 18 名が、インターネット上に存在する CT 画像から学習データとテストデータを作成し、学習活動を通じて、CT 画像の部位を自動識別する機械学習の動作原理の理解を試みた。その結果、学生の機械学習への関心が高まった。特徴量や学習データの追加が今後の課題である。

キーワード: 機械学習, 学習活動, サポートベクタマシン, 再学習

1. はじめに

昨今の医療分野では、医用画像を用いたコンピュータ支援診断 (Computer-aided detection(or diagnosis)) など、機械学習への期待が高まっている¹⁾。中でも、放射線業務に携わる診療放射線技師は、医師の負担軽減のための読影補助など、今後機械学習に触れる機会が増える予想される²⁾。そのため、教育的な視点から、医療関係者が機械学習の基本原則や限界などを体験的に理解できれば非常に意義深い。

そこで本論文では、診療放射線技師を目指す学生を対象に、インターネット上に存在する CT (Axial) 画像から学習データとテストデータを作成し、CT 画像の撮影部位 (例えば、頭部や上腹部) を自動識別する機械学習の動作原理を、学習活動にて理解する試みについて述べる。

2. 対象

インターネット上に存在する CT 画像 (頭部・胸部・上腹部) をそれぞれ任意に選択した 12 枚を用いる。しかし、これらの画像は、画像サイズ、解像度、コントラスト、病変等も見られ規格が定まらないものを用いた。そして、それぞれ 10 枚を学習データ、残り 2 枚をテスト画像とし、機械による撮影部位の識別を評価した。

また、機械学習には scikit-learn によるサポートベクタマシン (以下, SVM) を用いた。

3. 方法

本手法は、頭部 CT 画像と上腹部 CT 画像のそれぞれ 12 枚を用いて、以下に述べる特徴量にて、事前に作成したそれぞれ 9 枚の学習データと、残り 1 枚

の学習画像と 2 枚のテスト画像から、学生グループがフリーの画像処理ソフト (ImageJ) を用いて学習データとテストデータを作成する。学習データを SVM に取り込み機械学習させることで、頭部・上腹部のテスト画像がうまく識別できることを確認する。

次に、上記と同様の手順で、胸部 CT 画像も含めて再学習させ、同じ特徴量を用いて頭部・胸部・上腹部 CT 画像を識別する。その結果、識別率が下がる場合があったため、機械学習を行うために必要な条件など機械学習の課題を検討する。

ここで本手法における SVM のデータ学習時間は、2.2GHz の PC で約 2 秒程度であった。

3.1 特徴量の抽出

機械が頭部 CT と上腹部 CT 画像を識別するためには、両者を識別する特徴量を数値データとして機械に学習させる (教える) 必要がある。両者を識別する特徴量としてどのようなものが良いかを学生に考えさせた上で、頭部 CT と上腹部 CT の 2 クラスを識別する特徴量の例として、頭部と腹部陰影のそれぞれ縦幅と横幅の比 (アスペクト比)、左右対称性の 2 つを二次元の特徴量として採用し、機械学習のための学習データを取らせた (図 1)。具体的には、頭部と上腹部 CT 画像をそれぞれ 1, -1 にラベリングして機械学習する。ここで 2 つの特徴量を用いた理由は、1 つの特徴量では識別が必ずしも正しくできない可能性があるためである。

3.2 サポートベクタマシン (SVM) の利用

SVM の特長に、少ないサンプルによる学習でも高い識別能力が期待できる点が挙げられる。さらに 3.1 で述べた二次元の特徴量であれば、1 つの特徴量よりも情報量が多い。そこで、機械学習の方法として頭部・上腹部の 2 つのクラスを識別するために SVM

を用いた。非常に簡単な特徴量であっても、SVMの各種パラメータやカーネルを変化させることで領域を大きく二分することができる。SVMによる学習の結果、頭部と上腹部CT画像のテスト画像を両方とも正しく識別できた。

3.3 識別率への影響

一方、インターネット上の胸部CT画像10枚を“ラベル0”として、頭部・胸部・上腹部CT画像30枚全てのデータを再学習する。次に、それぞれのテスト画像の撮影部位を識別実験した結果、胸部CT画像は2枚とも正しく識別できたが、カーネルを変えても頭部や腹部CTテスト画像が誤識別となるものがあった。学習データセットが変更された際の、再学習手続が不適切だと、識別率が下がるなどの悪影響があることを認識する。

