

利用时间过程和内容特征的深度融合对在线文章进行人气预测

廖东亮,¹ 徐进,^{1*} 李公甫,¹ 黄伟杰,¹ 刘伟庆,² 李静²

¹微信, 腾讯公司。

{brightliao, jinxxu,

gongfuli, wainhuang}@tencent.com; ² 中国科学技术
大学, cslwqxx@mail.ustc.edu.cn, lj@ustc.edu.cn

摘要

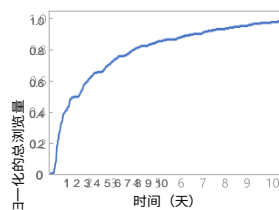
预测在线文章的受欢迎程度为许多应用提供了启示, 如推荐、广告和信息检索。然而, 要发展最佳的预测能力, 有几个技术难题需要解决。(1) 流行度在外部因素的影响下会发生波动, 而这些因素是不可预测的, 也是难以捕捉的。

(2) 内容和元数据特征在很大程度上决定了在线内容的受欢迎程度, 这些特征通常是多模式的, 而且非微不足道的模型。(3) 此外, 它还需要弄清楚如何融合时间过程和内容特征建模, 在网络文章的不同生命周期阶段进行流行度预测。在本文中, 我们提出了一种深度融合时间过程和内容特征 (DFTC) 的方法来解决

这些问题。对于时间流行过程的建模, 我们采用了递归神经网络和卷积神经网络。对于多模式的内容特征, 我们利用了分层注意网络和嵌入技术。最后, 一个临时性的注意力融合被用于动态地整合所有这些部分。使用从微信中收集的数据集, 我们表明所提出的模型在人气预测方面明显优于现有的方法。

简介

在线文章, 如门户网站的新闻和社交网络的博客, 已经成为最重要的信息来源。在线文章的受欢迎程度描述了它受到的关注程度, 可以用总浏览量来衡量。对于内容提供者来说, 受欢迎程度是衡量内容质量的一个标准, 对于内容消费者来说, 则是填补信息的一个途径。不幸的是, 我们只能在线文章的生命周期结束后才能获得总体人气。在早期阶段预测整体流行度对许多应用都有启示, 如推荐、广告和信息检索 (Gao 等人, 2018; Liu 等人, 2016)。此外, 它还具有学术价值和工业应用价值, 可以正确回答 "如何在任何时候



(a) 长期的增长趋势。



(b) 短期波动。

预测在线内容的整体流行度 "这样的问题?

最近, 人气预测引起了人们的极大关注。学者们用两个宽泛的类别来处理这项任务

*通讯作者: 金旭 <jinxxu@tencent.com>; 前两位作者贡献相同

图1：这两个子图显示了微信中文章受欢迎程度的时间动态。¹.(a) 所有文章的归一化总浏览量的累积过程。(b) 一篇文章的每小时浏览量的变化。

方法：时间过程建模和内容结构建模。时间过程模型是根据时间段内聚集的浏览量的时间演变过程来预测流行度的。累积的人气随着时间的推移而增加，同时在外部的影响下出现意外的爆发，如图1所示。大多数现有的工作都是基于外部影响的特殊假设来捕捉短期波动（Zhao 等人，2015；Cao 等人，2017；Rizoiu 等人，2017）。然而，许多外部因素是不可预测的。特定的假设限制了这些模型的预测能力。Yu 等
Copyright © 2019, Association for the Advancement of Artificial Engineers.
智力（www.aaai.org）。保留所有权利。

人试图根据手工制作的“阶段”，从时间过程本身提取人气波动（Yu 等人，2015）。然而，外部因素的影响可能涵盖不同的范围和持续时间。因此，我们很难假设波动的数量和形状。如何自动提取短期波动仍然是这个方法分支中尚未解决的问题。

另一方面，最近的工作证明了内容特征在流行度预测中的有效性，如短文描述、标题和图片（Zhang 等人，2018；Piotrkowicz 等人，2017；Sanjo 和Katsurai，2017）。然而，在线文章通常是长文本，对其进行建模并非易事，而且不同形式的元数据特征使内容特征的建模更加复杂。现有的工作都没有充分利用长文本和元数据特征来预测在线文章的受欢迎程度。

¹<http://www.wechat.com>

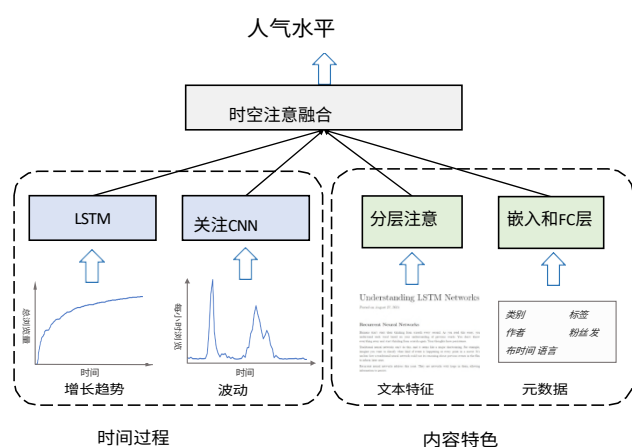


图2：节奏过程和内容特征模型的深度融合的整体框架。

同时，这两类方法在网络文章的不同生命周期阶段有各自的优势和劣势。时间过程模型依赖于一系列的历史事件，随着时间的推移，表现得越来越好，因为观察到的流行度越来越接近整体流行度。然而，在网上内容发布后的最初阶段，很难了解流行度的整体趋势。在实际应用中，在网络文章生命周期的早期阶段预测整体流行度是很有价值的，这样我们就可以推荐潜在的 "热门" 文章并过滤掉不受欢迎的文章。相比之下，内容特征不会随时间变化。因此，内容特征建模在早期阶段比较可靠，而它不能利用流行度的时间演变。因此，我们应该把时间过程和内容特征建模结合起来，以利用它们的再观察能力。然而，不同的文章显示出不同的增长速度和人气演变的波动。直观的融合方法，如矢量连接或线耳组合，在处理流行度演变过程的多样性方面缺乏灵活性。

