

文章编号: 2095-2163(2019)04-0128-04

中图分类号: TP181

文献标志码: A

基于深度 LSTM 的甲亢疾病发展预测及应用系统

杨意豪¹, 王梅¹, 左铭²

(1 东华大学 计算机科学与技术学院, 上海 201620; 2 上海交通大学 医学院附属瑞金医院, 上海 200025)

摘要: 针对甲亢疾病发展的阶段性特点, 建立 LSTM 深度学习模型, 利用患者的前期关键血检指标, 使用该模型对患者未来时刻指标值进行预测, 从而获得后期疾病发展预测情况。在此基础上, 基于 Vue.js 框架构建了发展预测应用原型系统, 实现了指标数据录入、模型调用、预测结果展示等功能, 为医生当前诊疗方案的制定和检验提供有效帮助。

关键词: 甲亢; 深度学习模型; 血检指标; Vue.js; 原型系统

Deep LSTM-based hyperthyroidism forecast and application

YANG Yihao¹, WANG Mei¹, ZUO Ming²

(1 School of Computer Science and Technology, Donghua University, Shanghai 201620, China;

2 Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China)

【Abstract】 According to the stage characteristics of the development of hyperthyroidism, the LSTM deep learning model is established. By using the learned model, the patient's key blood test indicators are input to predict the future value of the patient, so as to obtain the prediction of disease development in the later stage. In addition, the forecasting application prototype system is built based on Vue.js framework, which implements a series of functions such as index data input, model call and prediction result display. The research provides effective help for doctors' current diagnosis development and treatment plan evaluation.

【Key words】 hyperthyroidism; deep learning model; blood test indicators; Vue.js; prototype system

0 引言

随着医疗信息化的快速发展, 如何充分利用长期积累的大量临床医疗数据, 进行临床医疗数据分析与挖掘, 从而辅助医生临床决策, 提高医疗服务的智能水平和医院的就诊效率, 成为广泛关注的热点问题。

甲亢疾病在内分泌疾病中发病率较高, 因内分泌系统作用会对患者全身产生影响, 其患病时间越长, 治愈也更困难^[1]。因此, 甲亢疾病的早期干预至关重要。若能在早期阶段结合其临床检查数据对疾病未来发展情况进行预测, 不仅可以方便患者了解自身病情, 同时还可辅助医生评估和制定当前诊疗方案。如用药策略的选择、是否进行一些特殊和非常规的检查, 从而对病情进行有指导的提前干预。因此, 对甲亢病情预测的研究十分必要且具有重要的临床意义。

本文基于甲亢疾病发展的阶段性特点, 建立 LSTM 深度学习模型, 利用患者的前期关键血检指标, 使用该模型对患者未来时刻指标值进行预测, 从

而获得后期疾病发展预测情况。本文使用真实数据对模型的有效性进行了验证。在此基础上, 基于 Vue.js 框架开发了发展预测应用原型系统, 实现了学习模型与应用系统的有效结合, 从而为医生当前诊疗方案的制定和检验提供帮助。

1 相关工作

甲状腺疾病是内分泌系统最常见和多发的疾病, 常引起临床各学科的关注, 其中又以甲状腺功能亢进症(甲亢)最多见。在前期研究中, Strieder 等人^[2]通过数据分析建立了简单的预测评分, 用来评估自身免疫性甲状腺病(Autoimmune Thyroid Disease, AITD)患者女性亲属患病风险。Wang 等人^[3]采用 YOLOv2 神经网络深度学习建立甲状腺结节图像自动识别与诊断系统。Yen-kung 等人^[4]使用 Cox 比例风险模型评估 Graves 病患者发生癌症的风险。现有工作主要关注于甲亢疾病的相关性分析或自动诊断, 对疾病发展预测的相关工作较少。