4. 調査

1) 学生の機械学習認知度の事前調査

調査の結果、「知っている」、「聞いたことはある」、「知らない」、「興味なし」は、それぞれ6%、22%、72%、0%であった。

2) 学生の理解を計るために問題を出題

機械学習に関する概念や仕組みについて出題した記述式問題2問の正解率は80.5%であった。

3) 識別率の向上と機械学習への課題

学生から「判別するものが増えたため、特徴量の変更・増加や学習データ数の増加、アルゴリズムの変更を試みては」との意見が得られた。

5. 考察

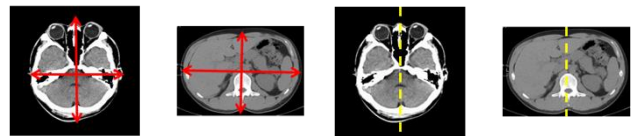
AI（人工知能）という言葉の認知度は高いが、機械学習については72%の学生が知らなかった。

本研究で用いた特徴量は、頭部CT画像では頭部陰影の縦横比（アスペクト比）が大きく左右対称性も強い。一方、上腹部CT画像では縦横比が小さく左右対称性も弱いという反対の傾向を示す。このため、インターネット上の画像では画像サイズなどの規格が様々であるが、頭部と上腹部CT画像を識別するための特徴量として、適当な特徴量やSVMの選択により、機械がデータをどのように学習し、どのように識別するか、調査4の2)の結果から、学生が体験することで、機械学習の概念やプロセスの理解が見られた。

また、SVMがもつパラメータやカーネルを変えても学習データから境界線が引けること（図2）を学び、識別結果（表1）と一連の流れを示すことで学習前よりも機械学習に対する高い関心が見られた。

一方、縦横比は小さく左右対称性の強い胸部CT画像を再学習させテストした場合、胸部CT画像は識別できても、現段階では頭部や上腹部画像を必ずしも正しく識別できるとは限らない。このことから、識別率向上のためには、特徴量や学習データ数の増加が必要などの意見が学生から上がり、機械学習が

もつ汎化性能や過学習にも触れることができた。



縦横比(アスペクト比)

左右対称性

図1 特徴量 (左:縦横比, 右:左右対称性)

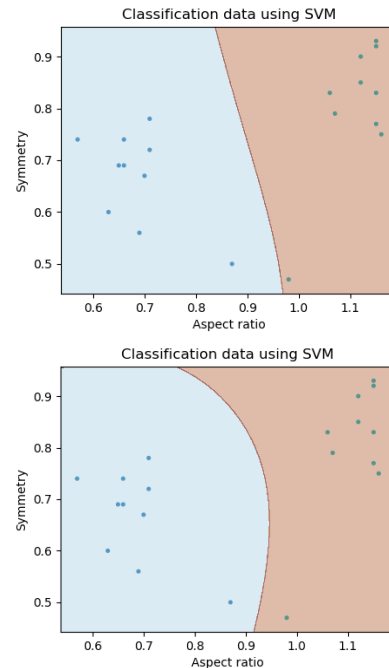


図2 頭部CT画像と上腹部CT画像のSVMによる識別 (上:ソフトマージン, 下:ハードマージン)

表1 頭部・上腹部CT画像のテスト結果

test画像	特徴量		識別結果 *
	縦横比	対称性	
頭部CT	1.28	0.79	1
頭部CT	1.12	0.43	1
上腹部CT	0.63	0.89	-1
上腹部CT	0.71	0.78	-1

* 1: 頭部CT画像, -1: 上腹部CT画像

6. まとめ

機械学習の代表例であるSVMを用いた学習活動により学生は機械学習の内容を理解できた。今後は、SVMの教材化をさらに進め、医療用データセットの整備や、もう少し簡単に学習できるコンテンツ整備を検討する。

参考文献

- (1) 木戸尚治: “ディープラーニング技術を用いたコンピュータ支援診断 (CAD)”, 臨床放射線, Vol.62, No.10, pp.1223-1228 (2017)
- (2) 土井邦雄: “診療放射線技師による読影支援に関する研究の必要性”, 日本放射線技術学会雑誌, Vol.70, No.1, pp.86-88 (2014)