在上述挑战的激励下，我们提出了一种神经网络工作方法，名为时间过程和内容特征的深度融合（DFTC）。DFTC的框架如图2所示。在我们的模型中，我们通过以下的建模技术来解决上述挑战：（1）对于时间过程的建模，我们采用了循环神经网络（RNN）来捕捉人气的长期增长趋势。对于短期波动，我们采用基于注意力的Convolutional Neural Network（CNN）来自动提取上升或下降的 "阶段" 结构。（2）对于内容特征的建模，我们利用层次注意网络（HAN）（Yang等人，2016）来捕捉文本特征，并采用嵌入技术将元数据特征嵌入到豪迈的密集向量。（3）对于动态融合，我们采用了一个时间注意力层。它利用注意力机制，根据上述所有建模技术的输出和时间背景，学习灵活的权

进行广泛的实验来评估不同阶段的预测效果，包括对CNN和注意力融合的影响的案例研究。

我们的主要贡献总结如下：

- 我们利用RNN的长期增长趋势和CNN的短期波动的时间过程自动，而不是具体假设外部因素或手工制作的 "阶段"。
- 我们采用HAN技术处理文本特征，采用嵌入技术处理元数据特征，采用时间注意力融合技术动态地整合时间过程和内容特征模型。
- 实验结果表明，所提出的模型明显优于最先进的方法，证明了我们方法的有效性和优越性。

重来组合这些技术。此外，我们从微信中收集了真实世界的数据集，并在此基础上对其进行了分析。

相关工作

几十年来，人气预测已经引起了人们的极大关注。学者们用两种广泛的方法来处理这项任务：时间过程建模和基于特征的建模。

时间过程建模。一些研究者将在线内容的流行度累积视为观点事件的一个微观的ar-rival点过程。他们通过对单个事件的微观点过程进行建模来预测流行性（Shen等人，2014），Hawkes点过程（Zhao等人，2015）或神经网络（Cao等人，2017；Gou等人，2018）。然而，在大规模应用中，事件的数量会在短时间内爆炸，这将导致微观节奏过程建模的性能问题。因此，我们认为基于事件量的宏观积累过程来预测流行度更有实际价值，因为宏观时间过程通常是几十个到几百个数据点。

人们提出了许多模型来描述宏观时间过程的演变。Hawkes intensity processes（Rizoiu等人，2017）为宏观时间过程扩展了hawkes point process，并在Youtube视频流行度预测中采用。HIP对时间过程的函数形式和外部因素的影响做了具体的假设，这可能限制了这些模型的表达能力（Du等人，2016）。Mishra等人提出了一个双RNN模型，用于建模微观和宏观的时间过程（Mishra, Rizoiu, and Xie 2018），并取得了最先进的性能。然而，它仍然需要微观事件来模拟由外部影响引起的波动。Yu等人从宏观时间过程中提取上升和下降的"阶段"来捕捉波动，并提出了一种基于阶段的线性回归方法来预测人气（Yu等人，2015）。然而，手工制作的"阶段"不能处理人气演变过程的多样性。在我们的模型中，我们采用注意力CNN来自动提取局部的上升和下降结构，采用LSTM来捕捉宏观时间过程的长期增长趋势。

基于特征的建模。其他一些研究者对在线内容的内容和元数据特征进行了修改，以达到以下目的

流行度预测。Piotrkowicz 等人仅使用标题特征预测了新闻文章的受欢迎程度 (Piotrkowicz 等人, 2017)。Sanjo 等人提出了一个用于在线食谱流行度预测的视觉-语义融合模型, 利用了食谱中的im-age和短文特征 (Sanjo和Katsurai 2017)。用户引导的分层注意力网络 (Zhang等人, 2018) 学习了模式内容和用户特征, 用于社会图像流行度预测。遗憾的是, 现有的工作中没有一个能充分利用长文本和元数据特征来预测在线文章的流行度。此外, 这些方法也忽略了在线内容的流行度演变过程。

其他一些方法提取了各种手工制作的特征, 用于预测流行度, 包括时间上的特征, 也有一些方法提取了一些其他的特征。

cess和内容特征。Keneshloo 等人提取了元数据、内容和时间特征, 并采用树状再回归法进行新闻流行度预测 (Keneshloo等人, 2016)。Shulman 等人增加了新的社会结构和早期采用者特征以提高预测性能 (Shulman, Sharma, and Cosley 2016)。他们的表现在很大程度上依赖于提取的特征。然而, 这些特征很难设计和测量, 而且往往与特定的数据集或应用绑定。受深度学习巨大成功的启发, 我们利用神经网络对内容特征和时间过程进行修改, 避免了费力的特征工程 (Xu等人, 2012; Xu等人, 2017)。