LSTM(长短时记忆网络)作为一种深度神经网络, 可通过隐层节点循环连接和具有门结构的神经

基金项目: 上海市科技创新行动计划(16JC1400803); 上海市信息化发展专项(201801027)。

作者简介: 杨意豪(1994-), 男, 硕士研究生, 主要研究方向: 数据库与信息检索; 王梅(1980-), 女, 博士, 教授, 主要研究方向: 数据库与信息检索; 左铭(1979-), 男, 博士研究生, 工程师, 主要研究方向: 医疗大数据、数据管理。

收稿日期: 2019-04-25

元捕捉长期时序依赖关系^[5],被广泛用来处理具有时间顺序发展特点的数据,获得了良好的效果。本文基于甲亢疾病发展的阶段性特点,构建 LSTM 深度学习模型,并基于该模型进行甲亢指标值的预测。

在系统实现上,本文使用 Vue.js 框架,该框架是一种轻量的渐进式框架。一方面,其核心库只关注视图层,易于上手,且便于与第三方库或其它项目整合;另一方面,当与单文件组件以及其它所支持各类库结合使用时,也能够为复杂的应用程序提供驱动^[6]。

2 甲亢病情预测模型

经甲状腺医学临床研究表明,甲状腺疾病的临床诊疗与 4 项脱敏血检指标密切相关,分别为游离三碘甲腺原氨酸 FT3、游离甲状腺素 FT4、促甲状腺素 TSH 和促甲状腺素受体抗体 TRAb^[7]。当 FT3、FT4 水平升高,同时 TSH 水平降低,即可诊断患有甲状腺功能亢进症。TRAb 水平作为检测由淋巴细胞产生并刺激甲状腺的免疫球蛋白-TSI 的标准,与甲亢的痊愈有着重要的关系。因此,该模型通过 4 项指标水平的变化规律预测未来的病情发展状况。

模型首先通过前若干个月的上述 4 个血检指标值预测未来某时刻的指标值,然后由该时刻的指标正常、异常判断,汇总得到病情发展状况预测。根据相关医学研究成果以及医生临床经验所知,患者在患病前六个月内的指标变化与病情未来发展有重大联系,患病后两年左右时间点的病情状况具有一定的代表意义。因此,该模型根据前六个月的 FT3、FT4、TSH、TRAb 4 项指标值预测 2 年时刻的指标值。

4 项指标的异常范围见表 1。依据表 1 可以发现 FT3、FT4、TSH 及 TRAb 指标具有不同的量纲和单位,数据范围各有差异,这会很大程度地影响模型的训练效果。因此,应当对每个指标统一范围、规范数据。标准化处理能够消除指标之间的量纲影响,使指标之间具有可比性,更适合模型训练,还可以减小指标缺失的影响。本文采用 0 均值标准化(zero_score)对每位患者的每项指标做标准化处理,该方式的计算公式如下:

$$x_{[:,j]}^* = \frac{x_{[:,j]} - \mu}{\sigma}, \quad (1)$$

其中, $x[:,j]$ 代表患者的一系列指标; μ 表示该列的均值; σ 表示该列的标准差。经过 zero_score 的处理,使数据的分布情况得以改变,将不同的特征

变换为均值为 0 且标准差为 1 的统一分布,提高了训练效率。

表 1 指标异常范围

Tab. 1 Abnormal range of indicators

指标	异常偏低	正常	异常偏高
FT3/(pmol/L)	0.5~2.62	2.63~5.70	5.71~46.08
FT4/(pmol/L)	5.15~9.00	9.01~19.04	19.05~77.22
TSH/(μ IU/mL)	0.0~0.35	0.352 9~4.935 5	4.944 2~100.0
TRAb/(U/L)	-	0.1~1.75	1.76~395.3

由于患者前六个月的指标检查记录按诊断时间先后形成了一种序列数据,为了沿着时序探索指标的变化规律,本文使用 LSTM 算法分析数据。LSTM 神经元结构如图 1 所示。

