

Medical Image Classification using Support Vector Machine

Your Name¹ and Teacher Author¹

¹University of Fukui, 3-9-1 Bunkyo, Fukui, 910-0019, Japan

2024.11.30

摘要

This is Abstract chapter.

1 引言

支持向量机（SVM）是一种强大的监督式学习算法，广泛应用于分类和回归任务中。该算法通过在特征空间中寻找最大间隔超平面来区分不同类别，有效地提高了模型的泛化能力。

2 相关工作

支持向量机早已被应用于文本分类任务中，如垃圾邮件检测和网页分类[1]。在人脸识别、手写识别[2]和医学图像分析[3]等领域，由于其高效的分类能力，SVM得到了广泛应用。支持向量机的发展与应用是机器学习领域中一个成功的案例，展示了理论研究如何被转化为实际应用的有效工具。

3 方法

本研究的目标是利用支持向量机（SVM）进行分类任务，找到一个能够最大化地分隔不同类别数据点的决策边界，即超平面。该超平面的表达式为：

$$f(x) = \mathbf{W}^T x + b$$

其中， $f(x)$ 是模型的预测输出，表示样本 x_i 落在特定类别的置信度。 \mathbf{W} 是超平面的法向量， b 是偏置项， x 是输入的特征向量。

SVM通过解决一个优化问题来确定最优的 \mathbf{W} 和 b ，该问题旨在最大化两个类别之间的间隔。

通常使用合页损失（Hinge Loss）来训练分类器，该损失函数鼓励找到具有最大边缘的决策边界，定义为：

$$L(y_i, f(x_i)) = \max(0, 1 - y_i f(x_i))$$

其中， $f(x_i) = \mathbf{W}^T x_i + b$ ， y_i 是实际的类标签，取值为 $\{-1, 1\}$ 。本研究构建了一个基本的支持向量机，利用合页损失对其进行优化，实现一个医学图像分类器。

4 实验

本研究利用SVM对医学图像进行分类，判断图像是PET还是CT。该数据集来源于国家癌症研究院的癌症成像计划（CIP），涵盖了355名受试者的肺部扫描图像，共251135张扫描图[4]。这些图像主要收集自2009年至2011年间，每位受试者的数据包括性别、年龄、体重、吸烟史及癌症诊断分类。所有扫描数据以DICOM格式存储。本实验选取了38名被诊断为小细胞癌的受试者的数据，这部分数据包括PET扫描、多种CT扫描以及融合增强后的扫描图像，共12930张图像。通过精确筛选，获得了928张配对扫描图像用于本研究。

5 结论

本研究深入探讨了支持向量机的发展历史、技术原理及其在医学图像分类任务中的应用。实验结果显示，使用本研究中的数据集对SVM进行

表 1: 实验数据划分

参数数量	测试数据集	训练数据集
肺部PET	24	440
肺部CT	24	440
总计	48	880

表 2: 不同输入图像尺寸下SVM决策函数需优化的参数数量

参数数量	128×128	256×256	512×512
通道=1	16385	65537	262145
通道=2	32769	131073	524289
通道=3	49153	196609	786433

训练和测试，模型不仅能有效完成分类任务，而且显示出优异的分类性能。

致谢

本文对国家癌症研究院癌症成像计划表示感谢，感谢其在互联网上公开并授权使用其高质量的医疗图像数据集，为本研究的顺利进行提供了重要资源。

参考文献

- [1] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28, July 1998.
- [2] C. Bahlmann, B. Haasdonk, and H. Burkhardt. Online handwriting recognition with support vector machines - a kernel approach. In *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, pages 49–54, Niagara on the Lake, Ont., Canada, 2002. IEEE Comput. Soc.

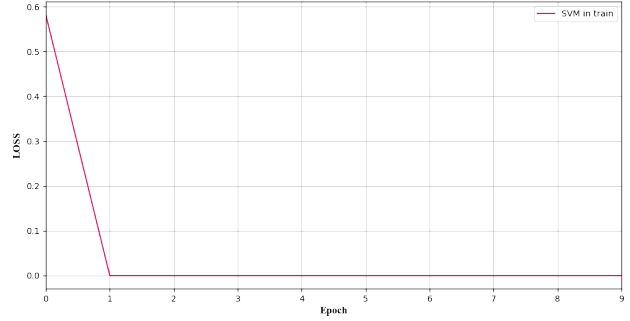


图 1: 训练过程中各周期的损失变化图

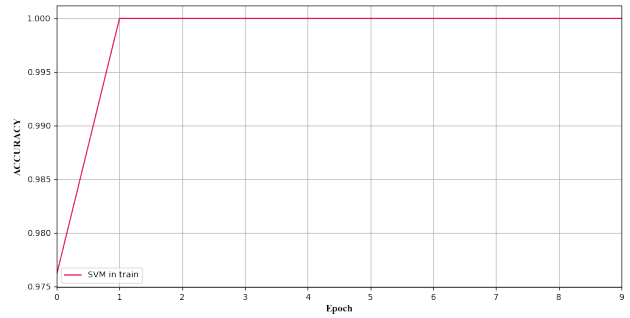


图 2: 训练集上各周期的准确率变化图

- [3] Neha Gautam, Avinash Singh, Kailash Kumar, Puneet Kumar Aggarwal, and Anupam. Investigation on performance analysis of support vector machine for classification of abnormal regions in medical image. *Journal of Ambient Intelligence and Humanized Computing*, February 2021.
- [4] Ping Li, Shuo Wang, Tang Li, Jingfeng Lu, Yunxin HuangFu, and Dongxue Wang. A Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis, 2020.

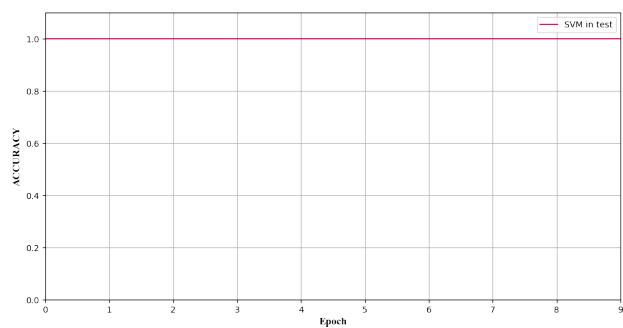


图 3: 测试集上各周期的准确率变化图