问题的提出

我们将人气预测任务视为一个分类问题, 将总浏览量离散为 n 个区间

$\{l_1, l_2, \dots, l_n\}$ 来代表在线文章的流行程度。我们的目标是在网上文章发表后的任何时候预测其受欢迎程度。为了便于计算, 我们将连续时间离散为时间段, 并将用户反馈事件量聚集为宏观时间序列。

这里, 用户反馈事件不仅包含 "查看", 而且在许多应用中还包含 "分享"、"评论" 或 "喜欢"。我们把这些事件在时间段 t 中的所有数量作为反馈向量 v_t 。更正式地说, 对于一个在线内容 c , 给定任何时间段 t 和历史反馈序列 $\{v_1, v_2, \dots, v_t\}$, 目标是预测 c 的整体流行水平。

模型

在这一节中, 我们介绍了所提出的时态过程和内容特征的深度融合 (DFTC) 模型。整体框架在图2中呈现。DFTC由三部分组成: 时间过程建模、内容特征建模和周到的融合。时间过程建模以历史反馈序列 $\{v_1, v_2, \dots, v_t\}$ 为输入, 并采用递归神经网络进行建模。

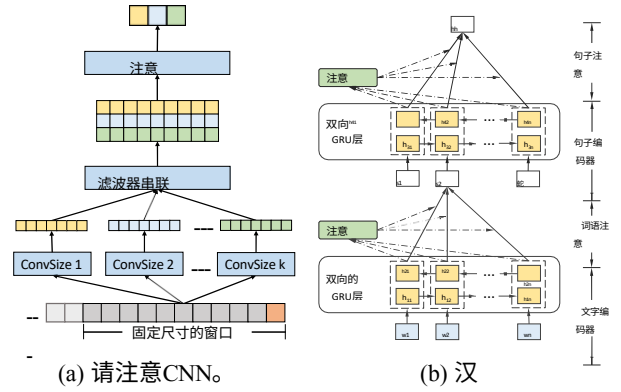


图3: (a) 用于捕捉短期波动的注意力CNN的结构。(b) 分层注意力网络(HAN)用于对文本内容特征进行建模。

时间过程建模

在这项工作中, 我们采用了循环神经网络 (RNN) 对流行的时间演变过程进行建模。长短时记忆 (LSTM) 是最广泛使用的RNN结构。我们重申了LSTM的表述:

$$i_t = \sigma(W x_{t-1}^i + U c_{t-1}^i + V h_{t-1}^i + b^i) \quad (1)$$

$$f_t = \sigma(W^f x_{t-1}^s + U^f c_{t-1} + V^f h_{t-1} + b^f) \quad (2)$$

$$c_t = f_t \odot h_{t-1} + i_t \odot \tanh(W x_{t-1}^c + V h_{t-1}^c + b^c) \quad (3)$$

$$o_t = \sigma(W x_{t-1}^o + U c_{t-1}^o + V h_{t-1}^o + b^o) \quad (4)$$

在内容特征建模中, 我们利用层次注意网络来学习文本特征, 并利用嵌入技术来提取元数据特征。在内容特征建模方面, 我们利用分层注意力网络工作来学习文本特征, 并利用嵌入技术来提取元数据特征。最后, 我们通过时空注意力融合动态地整合所有这些部分。

$$h_t = o_t \otimes \tanh(c)_t \quad (5)$$

RNN在时间建模方面的优越性在于，隐藏状态 h_t 涉及所有的历史信息，因此我们不需要对历史趋势的函数形式做出特定的假设（Du等人，2016）。更重要的是，LSTM中的记忆单元 c_t ，确保长期依赖性也能被捕捉到。因此，我们采用LSTM来学习人气的长期增长趋势。具体来说，我们将每个时隙的反馈向量 v 送入LSTM，并在输出向量 h 中获得历史增长模式 r 。

另一方面，网络内容的短期流行会受到外部事件的影响，呈现出未曾预料到的爆发（Zhao等人，2015；Rizoiu等人，2017；Cao等人，2017）。然而，要弄清所有的影响因素是非常困难的，因为其中许多因素是不可预测的。因此，我们假设从时间过程本身来捕捉短期的波动。考虑到一维时间轴上的短期人口数量曲线，由外部因素引起的波动使曲线由上升和下降阶段组成，看起来像"山"和"谷"，如图1b所示。这些"山"和"谷"是翻译不变的局部结构。因此，我们利用了1-D卷积神经网络，这已被证明是捕捉这种结构的最佳选择。此外，不同因素的影响在不同的时间跨度上持续，这意味着"山"有不同的宽度。受inception模块（Szegedy等人，2016）的启发，我们采用了多个

t

图3a中显示了不同大小的核，以捕捉不同规模的波动。然后，我们将所有卷积核的输出垂直堆叠。

请注意，CNN通常需要固定大小的输入。因此，我们在 t 之前取一个固定长度为 k 的剪切序列 $\{v_{t-k+1}, v_{t-k+2}, \dots, v_t\}$ 。然后我们应用同样的填充，得到一个输出系列 $\{c_{t-k+1}, c_{t-k+2}, \dots, c_t\}$ ，长度也为 k ，捕捉到最近历史的波动模式。最后，我们需要通过时间维度将输出序列合并为输出向量 h^c 。有几个广泛的合并操作者使用的方法，如矢量连接、最大/平均集合和线性组合。这里