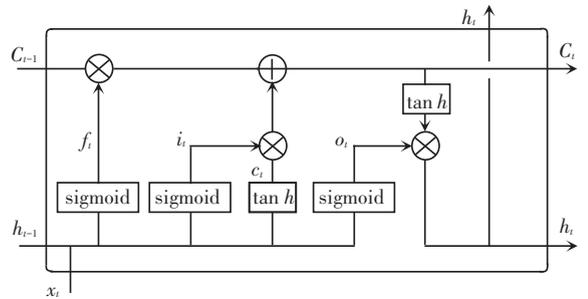


图 1 LSTM 神经元

Fig. 1 LSTM cell

其中, f_t 为遗忘门; i_t 和 c_t 为输入门; o_t 和 h_t 为输出门。多个神经元循环迭代,使信息可从当前步传递至下一步,同时通过上述门结构设计使 LSTM 可处理序列数据中的长期依赖问题。神经元具体计算公式如下:

$$f_t = \text{sigmoid} (W_f \cdot [h_{t-1}, x_t] + b_f), \quad (2)$$

$$i_t = \text{sigmoid} (W_i \cdot [h_{t-1}, x_t] + b_i), \quad (3)$$

$$c_t = \tan h (W_c \cdot [h_{t-1}, x_t] + b_c), \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * c_t, \quad (5)$$

$$o_t = \text{sigmoid} (W_o \cdot [h_{t-1}, x_t] + b_o), \quad (6)$$

$$h_t = o_t * \tan h (C_t). \quad (7)$$

将患者前六个月及 2 年时间点的指标值作为训练数据输入模型进行参数学习,完成模型训练后,可进行预测。基于该模型的甲亢病情预测算法如下:

Input: data

Output: result

set model = restore(LSTM_model)

standard_data = zero_score(data)

output = model.predict(standard_data)

result = inverse_zero_score(output)

return result

其中, *data* 表示待预测的数据, *LSTM_model* 为已经训练好的模型。首先从模型中读取已训练的参数, 其次将标准化后的数据输入模型并预测, 最后还原标准化得到预测结果。

3 应用系统框架实现

系统采用基于 Vue.js 框架的 B/S 架构。在成功启动系统并正常运行后, 可通过浏览器输入 IP 地址和端口号进行访问。本系统分为患者端和医生端, 患者端主要有用户信息管理、查看检查记录、查看诊断数据等功能; 医生端主要有用户信息管理、诊断患者病情等功能。查看诊断数据时, 患者的病情预测结果是使用基于 LSTM 算法的甲亢病情预测模型所得。该系统使用 MySQL 数据库存储用户信息和检查记录等数据。系统功能设计介绍如下:

(1) 用户管理。该功能主要是注册新用户或是修改用户名和密码, 用户分为患者和医生。

(2) 查看与增加患者检查记录。该功能主要是在患者进入系统后, 通过在线问诊上传自己的检查报告, 然后从数据库中读取该患者的血检记录, 显示到界面中供患者查看。此外, 患者还可以增加自己的血检记录, 添加成功的数据会保存到数据库中, 同时会显示到界面中。

(3) 查看患者诊断数据。该功能是在患者添加了检查记录和基本信息之后, 提交个人数据进行病情诊断, 系统读取并显示该患者的个人信息、检查报告。同时, 系统会调用保存并训练好的模型程序使用算法进行病情预测, 最后将模型自动预测结果及医生评估结果和治疗建议等显示到界面中供患者查看。

(4) 医生诊断。该功能主要是在医生进入系统后, 查看患者的检查记录并操作, 通过智能诊断, 系统便会将该患者的特征数据做相应的预处理, 然后调用模型使用算法进行预测, 并将预测结果反馈给医生。辅助医生对患者病情进行准确诊断, 并给出安全性评价和治疗建议, 诊断内容自动保存到数据库中, 以便患者查看。

