

基于卷积神经网络和 LSTM 循环神经网络的客户复购预测方法^①

施海昕^{②*} 诸建超^{**} 严骏驰^{③***} 程 栋^{**} 刘云锋^{*}

(* 云汉芯城(上海)互联网科技股份有限公司 上海 201612)

(** 华东师范大学计算机科学与软件工程学院 上海 200062)

(*** 上海交通大学电子信息与电气工程学院 上海 200240)

摘要 通常一个销售代表会有数百名客户。销售代表无法定量预测哪位客户最近有下单需求,所以多采用轮询或者主观直观的方式决定每天的回访客户名单。本文以深度学习的思路,把销售代表的历史回访记录作为输入数据,以卷积神经网络(CNN)和循环神经网络(RNN)为基础模型,让模型学习客户复购行为的内在逻辑以便指导初级销售代表的每日回访工作。相对传统回访策略,在深度学习算法辅助下的回访策略可以把回访相对成功率提高 120%,取得了显著的效果。

关键词 循环神经网络(RNN);长短期记忆网络(LSTM);卷积神经网络(CNN);深度学习

0 引言

随着互联网技术的发展与普及,国内电子分销行业在电商化方面取得了长足的发展。在一个典型的电商场景中,当客户访问平台网站时,系统后台会给每一位客户分配一位专属销售代表,以便为客户提供一对一的优质服务。在客户首次采购之后,专属销售代表除了保证首次采购的商品按时交付以外,也会在后续积极联系客户以便促成再次采购。

根据当前收集的客户反馈信息,客户采购行为是基于研发或者生产计划的,是有一定时间序列规律的。在客户即将采购时,销售代表和客户的积极沟通可以促成客户的采购决策。由于每位销售代表需要服务数百位客户,而每天销售代表的时间是有限的,若为每个客户平均分配时间又不切实际,因此,在日常工作中,应该把时间花在回访哪些客户上才能产出最多的价值是一个亟待解决的问题。实际工

作中,销售代表无法定量预测哪位客户最近会下单,因此通常采用名单轮询或者主观直观的方式来决定每天回访的客户名单,这种方式十分依赖销售本身的从业经验,效率较低,不利于业务的进一步发展。

因此,本文选择采用基于深度神经网络的监督学习方法,提出一种 CNN-LSTM 组合模型,通过时间序列预测出客户的复购可能性,即对每个客户在未来的下单可能性进行预测,最终实现得以接近或达到高级销售代表水平的模型,以此来指导初级销售代表的每日回访工作。

时间序列预测是本文所提出方法的本质所在,其基本思想是通过分析历史数据,学习出其中时序上的关联性,用以推断出下一个时刻或者未来某个时间段内目标客户的购买趋势。在深度学习^[1-2]中,循环神经网络^[3-4](recurrent neural network, RNN)经常被用来进行时序预测,但是早期的 RNN 具有梯度消失等问题^[5]导致其效果并不好,为了解决这些问题,长短期记忆网络^[6](long short-

① 科技创新 2030“新一代人工智能”重大项目(2018AAA0100704)和国家自然科学基金面上项目(61972250)资助。

② 男,1980 年生,硕士;研究方向:人工智能,数据挖掘;E-mail:shihx@ickey.cn

③ 通信作者,E-mail:yanjunchi@sju.edu.cn

(收稿日期:2020-10-22)

term memory, LSTM) 以及门循环单元网络^[7] (gate recurrent unit, GRU) 便应运而生, 在多种类型的任务中均取得了 SOTA (state-of-the-art) 的效果, 比如 Tree-LSTM^[8] 就在语义相关性和语义分类两个任务上均取得了超越当时其他算法的优异性能。

本文使用 LSTM 来提取时序相关特征, 使用卷积神经网络(convolutional neural network, CNN) 来提取时序无关特征。CNN 可以很好地识别出不同类型数据中的多种特征, 比如图像卷积指的是 2D-CNN, 用于对图像数据提取特征。LeNet^[9] 作为真正意义上的 CNN 用来对手写体数字进行识别, 其中浅层 CNN 提取比较基础的特征, 比如点、边缘, 而深层 CNN 可以将这些浅层特征组合成更加高级的特征, 比如由线条组成的眼睛、鼻子。对于客户画像等一维数据, 使用一维卷积神经网络(1D-CNN) 对其进行特征提取, 与 2D-CNN 类似, 二者的关键区别只在于输入数据的维度以及卷积核在数据上的滑动方式。

本文将 LSTM 和 CNN 作为基础模型架构, 把销售代表历史回访记录作为输入数据, 并将其划分为时序相关特征和时序无关特征, 分别使用 LSTM 和 CNN 得到两种类型的特征表示, 而后将二者进行整合, 最终通过一个简单的分类器得到预测结果, 即对于每个客户, 预测其在未来可能会下单的概率。通过将算法实际部署上线, 其有效性得到了进一步证明, 展示了深度学习技术在这一方面的可行性和进一步发展的潜力, 对于该领域甚至其他领域的相关问题具有指导意义。