我们采用注意力机制（Vaswani等人，2017）来合并 $\{c_{t-k+1}, c_{t-k+2}, \dots, c_t\}$ 。注意机制通过将不同的注意权重 α^c 乘以不同的向量 c ，帮助输出 h^c 集中于这种受外部因素影响的时间段。注意权重 α^c 和输出向量 h^c 的计算如下：

$$\alpha_i^c = V_i^c \tanh\left(\sum_{i=1}^c W_j c_{t-k+i}^c + b\right)^c \quad (6)$$

$$\alpha^f = \frac{\exp(a_i)}{\sum_{j=1}^k \exp(a_j)^c} \quad (7)$$

$$h^c = \sum_{i=1}^t \alpha^c c_{t-k+i}^c \quad (8)$$

内容特征建模

在线文章的内容特征，包括文本和元数据特征，在很大程度上决定了它们的流行。在线文章通常是长文本文件，如新闻文章和博客。受神经网络在自然语言过程中的巨大成功启发（Lai等人，2015；Yang等人，2016；Huo, Li和Zhou, 2016），我们采用了层次化注意力网络（HAN）（Yang等人，2016）来对文本内容特征进行建模。HAN的框架如图3b所示。考虑到文档固有的层次结构（即单词构成句子，句子构成一个文档），HAN将一个文档编码为一个具有

编码器和注意力的水平，应用于词级和句级。词级和句级的编码器都是双向的Gated Recurrent Unit。关于HAN的更多细节，请参考他们的研究文章（Yang等人，2016）。此外，标题是文章的高级概述

并显示主要印象。我们还学习了一个标题代表向量作为补充。由于标题通常是

只用词级编码器和注意力。然后，我们将文件向量和标题向量一起作为文本进行分类。

特征 h^h 。

元数据特征包括一次性特征，如

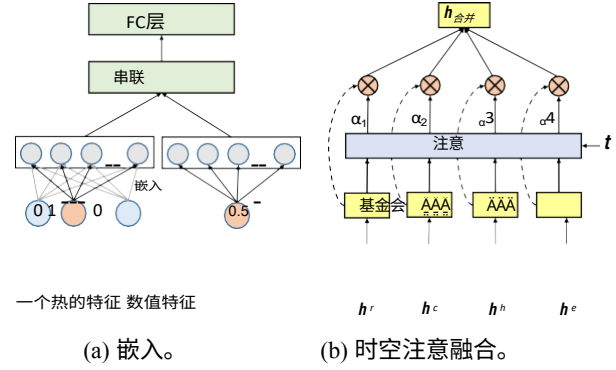


图4：（a）元数据特征的嵌入技术的说明。（b）时空注意层的结构。

如图4a所示，我们通过嵌入矩阵将单热特征嵌入到向量中。另一方面，我们将数字特征与嵌入向量相乘，提取视觉感知的源的密集向量。然后我们把

确定嵌入向量，并应用全连接层将所有元数据特征组合到 h^e 。

专心致志的融合

对于融合上述建模技术，一个直接的方法是将所有这些部分的输出串联起来，并将其送入

连接结果进入输出层进行预测。设 h^r, h^c, h^h, h^e 代表RNN、CNN、HAN和HAN的输出。元特征嵌入。然后，我们可以得到在 t 的预测结果为： $Y_t^c = f(W[h^r, h^c, h^h, h^e] + b)$ 。这样一来， h^r, h^c, h^h, h^e

与固定权重 w 相结合。如第1节所述，我们认为它在处理时间过程的动态演变方面缺乏灵活性。在网上文章发表后的最初阶段，时间过程模型很难了解到流行的整体增长趋势。因此，预测应该主要依赖于内容特征建模。随着时间的推移，观察到的流行度越来越接近整体流行度，所以时间过程模型应该在预测中占主要部分。在此基础上，我们假设将 h^r, h^c, h^h, h^e 与一个灵活的模型进行整合，twoable权重 α^m 的向量。 α 应该是 h^r, h^c, h^h, h^e 的一个函数。

和时间背景 t ，这样它就能自动适应不同的输出和时间背景。

在这项工作中，我们采用了一种注意力机制来实现动态整合，如图4b所示。注意机制是一个元素的组合，因此我们将 h^r, h^c, h^h, h^e 形成全连接层，用于特征组合和交流。需要元素对齐的向量 h^r, h^c, h^h, h^e 。那么我们

使用两层神经网络来计算周到的权重 α^m ，因为：

$$\alpha^m = V^m \tanh\left(\sum_{i=1}^i W_j h_t + w_t t + b\right) \quad (9)$$

$$i \quad i \quad W_j \quad h_t + w_t t + b \quad j \in \{r, c, h, e\}$$

作为类别，以及数字特征，如作者的粉丝数量。而不是手工制作的选择和组合
 这些特征，我们利用嵌入技术，以实现对这些特征的管理。
 将这些特征分解成同源的密集向量，并应用全连接层进行特征组合。

$$\alpha^m = \sum_{k \in \{r, c, h, e\}} \frac{\exp(a_i)^m}{\exp(a)^m} \quad (10)$$