图 2 为该原型系统的总体框架设计图, 该原型系统是基于 MVVM 体系结构的 Vue 项目, MVVM 体系结构具体指 View、Model 和 ViewModel。

View 层是用户界面, 包括登录界面、注册界面和诊断界面等。当用户在界面中进行登录、注册、添加数据和病情诊断等操作时, 会将控制命令传递给 ViewModel 层。ViewModel 层将按用户的操作处理

相应的数据并将结果反馈给 View 层, 待操作的数据由 Model 层从数据库中存取。整个系统通过数据层与 Model 层、Model 层与 ViewModel 层和 ViewModel 层与 View 层间的互相通信传递信息与请求。

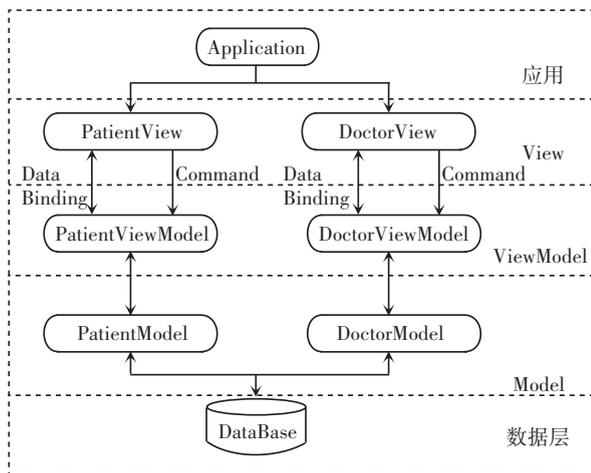


图 2 系统总体框架设计

Fig. 2 Overall frame design of the system

4 应用系统有效性评估

本文所用真实数据 (仅含指标数据, 无患者个人信息) 共包含 2 460 位患者的临床诊断数据, 将其随机划分为 2 部分, 1 960 位患者作为训练数据用于训练模型。余下 500 位患者数据输入应用系统用于系统评测。依次将 500 位患者的前六个月数据输入应用系统, 调用算法, 得到 2 年预测输出, 根据表 1 计算其正常、异常判断, 并与其 2 年附近的真实的正常、异常值判断作对比, 评估模型与系统实用性。这里偏低、正常、偏高可视为类标签, 则评估标准采用分类评价指标, 评价指标计算如下:

$$\text{准确率 Precision: } P = \frac{TP}{TP + FP}, \quad (8)$$

$$\text{召回率 Recall: } R = \frac{TP}{TP + FN}, \quad (9)$$

$$\text{F1 分值 F1-Score: } F = \frac{2 \times P \times R}{P + R}, \quad (10)$$

$$\text{正确率 Accuracy: } A = \frac{TP + TN}{TP + FP + FN + TN}, \quad (11)$$

其中, *TP* 表示为正例且预测为正例的数量; *FP* 表示为负例且预测为正例的数量; *FN* 表示为正例且预测为负例的数量; *TN* 表示为负例且预测为负例的数量。例如对于异常偏低的指标, 其正例为指标异常偏低, 负例为指标正常或指标异常偏高。评估结果见表 2。

表2 系统评测结果

Tab. 2 System assessment results %

指标	准确率	召回率	F1 分值	正确率
FT3	73	84	77	91.74
FT4	88	80	81	94.62
TSH	82	87	84	87.75
TRAb	90	90	90	89.23
AVG	83	85	83	90.84

表2所示的评测结果中,4项指标的平均正确率达到90.84%。其中,FT4指标的预测正确率达到94.62%。由表1可知,与TRAb的整个数值分布范围相比,其正常范围较小,因此TRAb的预测正确率在4个指标中相对较低,但也达到了89.23%。上述实验结果表明本文学习模型具有较高的预测准确率,本文应用系统具有一定的实际临床意义。