1 相关工作

对于用户预测分析的研究早已屡见不鲜, 其中包括基于传统机器学习的算法, 也有基于神经网络等深度学习算法的相关研究。

在传统机器学习算法中, Hawarah 等人^[10] 研究了在家庭自动化系统使用过程中的用户行为预测问题, 通过贝叶斯网络来预测和诊断用户的居住情况。Nesi 等人^[11] 通过递归划分过程提出了一种分类树, 用于对 Twitter 推文的转发情况进行分析和预测。

Gündüz 等人^[12] 提出了一种基于点击流树的模型, 用以预测用户访问网页时的下一个请求。Tseng 等人^[13] 通过一种新颖的数据挖掘方法来预测移动网络系统中的用户行为模式, 它可以高效地发现与请求的服务相关的移动用户的序列移动模式, 在准确性、执行效率和可扩展性方面均表现突出。

另一方面, 随着神经网络表现出极其优异的特征提取能力, 基于神经网络相关算法的用户行为预测更是取得了卓越的性能。Zhou 等人^[14] 提出了一种基于注意力机制的深度学习预测框架来预测用户行为, 考虑使用异构用户行为, 将所有类型的行为投影到多个语义空间中, 并在每个空间中通过自我注意力机制在行为之间产生关联和影响。Li 等人^[15] 提出了一种使用双通道 CNN-LSTM 组合模型对用户评论进行情感分析的方法。Dutta 等人^[16] 同样提出了一种 CNN-LSTM 组合模型, 用于帮助临床医生们在短时间内快速验证 COVID-19 患病情况, 并对 COVID-19 未来疫情的发展趋势进行有效预测。

需要指出的是, CNN-LSTM 组合模型虽然并不是本文首创, 但却是最早被使用对销售回访数据进行特征工程和相关建模, 进而对客户复购行为进行预测的研究成果。

2 数据准备

本文采用了云汉芯城(上海)互联网科技股份有限公司(下简称云汉芯城)的实际运营数据。云汉芯城是电子分销行业的重要互联网平台。

2.1 原始特征

云汉芯城对销售代表的要求是回访客户后需要在 ERP 中填写回访记录, 后续还有销售运营人员把销售代表的回访记录与实际电话录音进行抽检对比, 因此数据库中的回访记录数据质量较高。

回访记录中的字段包括回访方式(如电话、QQ、上门拜访等)、回访对象(首单客户、3 个月内未下单、3 个月内未下单等)、回访目的(挖掘需求等)和回访结果(暂无采购需求、成功拿到订单)。

除了回访记录以外, 数据库中还有客户的画像数据, 其中的字段包括客户类型(贸易商、生产厂

家、研发方案提供商等)、客户行业(消费电子、汽车电子、医疗电子等)和客户标签(重点客户、高潜质客户、中小型客户等)。

当用这些历史数据构造训练和测试集时,本文将一次回访的记录作为一个样本,上述提到的字段中的回访结果(能/否成单)作为样本标签,其余所有字段作为样本特征。

2.2 生成新特征

由于回访行为是发生在时间维度上的,以原始数据为基础还可以构造一些新特征来帮助深度学习捕捉更多信息,增强模型的泛化能力。所构造的新特征包括回访问隔天数、此事件本次回访结果、此事件上次回访结果、是否成功拿到过订单、成功拿到订单数、回访次数、回访过的销售代表人数。这些新特征是以回访记录作为基本单位的,即对于每一条记录都将会扩展出以下 7 个新特征。

(1) 间隔天数 = 本次回访日期 - 上次回访日期;

(2) 此事件本次回访结果 = 给定一个历史回访事件,它的回访结果;

(3) 此事件上次回访结果 = 给定一个历史回访事件,此事件的上一个事件的回访结果(同一客户);

(4) 是否成功拿到过订单 = 到本次回访为止,历史上该客户是否有因回访下单;

(5) 成功拿到订单数 = 到本次回访为止,该客户因回访下单的累计次数;

(6) 回访次数 = 到本次回访为止,该客户累计被回访的次数;

(7) 回访过的销售代表人数 = 到本次回访为止,回访过该客户的不同销售代表的累计人数。

3 训练模型

基于上述实际数据以及按照业务经验和专家知识所抽取的相关特征,可以进行下一步的机器学习建模和预测。

3.1 特征分类

根据特征是否时序相关,把特征分成时序相关

特征和时序无关特征,分别用于 LSTM 和 CNN 网络的输入。

时序相关特征是与时序信息高度相关的特征,该特征在实际每一次回访后会不断产生。时序相关特征包括回访过的销售代表人数、回访次数、成功拿到订单数、回访问隔天数、此事件本次回访结果、此事件上次回访结果、该销售是否成功拿到过订单。