时间背景*t*包括给定时间段*t*的周期性属性（即一天中的小时和星期），时间

t 的时间间隔和发布时间。这里，周期性特征是一个热点特征，时间间隔是一个数字特征。我们采用与嵌入元数据特征相同的策略，将时间背景嵌入到矢量中。有了周到的权重 α^m ，我们将所有的子网络动态地组合为 h^{merge}_t ，得到一个概率分布 $P_t = \{p_t(l_1), p_t(l_2), \dots, p_t(l_n)\}$ 的人气水平完全后连接层和软性输出层。然后，我们将具有最大概率的受欢迎程度作为预测结果 y_t 。

$$h^{merge}_t = \sum_{i \in \{r, c, h, e\}} \alpha_i \tilde{h}_i^t \quad (11)$$

$$P_t = \text{softmax}(f(h^{merge}_t)) \quad (12)$$

$$y_t = \arg \max_l p_t(l) \quad (13)$$

时间性衰减的损失

通过所提出的模型，我们得到在线文章 c 在时间段 t 的流行程度的概率分布 P_t 。假设 c 的真实程度为 l_c ， t 的单步损失由交叉熵定义为 $L_t = -\log p_t(l_c)$ 。通过为整个时间序列，我们可以定义整个损失的

我们的模型为 $J = \sum_t L_t$ 。

在实际应用中，在早期阶段预判整体流行度是更有价值的。此外，观察到的流行度和总体流行度之间的内在关系使我们更容易在后期进行预测。为了帮助我们的模型在早期阶段投入更多的精力来优化预测性能，我们对单步损失乘以一个时间衰减因子：

$$J = \sum_t D(\Delta t) L_t = - \sum_t D(\Delta t) \log p_t(l)_c \quad (14)$$

时间衰减因子 $D(\Delta t)$ 应该是一个单调的、非增加的时间间隔函数 Δt 在 t 和发布时间之间。在这项工作中，我们选择一个函数如下：

$$D(\Delta t) = \lceil_{\log \nu} (\Delta t + 1) \rceil^{-1} \quad (15)$$

这里， $\lceil \cdot \rceil$ 代表向上舍入运算。 $\Delta \Delta t$ 是指

从发布时间到 t 的时隙，因此 Δt 和 $\lceil_{\log \nu} (\Delta t + 1) \rceil^{-1}$ 都是正整数。 $\nu > 1$ 是一个控制衰减率的超参数。我们采用对数函数来保证 $D(\Delta t)$ 的衰减率随着时间的推移越来越小。为了限制对数函数的初始衰减率，我们采用了向上取整算子。

| | 热的 | 正常 | | 冷 | |
|---------|--------|--------|--------|-------|--------|
| | | >1,000 | ≤1,000 | >10 | ≤10 |
| 培训#文章 | 18,832 | 9,159 | 9,243 | 8,946 | 8,884 |
| 平衡测试#文章 | 2,093 | 1,020 | 989 | 1,009 | 1,007 |
| 随机测试# | 78 | 467 | 2,060 | 4,308 | 23,087 |

表1：数据集统计

来订阅文章更新。微信为用户提供文章推荐和搜索功能。在阅读文章时，用户还可以采取 "分享"、"保存"、"喜欢" 和 "提示" 行动。现在，微信中每天有超过50万篇新文章和20亿次 "浏览"。

我们将文章的总体受欢迎程度分为三个类别，"热门"（超过10,000次浏览），"冷门"（少于100次浏览）和 "正常"（其他）。在这里，我们把文章发表后15天内的浏览量作为整体流行度的一个近似值。文章浏览量的分布是一个典型的幂律分布。只有0.08%的文章是 "热" 的，而93%以上的文章是 "冷" 的。因此，我们收集了5月25日至7月25日的所有 "热门" 文章。

和其他两类文章的样本下。为了保证训练数据的多样性，我们对对数的浏览量采用了片状均匀抽样。我们计算每篇文章每5分钟的 "查看"、"分享"、"保存"、"喜欢" 和 "提示" 的动作量作为宏观时间序列。然后，我们在观察到的受欢迎程度达到整体受欢迎程度的80%或80%的 "热门" 阈值之前剪辑这些时间序列。时间序列长度小于12的文章被过滤掉，因为它们要么立即成为 "热门"。偶尔或一直保留 "冷"。最后，我们得到61,178。在下面的实验中，我们将85%的文章用于训练我们的模型，5%的文章用于评估，10%的文章用于评价。在接下来的实验中，我们取85%的文章用于训练我们的模型，5%的文章用于评估，10%的文章用于评价，称为 *平衡测试集*。为了评估我们的模型在现实中的分布情况，我们从7月26日至8月10日随机抽取了其他30,000篇文章作为 *随机测试集*。数据集的统计信息如表1所示。此外，元数据的特征包括

实验

数据集

我们从一个广泛使用的移动社交应用微信中收集了一个

在线文章数据集。²媒体机构和个人用户都可以建立自己的官方账户来发布新闻和文章。用户可以关注官方账号

²<http://www.wechat.com>

"类别"、"发布时间"、"内容长度"、"视频数量"。ber "和"出版商的粉丝数"。

实验设置

基线 我们将提议的DFTC方法与以下基线进行比较：

- **基于特征的分类器**。我们采用逻辑回归（LR）和随机森林（RF）作为流行分类任务的基线。这些分类器把时间过程特征和内容特征都作为输入。
- **HIP（Rizoiu等人，2017）**。霍克斯强度过程以众所周知的霍克斯点过程为基础，对宏观时间过程进行建模，并被应用于预测视频的流行度。
- **VoRNN-TS（Mishra, Rizoiu, and Xie 2018）**。栏目