5 结束语

本文提出了一种基于LSTM算法的甲亢血检指标预测模型,并开发了模型应用原型系统,实现了基于模型的病情预测功能。目前,本文实现的原型系统还存在一些不足,如尚未根据预测结果为患者推荐合适的医疗资源和诊疗策略。接下来的研究中,将会关注用药分析及推荐,进一步优化系统的功能

(上接第127页)

3 结束语

本文针对近年来新提出的MNW模型采用定向搜索变异的改进人工鱼群算法进行求解,取得了预期的结果,结果表明该算法具有良好的收敛表现和应用价值。同时,由于该算法在应用过程中并未对路段容量进行限制,所以分配结果中有可能出现流量超出路段容量的情况,这也是下一步的研究重点。

参考文献

- [1] WARDROP J G. Some theoretical aspects of road traffic research [J]. Proceeding of the Institute of Civil Engineers, 1956, 1(2): 57-68.
- [2] DAGANZO C F, SHEFFI Y. On stochastic models of traffic assignment [J]. Transportation Science, 1977, 11(3): 253-274.
- [3] 黄海军. 城市交通网络平衡分析: 理论与实践[M]. 北京: 人民

与性能。

参考文献

- [1] WANG Y, SUN X L, WANG C L, et al. Influence of screening and intervention of hyperthyroidism on pregnancy outcome [J]. European Review for Medical & Pharmacological Sciences, 2017, 21(8): 1932-1937.
- [2] STRIEDER T G A, TIJSSSEN J G P, WENZEL B E, et al. Prediction of progression to overt hypothyroidism or hyperthyroidism in female relatives of patients with autoimmune thyroid disease using the Thyroid Events Amsterdam (THEA) Score [J]. Archives of Internal Medicine, 2008, 168(15): 1657-1663.
- [3] WANG Lei, YANG Shujian, YANG Shan, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network [J]. World Journal of Surgical Oncology, 2019, 17(1): 12.
- [4] CHEN Y K, LIN Chengli, CHANG Y J, et al. Cancer risk in patients with graves' disease: A nationwide cohort study [J]. Thyroid, 2013, 23(7): 879-884.
- [5] SARRAF S. French word recognition through a quick survey on recurrent neural networks using long-short term memory RNN-LSTM [J]. American Scientific Research Journal for Engineering Technology & Sciences, 2018, 39(1): 250-267.
- [6] 王志任. 基于Vue.js的开发平台的设计与实现[D]. 广州: 广东工业大学, 2018.
- [7] 张红菊, 秦东平. 甲硫咪唑治疗甲状腺功能亢进症的效果及对FT3、FT4、TSH水平的影响 [J]. 临床医学研究与实践, 2019, 4(9): 42-44.
- [8] 交通出版社, 1994.
- [4] CASTILLO E, MENENDEZ J M, JIMÉNEZ P, et al. Closed form expressions for choice probabilities in the Weibull case [J]. Transportation Research Part B: Methodological, 2008, 42(4): 373-380.
- [5] KITTHAMKESORN S, CHEN A. A path-size weibit stochastic user equilibrium model [J]. Transportation Research Part B: Methodological, 2013, 57: 378-397.
- [6] 刘炳全, 孙广才. 基于Logit分配的交通网络设计模型的改进粒子群算法 [J]. 科学技术与工程, 2008, 8(19): 5446-5450, 5456.
- [7] 张福龙. 基于最大最小蚁群算法的随机用户交通分配模型研究 [D]. 西安: 长安大学, 2016.
- [8] 刘建美, 马帅奇. 关于求解随机用户均衡问题的截断拟牛顿型信赖域法研究 [J]. 数学的实践与认识, 2017, 47(10): 184-190.
- [9] 魏秋月. 基于蚁群优化的随机用户均衡模型研究 [D]. 西安: 长安大学, 2017.
- [10] 姜山, 季业飞. 改进的人工鱼群混合算法在交通分配中的应用 [J]. 计算机仿真, 2011, 28(6): 326-329.