时序无关特征则是一类相对稳定特征,值很少随着每次回访事件而更新。主要包括客户画像数据所构造的特征,如客户标签、客户行业、客户类型、回访对象等。

3.2 数据预处理

假设每个客户的采购决策互相独立,把同一个客户在时间维度上的所有记录进行聚合,最初的训练集首先按照客户 id 进行分组。

对于序列长度的取值,从 3 到 15 做了遍历测试,最终取得最佳序列长度为 10,原因分析如下:如果序列过短,则表示时间跨度较短,这会导致模型无法提取到足够的时序信息;而如果序列过长,则容易丢失前后依赖关系,这与 LSTM 对传统 RNN 的改进类似,即序列后部的梯度很难反向传播到前面的序列,导致梯度消失。因此,把每一个客户的回访数据都尽量标准化为 10 条记录,即对同一个客户,将他的回访记录按照时间顺序排列后,以滑动窗口的方式取出每 10 条记录作为一个事件序列。

具体而言,本文考虑如下两种情况。

(1) 如果对于单一客户的回访记录不足 10 条,将会直接舍弃这些记录,因为这种数据对于训练模型来说是有害的。

(2) 对于单一客户回访记录大于 10 条的情况,将按照长度为 10、步长为 1 的滑动窗口将其分割为多个事件序列,如图 1 所示。假设客户 A 总共有 12 条记录,则可以将其分成 1~10、2~11 以及 3~12 这 3 个事件序列。

通过大量的实验发现,上述的数据质量控制策略取得了一定的成效,保证了结果的可信性和模型质量。

值得注意的是,对于连续型特征,可以直接使用数据库中的字段值作为特征值。对于离散型特征,

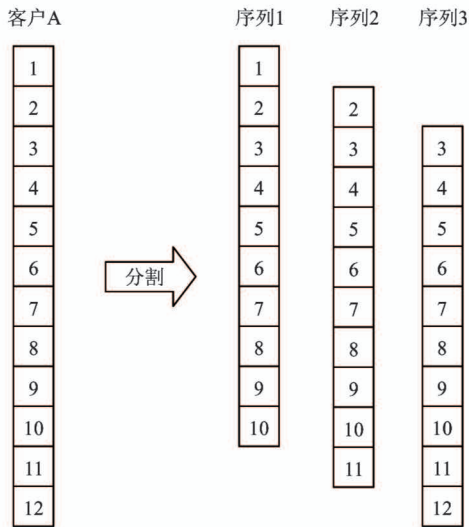


图1 将长度为12的序列分割为3个长度为10的序列

则需要将其进行转换,将文字映射到离散型数值上,以使计算机理解这个特征的意义,如此才能输入到网络中进行训练。

此外,考虑到不同特征之间的数值取值范围有较大差异,比如“间隔天数”这个连续型特征,少则几天,多则上百天,而“回访结果”这个离散型特征,只有1和0两个取值,表示回访是否成功。对于各种不同类型的数据,如果直接输入模型中进行训练的话,由于数值跨度太大,会导致模型梯度回传过程中出现各种问题,不利于模型的收敛。因此,需要通过某些方法将所有数据归一化到0和1之间,本文中使用的是Min-Max标准化方法,具体来说,对于某个特征的每个数据 $x \in X$,通过式(1)进行映射。

$$x' = (x - \min(X)) / (\max(X) - \min(X)) \quad (1)$$

其中 X 表示当前特征的所有数据。假设数据集有 N 个样本,每个样本有 M 个特征,那么该数据集就对应一个 $N \times M$ 的输入矩阵,每一行表示一个样本,每一列表示某个特定的特征,而Min-Max标准化就是针对 M 列中的每一列所进行的操作。

当所有特征都标准化之后,将数据集按照64个事件序列作为一个批次进行迭代训练。这里批次大小是一个超参数,值越大,所需的内存空间就越大,训练越快,且所提取到的特征越全面,反之亦然。此处根据多次实验测试,将其设置为64时取得最优效果。

3.3 定义模型

本文使用的模型结合了经典的LSTM和CNN模型,如图2所示。由于输入数据分为时序相关特征和时序无关特征,对这两部分数据分别进行处理。对于时序相关数据,先用CNN预处理提取一次特征,然后再用特别适合处理时序信息的LSTM进行二次处理。而时序无关特征则直接用CNN来处理。

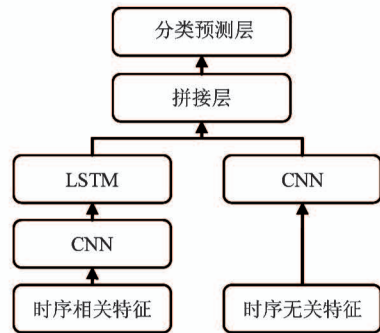


图2 神经网络结构示意图

处理时序相关特征用到了一个CNN-LSTM的组合模型,其中CNN部分是一维卷积,详细结构见图3。训练集中的每条回访记录都是一个事件,每个事件有7种特征,将每10个事件构成了一个 10×7 的事件序列矩阵,另外,每个训练批次的大小为64,故CNN的输入数据维度为 $64 \times 10 \times 7$,图3中省略了训练批次这一维度。