| 方法 | 平衡测试组的结果 | | | | 随机测试集的结果 | | | |
|---------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | 准确度 | 热门 | 正常F1 | 寒冷的F1 | 准确度 | 热门 | 正常F1 | 寒冷的F1 |
| LR | 0.6441 | 0.3575 | 0.6446 | 0.7088 | 0.7551 | 0.4248 | 0.8272 | 0.8973 |
| 射频 | 0.6587 | 0.4246 | 0.6506 | 0.7277 | 0.8086 | 0.4743 | 0.8454 | 0.8909 |
| HIP | 0.6502 | 0.4353 | 0.6330 | 0.7182 | 0.7860 | 0.4342 | 0.7742 | 0.9217 |
| 呼叫中心- | 0.6709 | 0.4447 | 0.6530 | 0.7366 | 0.8569 | 0.4581 | 0.8505 | 0.9540 |
| CACNN | 0.6965 | 0.4018 | 0.7040 | 0.7394 | 0.8498 | 0.4825 | 0.8472 | 0.9493 |
| DFTC-TS | 0.7278 | 0.4858 | 0.7203 | 0.7638 | 0.8863 | 0.5253 | 0.8592 | 0.9698 |
| DFTC-SF | 0.6542 | 0.5343 | 0.6754 | 0.6212 | 0.6879 | 0.5536 | 0.6926 | 0.7869 |
| DFTC-SM | 0.7559 | 0.5554 | 0.7489 | 0.7812 | 0.9301 | 0.5649 | 0.8625 | 0.9759 |
| DFTC | 0.8147 | 0.6110 | 0.7822 | 0.8393 | 0.9653 | 0.6292 | 0.8729 | 0.9916 |

表2：总体预测性能

RNN利用LSTM的优势，在宏观时间过程中实现了最先进的性能。

- **CACNN（Gao 等人，2018）**。Context Attention Convolutional Neural Network被提出来进行点击率预测，它用attention CNN对时间过程进行建模，并纳入元数据特征。
- **DFTC-TS & DFTC-SF**。DFTC-TS是我们模型的时间过程建模部分。DFTC-SF是我们模型的内容特征建模部分。
- **DFTC-SM**。DFTC-SM通过向量连接将时间过程和content特征建模合并起来。

衡量标准 我们采用准确率和F1分数作为分类性能的衡量标准。准确率是指准确预测与所有预测的比率。在我们的实验中，我们研究了流行程度的F1分数，分别把它们作为正数和其他两个级别作为负数。

参数设置 在时间过程建模中，我们采用隐藏大小为512的单LSTM层。我们采用了4种内核，其大小分别为1、3、7、11。每种核的数量被设定为128。CNN的输入窗口大小 k 为3小时。在内容特征建模中，我们从相关分类任务的预训练的HAN开始，用我们的数据集微调参数。元数据特征的嵌入大小被设定为32。我们采用2个FC层进行元数据嵌入组合，1个FC层用于对齐 h^r, h^c, h^h, h^e ，以及在关注层之后的1个FC层。

所有的FC层都采用ReLU作为主动函数，并有统一的隐藏大小512。在衰减损失函数中，我们设定 γ 为12。最后，我们利用Xavier初始化和Adam优化器进行参数学习，并在每个FC层和RNN层采用dropout进行正则化。

复杂度分析 对于我们的神经网络，主要的空间消耗是权重矩阵的存储，主要的时间成本是线性变换的跳动。让 n 代表网络中的最大隐藏大小，我们模型的空间和

这一点不需要大加考虑。具体来说，在我们的实验设置中，参数的总数为

6.2M。在在线预测中，我们模型的HAN和嵌入部分对每篇具体文章只计算一次。我们模型的LSTM、注意力CNN和注意力融合部分将在每个时间步骤中执行。为了比较所提出的模型与LSTM和注意力CNN的时间成本，我们在Tesla P40 GPU上进行了基于TensorFlow的实验。在批次大小为64的情况下，向前一步预测，CNN消耗0.29ms，RNN消耗0.49ms，我们提出的DFTC消耗0.97ms。

结果分析

与基线的比较 表2显示了用准确性和F1得分衡量的预测性能比较。我们可以看到，所提出的方法在两个测试集上的所有指标都达到了最佳结果。请注意，所有的方法在随机测试集上都得到了更好的结果。在现实的分布下，大多数文章的浏览量与其他类别的文章的浏览量相差甚远。大多数“冷门”文章的浏览量低于10，大多数“冷门”文章的浏览量低于1000，如表所示

1.在平衡的测试集中，有更多的例子接近分类边界，这对性能评估更有说服力。

具体来说，基于特征的分类器的性能在很大程度上取决于提取的特征，而这些特征很难设计和测量，特别是对于时间序列任务而言。鹰的强度时间复杂度都是 $O(n^2)$ 。在其他一些神经网络中，如用于CV任务的2D-CNN，中间的结果也消耗了大量的空间。然而，中间的结果我们模型的中介结果是最大尺寸为 n 的一维向量、

由于流行度动态的特定假设限制了表达能力，因此RNN流程只能达到与基于特征的分类器类似的性能。卷式RNN用LSTM而不是经验函数形式来模拟人气增长趋势。它优于基于特征的分类器和HIP，而在不考虑文本和元数据特征的情况下，它仍然缺乏预测能力。CACNN在平衡测试集中的表现比其他最先进的基线好一点，它整合了时间过程和元数据特征建模。然而，CNN的输入只是时间过程的一部分，它不能捕捉到人气的长期增长趋势。与他们相比，我们提出的DFTC