可以看到,图中的卷积核实际是二维的,只需要指定纵向的卷积核长度即可,而卷积核横向的维度是固定的,与输入事件的特征数相同。因为对于事件序列来说,每个事件的所有特征作为一个整体,卷积的移动方向只能沿着纵向的列维度进行移动。本文定义了64个卷积核,表示在网络的第1层中可以提取到64个不同的特征,其中每个卷积核的长度为2,以步长为1向下平移。因此,CNN的输出特征尺寸为 $64 \times 10 \times 64$,其中第1个64表示训练批次的大小,第2个64表示提取了64个特征。

在每一层CNN之后,还需要再对所提取的特征进行标准化操作,使得每一层输出数据的分布处于一个大致相同的范围,以加速模型的收敛。此处选择批标准化^[17](batch normalization),即对一个批次内的所有数据计算每个维度上的均值 μ 和方差 σ^2 ,

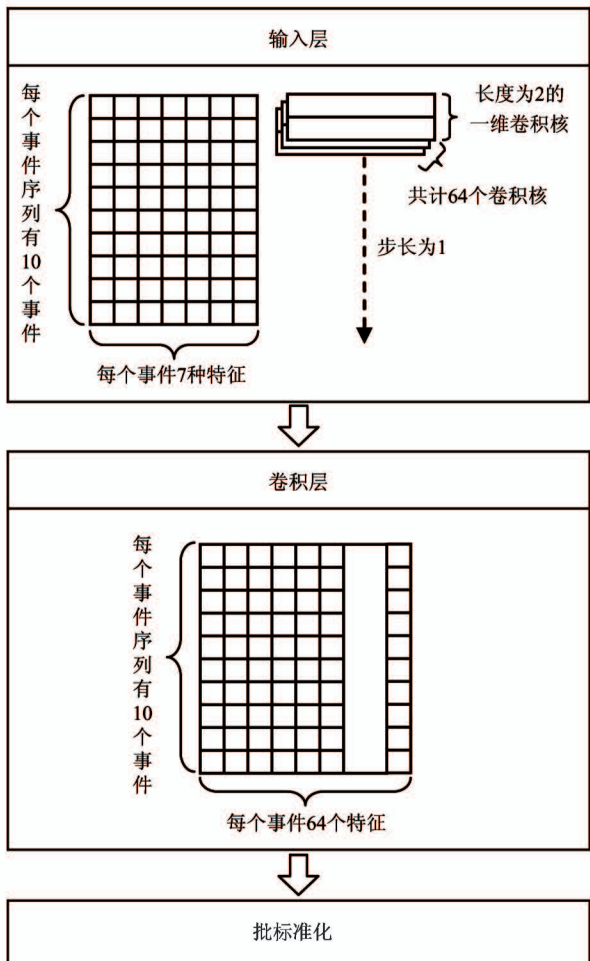


图 3 处理时序相关特征的 CNN

并在模型训练过程中学习出缩放系数 γ 和偏移系数 β ,按照式(2)将输入 x 标准化为输出 y 。

$$y = \gamma \times (x - \mu) / (\sigma^2 + \varepsilon)^{1/2} + \beta \quad (2)$$

其中的 ε 是一个极小的值,为了防止分母为 0。

批标准化之后的数据将会接着进入到如图 4 所示的 LSTM 中,输入数据尺寸为 $64 \times 10 \times 64$,图中同样省略了批次这个维度。按照事件序列维度将数据切分为 10 组, $T_1, T_2, T_3, \dots, T_{10}$,每组的特征长度为 64,将其送入到一个有 3 层隐藏层的 LSTM 网络中,隐藏层单元数为 128,故可以得到尺寸为 $64 \times 10 \times 128$ 的时序特征输出。最后,将其输入到一个全连接层得到长度为 2 的向量,用来表示时序相关部分数据的整体特征,作为 CNN-LSTM 模块的总输出。

另一方面,时序无关特征会直接通过另一个一维 CNN 模型进行处理,其与图 3 中时序相关特征所

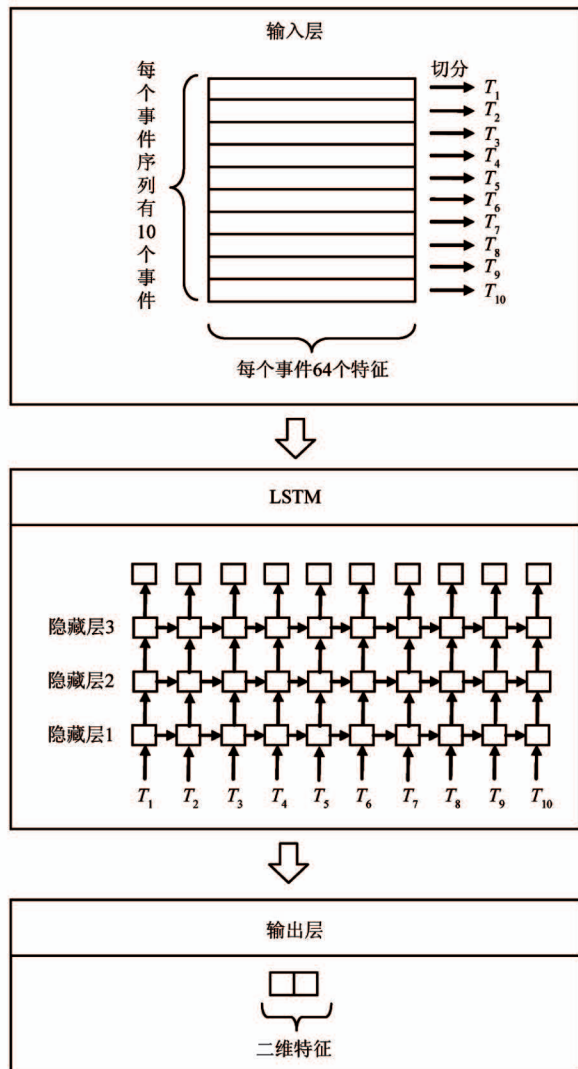


图 4 处理时序相关特征的 LSTM

用的 CNN 非常相似。此处时序无关特征一共有 4 个,本文同样定义了 64 个长度为 2 的一维卷积核,以步长为 1 向下平移,因此,输入数据的维度就是 $64 \times 10 \times 4$ 。最后通过一个全连接层得到长度为 2 的向量,作为时序无关部分数据的特征表示。

如图 2 所示,将时序相关的 CNN-LSTM 模块的输出向量与时序无关的 CNN 模块的输出向量进行拼接,再通过一个全连接层,接一个 Softmax 激活函数得到两个概率值,分别表示当前时刻客户是/否下单的概率。

3.4 损失函数

训练集样本中每一条记录都是一次回访结果。根据实际情况可以预见,回访结果为失败的记录将会远多于回访结果为成功的记录。这就是典型的正

负样本不均衡案例。如果不妥善解决,大量的负样本会对损失函数起主要贡献,从而主导梯度的更新方向,致使模型无法学习到充分的正样本特征。为了解决这个问题,选用 Focal Loss^[18] 损失函数,其定义如式(3)所示。

$$FL(p_i) = -\alpha_i(1-p_i)^\gamma \log(p_i) \quad (3)$$

其中 p_i 表示模型对于样本的预测概率;加权因子 α_i 用来平衡正负样本的重要性,当正样本远多于负样本时,可以取 $\alpha_i = 0.25$,以减小正样本对于损失函数的影响;此外,公式中还有一个参数 γ 被称为调制因子,用以降低易分类样本的权重,使模型更加专注于训练难分类样本,实验中发现 $\gamma = 2$ 时取得最优效果。

3.5 优化算法

本文整体客户复购预测的算法如算法1所示。使用自适应矩估计^[19](adaptive moment estimation, Adam)作为本文模型的优化算法,这是一种可以替代传统随机梯度下降过程的一阶优化算法,它能够根据训练数据迭代更新神经网络的权重参数。Adam 通过计算梯度的一阶矩估计和二阶矩估计为不同的参数设计独立的自适应学习率,每次迭代参数的学习率都会有一定的范围,因此不会由于梯度很大而导致学习率也变得很大,进而保证参数的值相对比较稳定。

4 评估模型

4.1 用历史数据评估理论效果

在模型研发过程中,由于其还无法正式上线与传统回访策略的结果进行有效的对比和评估,本文调取了数据库中的历史回访数据,试图通过历史数据去评估、调试并优化模型性能,其整体思路如图5所示。

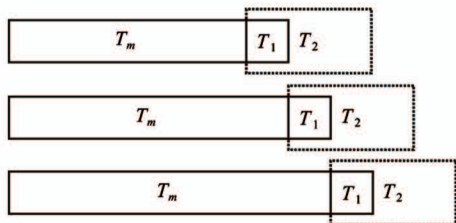


图5 基于滚动预测的评估模型的机制

算法1 客户复购预测算法

输入:学习率 $\alpha = 0.001$;一阶矩估计指数衰减率 $\beta_1 = 0.9$,一阶矩估计指数衰减率 $\beta_2 = 0.999$; $\varepsilon = 10^{-8}$;目标函数 $f(\theta)$
 初始化:目标函数初始参数 θ_0 ;初始时间 $t \leftarrow 0$;初始一阶矩 $m_0 \leftarrow 0$;初始二阶矩 $v_0 \leftarrow 0$
 1: 若 θ 未收敛,则循环以下步骤:
 2: 输入一批训练数据 x 及对应标签 y
 3: 通过图2中分类预测层的输出来计算目标函数 $f(\theta)$ 的值
 4: 通过链式法则计算神经网络中每个参数的梯度 $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$
 5: 更新带偏置的一阶矩估计 $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$
 6: 更新带偏置的二阶矩估计 $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$
 7: 计算偏置修正的一阶矩估计 $m'_t \leftarrow m_t / (1 - \beta_1^t)$
 8: 计算偏置修正的二阶矩估计 $v'_t \leftarrow v_t / (1 - \beta_2^t)$
 9: 更新参数 $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot m'_t / ((v'_t)^{1/2} + \varepsilon)$
 10: θ 收敛后可得到最优神经网络模型 $M(\theta)$
 11: 将真实数据输入网络 $M(\theta)$ 中即可得到对于现在或未来某一时刻,客户是否会下单的预测结果