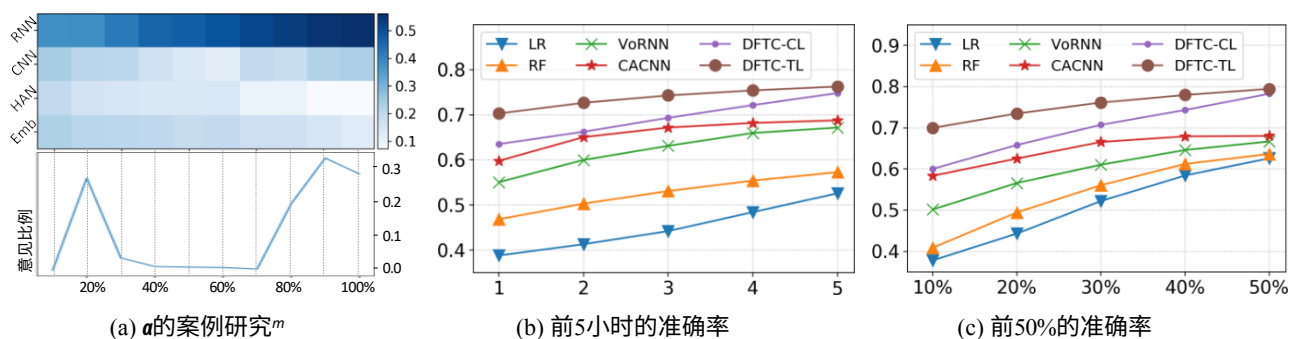


图5：性能分析。(a) 平均殷勤权重 α 的热图^m和每10%的时间序列的短期波动的线图。RNN的权重随时间增加，CNN的权重与波动相对应。(b) 文章发布后前5小时的平均准确率。(c) 当观察到的流行度达到整体流行度的10%-50%时的平均准确度。

该模型捕捉了时间过程的长期趋势和短期波动，并动态地整合了文本和元数据特征。基于这些建模技术，DFTC方法明显优于最先进的方法。在平衡测试集中，它显示了17.0%的准确率和37.4%、12.0%、13.5%的热、正常和冷F1分数的改善。我们的模型目前已被应用于微信的文章推荐中。

消融分析 建议的模型也优于其变体，即DFTC-TS、DFTC-SF和DFTC-SM。DFTC-TS和DFTC-SF只利用了时间过程或内容特征进行预测。他们的表现不理想也就不足为奇了。请注意，DFTC-TS的表现优于其他时间过程方法，即VoRNN-TS和HIP，说明LSTM对长期趋势的设计和CNN对短期波动的设计是有效的，也是最好的。DFTC-SM通过向量连接将时间过程和内容特征建模结合起来，这对于处理时间序列的动态增长缺乏灵活性。相比之下，DFTC方法采用了注意力机制来学习动态融合的灵活权重。因此，在平衡测试集上，它的准确率比DFTC-SM高出7.8%。

案例研究 为了研究注意力CNN和时间性注意力融合的效果，我们随机选择了一个“热门”艺术作品作为案例研究。我们对其关注权重 α^m ，并对其每10%的时间序列的浏览量进行汇总，如图5a所示。上方热图的四行分别代表RNN、CNN、HAN和嵌入的关注权重 α^m 。权重越大，颜色越深。可以看出，所有这些部分在开始时对预测的贡献是相似的。随着观察到的人气越来越接近整体人气，RNN可以更准确地学习长期的增长趋势。因此，RNN模型在后期的预测中发挥了重要作用。下面的线图显示了这篇文章的短期波动。我们可以观察到，浏览量在时间序列的开始和结束时都有爆发。相应地，关注度CNN在开始和结束时有很

高的关注度权重，这意味着它有效地捕捉到了这种局部的上升和下降结构，这对整个人气预测很重要。

早期阶段的表现 在实际应用中，预测早期阶段的整体人气是非常有价值的。图5b显示了文章发布后前5小时的平均准确率，图5c显示了当观察到的流行度达到平衡测试集中整体流行度的10%-50%时的平均准确率。我们可以看到，从时间和观察到的流行度两方面来看，我们的模型对早期阶段的预测性能都有明显的改善。请注意，这里没有显示HIP的性能，因为它需要一个足够长的时间过程来很好地估计参数（Rizoiu等人，2017）。基于特征的方法在早期阶段表现很差，因为手工制作的特征很难捕捉到时间上的动态特征。CACNN整合了元数据特征进行预测，因此它在早期阶段比VoRNN-TS表现更好。

DFTC-CL是我们的模型，具有共同的序列损失。内容特征建模使我们的方法能够在时间过程缺乏足够信息的情况下做出重新负责的预测，而贴心的融合确保内容特征建模在早期阶段发挥主要作用。因此，DFTC-CL明显优于最先进的方法。损失函数的时间衰减因子有助于我们的模型投入更多的精力来优化早期阶段的预测性能。当我们应用时间衰减损失函数进行优化时，即DFTC-TL，早期阶段的性能得到进一步改善。此外，由于观察到的流行度随着时间的推移越来越接近整体流行度，即使我们降低它们的权重，它也能在后期获得一个理想的性能。