使用 T_m 时间段内的回访数据作为训练集, T_1 时间段内的回访数据作为测试集。对于在 T_1 时间段内的客户,使用算法预测出其在 T_2 时间段内可能回访成功的用户列表 L_{pred} ,并根据 T_2 时间段真实成单的用户列表 L_{gt} ,计算出算法的查准率(precision),即算法预测出的正样本中真正的正样本所占的比例,将式(4)作为算法预测成功率的评估指标。

$$P = (L_{pred} \cap L_{gt}) / L_{pred} \quad (4)$$

在用历史数据评估模型的阶段,经过大量实验验证,最终选取 $T_1 = 1$ 周, $T_2 = 4$ 周较为合适,即对某一周中的所有用户,去预测当前周以及后3周中这些用户的成单可能性,以此来评估该模型的准确率。具体来说,假设用2018年3月1日到2018年12月30日的历史数据作为训练集,预测出2019年第1周预计回访成功的客户名单,结合2019年前4周的真实下单数据计算算法的预测成功率(A_1),将其与2019年第1周销售们自主策略的回访成功率(B_1)进行对比。同理,将2018年3月1日到2019年1月6日的历史数据作为训练集,来预测2019年第2周的回访名单,分别记录算法成功率(A_2)和实际成功率(B_2),以此类推。

表 1 记录了这 14 次评估中的详细对比数据,并计算得到算法相对效果提升的百分比,据此在图 6 中绘制了历史数据评估中的回访成功率曲线。其中虚线为算法预测成功率,实线为销售自主回访的实际成功率。从中可以看到,相比销售按照自己的策略进行回访,该算法的相对成功率平均提高了 60% 左右。

表 1 算法成功率 vs. 实际成功率

算法成功率	实际成功率	算法相对效果提升
16.80%	6.40%	162.50%
16.17%	7.93%	103.90%
10.58%	6.18%	71.20%
8.66%	8.15%	6.30%
4.93%	8.05%	-38.80%
5.84%	3.41%	71.40%
10.66%	5.33%	100%
10.91%	9.48%	15.10%
11.55%	4.63%	149.40%
12.91%	5.84%	121%
11.98%	7.91%	51.40%
13.47%	12.38%	8.80%
9.10%	9.67%	-5.90%
5.26%	4.19%	25.60%
算法相对效果提升的平均值		60.14%

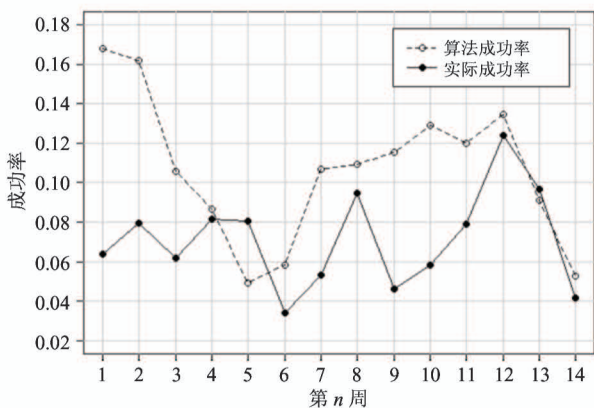


图 6 历史数据评估中的回访成功率曲线

4.2 增加训练集数据优化模型

通过历史回访数据证明模型的有效性之后,正式模型第一版于 2020 年 2 月 24 日部署上线,在实际工作过程中采用了本文所提出的基于深度学习的

回访策略推荐算法。由于时间仓促,第一版模型的数据训练集是固定的,采用从 2018 年 3 月到 2019 年 10 月的初级销售的历史回访记录。

2020 年 3 月 24 日,第二版上线,实现了一个脚本每天凌晨自动将新数据加入训练集来训练模型。这样训练集的时间范围变为 2018 年 3 月到当天的前一天。理论上这样做可以不断增强模型的鲁棒性,因为对于深度学习来说,更多的训练数据可以更有效地防止模型的过拟合问题,从而使模型能够学习到更多回访数据的时序特征并做出更有效的预测。同时,训练集的空间范围也从仅初级销售扩展到所有级别的销售代表。因为中级销售和高级销售所具备的筛选客户的能力也是可以为初级销售所用的。

安排 58 名初级销售全程进行自主回访,比较第一版和第二版模型的名单回访成功率。针对第一版模型,计算在名单中并且 7 天内回访成功的客户人数 a_1 占出现在名单中并且 7 天内回访总人数 b_1 的比例,将 a_1/b_1 作为第一版模型的成功率;同理,针对第二版模型,定义名单中并且 7 天内回访成功的客户人数为 a_2 ,名单中且 7 天内回访总人数为 b_2 ,比例 a_2/b_2 就作为第二版模型的成功率。完整的统计结果如表 2 所示。

表 2 两个版本模型的对比

	时间段	成功率	平均成功率
第 一 版	2020/2/24 - 2020/3/1	14.6%	10.9%
	2020/3/2 - 2020/3/8	14.5%	
	2020/3/9 - 2020/3/14	6.8%	
	2020/3/15 - 2020/3/21	10.2%	
第 二 版	2020/3/22 - 2020/3/23	8.7%	24.2%
	2020/3/24 - 2020/3/30	26.9%	
	2020/4/1 - 2020/4/7	26.0%	
	2020/4/8 - 2020/4/14	21.5%	
	2020/4/15 - 2020/4/22	24.5%	
	2020/4/23 - 2020/4/24	28.6%	

可见,第一版模型的成功率都在 15% 以内,平均成功率为 10.9%;第二版模型的回访成功率都维持在 20% 以上,平均成功率为 24.2%,相对效果提升 122%,效果显著。此外,从每周的数据来看,并无按周递增的情况,说明成功率的提升并非自然增长的结果,而是模型带来的增长。

4.3 评估实际效果

随后,开始评估第二版模型的实际使用效果。选择从2020年3月24日到2020年4月24日,安排58名初级销售全程进行自主回访,比较在名单中与名单外客户的回访成功率。

已知一个客户被回访过,看从回访当天到7天前是否出现在推荐名单中。如果出现过则为名单内客户,总数为 b_3 ;否则为名单外客户,总数为 b_4 。在名单内且成功下单,定义其为名单内回访成功的客户,总数记为 a_3 ;同理,名单外回访成功的客户总数记为 a_4 。则 a_3/b_3 和 a_4/b_4 分别表示算法使用后名单内、名单外客户的回访成功率。详细统计结果记录于表3中。

表3 评估效果详细统计结果

	推荐名单内 回访成功率	销售 人数	推荐名单外 回访成功率
总计	58	11.05%	24.17%
入职年数	1年以内	34	4.73%
	1~2年	20	11.37%
	2年以上	4	6.12%
年龄段	21~25	27	5.92%
	26~30	21	7.66%
	30以上	10	6.46%
性别	男	33	6.25%
	女	25	7.19%

从统计结果可见,推荐名单外的回访成功率仅为11.05%,而推荐名单内的回访成功率高达24.17%,相对效果提升119%。

除此之外,为了细致分析这58名参与算法评估的初级销售本身的能力对于回访成功率的影响,本文详细统计了他们的相关信息以及对应的回访成功率于表3中展示,从入职年数、年龄段以及性别这三方面进行了统计。

首先是“入职年数”角度,入职1~2年的销售具有更丰富的回访经验,使用算法后将原先本就已非常高的成功率再提升1倍;而刚入职一年以内的销售因为经验的欠缺导致原先的回访成功率极低,所以算法对他们的回访带来的收益巨大,在算法辅

助下绩效提升最明显,直接将成功率提升了2倍以上;同时,对于入职2年以上的初级销售业绩也有1倍的提升。

其次,从不同“年龄段”可以分析出,较年轻(30岁及以下)的销售具有较大的发展潜力和成长空间,算法的帮助对他们来说如虎添翼;而对于年长(30岁以上)却还处于初级的销售,他们属于绩效不佳者,算法带来的收益有限。

最后,按照“性别”分类,可以看到原先男女销售对于成功回访的转化率上,女性略占优势;更进一步地,在算法的帮助下,二者均有较大程度的提升,其中男性销售收益更大,甚至反超女性成功率。

5 结论

本文提出了一种基于CNN-LSTM组合神经网络的客户复购预测方法,通过实验环境测试以及实际生产环境中的评估,共同证明其具有较高实用价值。

具体而言,原有的客户回访策略为随机回访,销售们对他们每次回访的结果都缺乏统一、专业、量化的评估结果,因此采取的措施往往有较大的主观性;而本算法可以预测出的高潜力的客户名单,销售们可以据此有针对性地进行话术的准备。同时这份名单会在他们潜意识中植入一个观念:只要身体力行,在客户回访上投入更多精力,就有望得到预期中乃至超出预期的回报提升销售的工作热情。所以,本文的模型评估结果,一方面证明了算法本身具有足够的有效性,另一方面也给销售们心理上带来了较大的积极作用。

本文通过深度学习,发掘出销售代表回访过程以及客户本身的一些潜在规律,验证表明初级销售按照模型提供的回访名单可以显著地提高回访成功率。后续考虑采用时序点过程^[20-25]等方法进一步改进模型,从另一角度提取出更有效的时序特征用于客户的复购预测。