总结

在这项工作中，我们提出了一种新的在线文章流行度预测方法。我们采用RNN来捕捉长期趋势，CNN来提取短期波动。我们利用HAN来模拟文本特征，并采用嵌入技术来学习元数据特征。最后，一个时间关注融合层被用来动态地整合所有这些部分。基于真实世界的在线文章数据集的评估结果证明了所提模型的有效性和优越性。

鸣谢

这项工作是廖东亮在腾讯公司的微信公司实习时完成的。我们要感谢我们的微信同事Shen Huang在实验和本文写作中的巨大帮助。我们还要感谢我们的微信同事冯喆、邓月堂、王志平和白延东对这项工作的有益讨论和支持。我们还要感谢西北工业大学的聂飞平教授和科内尔大学的陈玉东教授提出的宝贵建议。

参考文献

- Cao, Q.; Shen, H.; Cen, K.; Ouyang, W.; and Cheng, X. 2017. Deephawkes: 弥合信息级联的预测和理解之间的差距。In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1149-1158.ACM.
- Du, N.; Dai, H.; Trivedi, R.; Upadhyay, U.; Gomez-Rodriguez, M.; and Song, L. 2016. 循环标记的临时点过程：将事件历史嵌入到矢量。In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1555- 1564.ACM.
- Gao, H.; Kong, D.; Lu, M.; Bai, X.; and Yang, J. 2018. 用于广告商层面点击率预测的注意卷积神经网络。1855-1864.
- Gou, C.; Shen, H.; Du, P.; Wu, D.; Liu, Y.; and Cheng, X. 2018. 学习序列特征用于级联爆发预判。 *Knowledge and Information Systems* 1-19.
- Huo, X.; Li, M.; and Zhou, Z.-H.2016.从自然语言和编程语言中学习统一的特征来定位有缺陷的源代码。In *IJCAI*, 1606-1612.
- Keneshloo, Y.; Wang, S.; Han, E.-H.; and Ramakrishnan, N. 2016. 预测新闻文章的受欢迎程度。In *Proceedings of the 2016 SIAM International Conference on Data Mining*, 441-449.SIAM.
- Lai, S.; Xu, L.; Liu, K.; and Zhao, J. 2015. 用于文本分类的递归卷积神经网络。在 *AAAI* , 第333卷, 2267-2273。
- Liu, A.-A.; Nie, W.-Z.; Gao, Y.; and Su, Y.-T. 2016. 基于视图的三维模型检索的多模式剪贴图匹配. *IEEE Transactions on Image Processing* 25 (5) : 2103-2116.
- Mishra, S.; Rizoïu, M.-A.; and Xie, L. 2018.用递归神经网络对异步社交媒体流中的流行性进行建模。 *arXiv预印本* *arXiv:1804.02101*。
- Piotrkowicz, A.; Dimitrova, V.; Otterbacher, J.; and Markert, K. 2017. 头条很重要：使用标题来预测twitter和facebook上的新闻文章的受欢迎程度。在 *ICWSM* , 656-659。

Rizoïu, M.-A.; Xie, L.; Sanner, S.; Cebrian, M.; Yu, H.; and Van Hentenryck, P. 2017. 期待时髦：社交媒体受欢迎程度的Hawkes intensity过程。In *Proceedings of the 26th International Conference on World Wide Web*, 735-744.国际万维网会议指导委员会。

Sanjo, S., and Katsurai, M. 2017.采用深度视觉语义融合的食谱流行度预测。In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* , 2279-2282.ACM.

Shen, H.-W.; Wang, D.; Song, C.; and Barabási, A.-L. 2014.通过强化的 poisson过程来建模和预测流行动态。在 *AAAI* , 第14卷, 291-297。

Shulman, B.; Sharma, A.; and Cosley, D. 2016 。Predictability of popularity: 预测和理解之间的差距。In *ICWSM*, 348-357.

Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. 重新思考计算机视觉的初始架构。In *Proceedings of the IEEE conference on computer vision and pattern recognition* , 2818-2826.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017。Attention is all you need.In *Advances in Neural Information Processing Systems* , 5998-6008.

Xu, J.; Yin, Y.; Man, H.; and He, H. 2012.基于稀疏归因的特征选择。In *Neural Networks (IJCNN), The 2012 International Joint Conference on* , 1-7.IEEE.

Xu, J.; Tang, B.; He, H.; and Man, H. 2017。基于相关性和冗余标准的半监督性特征选择。*IEEE关于神经网络和学习系统的交易* 28(9):1974-1984.

Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; and Hovy, E. 2016.用于文档分类的层次注意网络。在 *计算语言学协会北美分会的2016年会议上: Human Language Technologies*, 1480- 1489.

Yu, H.; Xie, L.; Sanner, S.; et al. 2015。一个YouTube视频的生命周期: 阶段、内容和流行。在 *ICWSM* , 533-542。

Zhang, W.; Wang, W.; Wang, J.; and Zha, H. 2018。用于多模态社会图像流行度预测的用户引导的分层注意力网络。In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, 1277- 1286.国际万维网会议指导委员会。

Zhao, Q.; Erdogdu, M. A.; He, H. Y.; Rajaraman, A.; and Leskovec, J. 2015。Seismic: 用于预测推文流行度的自激点过程模型。在 *第21届ACM SIGKDD国际知识发现和数据挖掘会议上*, 1513-1522。ACM.