参考文献

[1] Goodfellow I, Bengio Y, Courville A. Deep Learning [M]. Cambridge: MIT Press, 2016

- [2] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444
- [3] Jordan M I. Serial order: a parallel distributed processing approach[J]. *Advances in Psychology*, 1997, 121(97): 471-495
- [4] Elman J L. Finding structure in time[J]. *Cognitive Science*, 1990, 14(2): 179-211
- [5] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks [C] // International Conference on Machine Learning, Atlanta, USA, 2013: 1310-1318
- [6] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780
- [7] Cho K, van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014: 1724-1734
- [8] Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks [C] // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015: 1556-1566
- [9] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition [J]. *Neural Computation*, 1989, 1(4): 541-551
- [10] Hawarah L, Ploix S, Jacomino M. User behavior prediction in energy consumption in housing using Bayesian networks [C] // International Conference on Artificial Intelligence and Soft Computing, Berlin, Germany, 2010: 372-379
- [11] Nesi P, Pantaleo G, Paoli I, et al. Assessing the retweet proneness of tweets: predictive models for retweeting [J]. *Multimedia Tools and Applications*, 2018, 77(20): 26371-26396
- [12] Gündüz Ş, Özsu M T. A web page prediction model based on click-stream tree representation of user behavior [C] // Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington DC, USA, 2003: 535-540
- [13] Tseng V S, Lin K W. Efficient mining and prediction of user behavior patterns in mobile web systems [J]. *Information and Software Technology*, 2006, 48(6): 357-369
- [14] Zhou C, Bai J, Song J, et al. Atrank: an attention-based user behavior modeling framework for recommendation [J]. *arXiv:1711.06632*, 2017
- [15] Li W, Zhu L, Shi Y, et al. User reviews: sentiment analysis using lexicon integrated two-channel CNN-LSTM family models [J]. *Applied Soft Computing*, 2020, 94: 106435
- [16] Dutta S, Bandyopadhyay S K, Kim T H. CNN-LSTM model for verifying predictions of Covid-19 cases [J]. *Asian Journal of Research in Computer Science*, 2020, 5(4): 25-32
- [17] Ioffe S, Szegedy C. Batch Normalization: accelerating deep network training by reducing internal covariate shift [C] // International Conference on Machine Learning, Lille, France, 2015: 448-456
- [18] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017 (99): 2999-3007
- [19] Kingma D P, Ba J. Adam: a method for stochastic optimization [J]. *arXiv:1412.6980*, 2014
- [20] Du N, Dai H, Trivedi R, et al. Recurrent marked temporal point processes: embedding event history to vector [C] // The 22nd ACM SIGKDD International Conference, New York, USA, 2016: 1555-1564
- [21] Yan J, Xiao S, Li C, et al. Modeling contagious merger and acquisition via point processes with a profile regression prior [C] // International Joint Conference on Artificial Intelligence, New York, USA, 2016: 2690-2696
- [22] Xiao S, Yan J C, Yang X K, et al. Modeling the intensity function of point process via recurrent neural networks [C] // The 31st AAAI Conference on Artificial Intelligence, San Francisco, USA, 2017: 1597-1603
- [23] Yan J, Liu X, Shi L, et al. Improving maximum likelihood estimation of temporal point process via discriminative and adversarial learning [C] // International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 2948-2954
- [24] Wu W, Yan J, Yang X, et al. Decoupled learning for factorial marked temporal point processes [C] // Proceedings of the 24th ACM SIGKDD International Conference

on Knowledge Discovery and Data Mining, London, UK,
2018: 2516-2525

[25] Xiao S, Yan J, Farajtabar M, et al. Learning time series

associated event sequences with recurrent point process
networks[J]. *IEEE Transactions on Neural Networks and
Learning Systems*, 2019, 30(10): 3124-3136

A prediction method of clients' repurchase based on CNN and LSTM RNN

Shi Haixin^{*}, Zhu Jianchao^{**}, Yan Junchi^{***}, Cheng Dong^{**}, Liu Yunfeng^{*}

(^{*} ICkey (Shanghai) Internet and Technology Co. Ltd., Shanghai 201612)

(^{**} School of Computer Science and Software Engineering, East China Normal University, Shanghai 200062)

(^{***} School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240)

Abstract

Each sales representative usually has hundreds of customers. Sales representatives cannot quantitatively predict which customer will place an order recently, so they often use polling or intuitive methods to decide the list of visiting customers every day. This paper takes the historical visit records of sales representatives as input data, and uses convolutional neural network (CNN) and recurrent neural network (RNN) to make the model learn the internal logic of customer repurchase behavior, in order to guide the daily jobs of junior sales representatives' jobs. Compared with the traditional visit strategy, the new visit strategy of deep learning has achieved good results and increases the relative success rate of visit by about 120%.

Key words: recurrent neural network (RNN), long short-term memory (LSTM) network, convolutional neural network (CNN), deep learning