文章编号：1003-0077(2021)02-0001-18

基于深度学习的流行度预测研究综述

曹 婧[1,2]，沈华伟[1,2]，高金华[1]，程学旗[1]

（1. 中国科学院计算技术研究所 网络数据科学与技术重点实验室，北京 100190；
2. 中国科学院大学，北京 101408）

摘 要：在线社交网络中的消息流行度预测研究，对推荐、广告、检索等应用场景都具有非常重要的作用。近年来，深度学习的蓬勃发展和消息传播数据的积累，为基于深度学习的流行度预测研究提供了坚实的发展基础。现有的流行度预测研究综述，主要是围绕传统的流行度预测方法展开的，而基于深度学习的流行度预测方法目前仍未得到系统性地归纳和梳理，不利于流行度预测领域的持续发展。鉴于此，该文重点论述和分析现有的基于深度学习的流行度预测相关研究，对近年来基于深度学习的流行度预测研究进行了归纳梳理，将其分为基于深度表示和基于深度融合的流行度预测方法，并对该研究方向的发展现状和未来趋势进行了分析展望。

关键词：流行度预测；深度学习；信息传播；综述

中图分类号：TP391 **文献标识码**：A

Survey on Deep Learning Based Popularity Prediction

CAO Qi[1,2], SHEN Huawei[1,2], GAO Jinhua[1], CHENG Xueqi[1]

(1.CAS Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 101408, China)

Abstract: Popularity prediction over online social networks plays an important role in various applications, e.g., recommendation, advertising, and information retrieval. Recently, the rapid development of deep learning and the availability of information diffusion data provide a solid foundation for deep learning based popularity prediction research. Existing surveys of popularity prediction mainly focus on traditional popularity prediction methods. To systematically summarize the deep learning based popularity prediction methods, this paper reviews existing popularity prediction methods based on deep learning, categorizes the recent deep learning based popularity prediction research into deep representation based and deep fusion based methods, and discusses the future researches.

Keywords: popularity prediction; deep learning; information diffusion; survey

0 引言

随着互联网的繁荣发展，由用户原创内容（user generated content，UGC）主导的各种社交网站也随之在互联网中兴起，包括国外的脸书（Facebook）、推特（Twitter）、照片墙（Instagram），以及国内的新浪微博、微信朋友圈等。在这些社交网站上，用户之间组成的在线社交网络（online social networks，OSN），为信息在网络空间的传播带来了前所未有的便利。每天有数千万的原创内容在这些平台上被产生和传播[1]。如此纷杂巨量的信息，不仅为平台的管理带来了巨大的挑战，也容易对人们造成信息过载的困扰。在线社交网路中的消息流行度预测研究，旨在提前从海量的信息中预测并识别出未来的热门消息，从而为平台的质量把控提供辅助，并帮助人们从信息过载的困境中解脱出来。但由于社交网络的开放性以及社交网络中级联传播效应所带来的不确定性，准确地预测消息未来的流行度是一项极具困难和挑战的任务。

收稿日期：2019-11-06 **定稿日期**：2020-01-12

基金项目：国家自然科学基金（91746301，61472400，61425016，62002347）

传统的消息流行度预测方法主要分为两类，分别是基于特征提取的方法[2-4]和基于点过程建模的方法[5-7]。前者通常会人工提取内容、用户、时序、结构等方面的特征，然后根据这些提取的特征，结合传统的机器学习方法来对消息未来的流行度进行回归或分类预测。这类方法的性能依赖于人工提取特征的质量，而这些人工特征通常又是启发式提取的，很难全面且有效地捕获各类有效预测因素。而后者，基于点过程建模的方法，将消息传播过程看成是用户转发行为的一个到达点过程。其核心在于根据特定的假设，对点过程的速率函数进行不同的建模。但在实际场景中，我们很难知道速率函数的真正假设或形式，从而限制了点过程模型的能力。

随着近年来深度学习在文本、语音、图像等领域的成功应用，基于深度学习的流行度预测方法也开始逐渐被研究者们关注并提出[1,8-9]。借助于深度学习强大的表示能力，研究者们对内容、用户、时序、结构等因素进行了有效的建模表示[1,8]。此外，通过深度融合技术，实现了多因素以及多模型的融合[9]，大大提升了模型对消息未来流行度的预测性能。

为了更好地帮助流行度预测领域的相关研究人员，也有研究者梳理了消息流行度预测研究的相关工作，并形成了综述[10-12]。但已有的这些综述，主要都是围绕传统的流行度预测方法展开的，而未对基于深度学习的流行度预测方法进行系统性地归纳和梳理。因此，与已有综述不同，本文将重点论述和分析基于深度学习的现有流行度预测研究，并梳理出这些研究之间的相互关系。

本文组织结构如下：第1节介绍在线社交网络中流行度预测研究的相关背景，第2节和第3节分别梳理基于深度表示的流行度预测研究和基于深度融合的流行度预测研究。第4节从评价数据集、评价指标出发，总结了现有方法的评价体系。第5节对全文进行了总结。

1 问题定义与相关背景

本节对在线社交网络中的消息流行度预测这一研究问题所涉及的相关背景进行简要介绍，包括流行度预测问题的基本定义、问题与挑战，以及传统的流行度预测方法。

1.1 流行度预测问题定义

在线社交网络中的消息流行度预测研究，根据预测时机、预测任务的不同，可以定义为不同的问题并分别适用于不同的应用场景和需求。本节从上述这两个角度出发，总结现有流行度预测研究的问题定义。

1.1.1 预测时机

消息流行度预测问题，其主要任务是根据消息的内容或早期传播观测，对其未来的流行度进行有效预测。根据预测时机的不同，我们可以将在线社交网络中的消息流行度预测分为事前预测和早期预测这两类[13]。

1）事前预测(ex-ante prediction)

事前预测一般指在消息发布之前，根据消息本身的内容、发布用户等可获得的信息进行未来的流行度预测[8-9]。这类方法通常能为消息的传播提供诸多解释，即什么样的消息或者具备哪些特征的消息在未来能有什么样的流行度。但由于这类方法只利用了消息发布前可获得的信息来进行预测，一般在预测性能上表现欠佳。

2）早期预测(early prediction)

为了更好地预测消息在未来的流行度，一类更流行的做法是采用“窥视”策略(peeking strategy)，即在事前预测的基础上，结合消息在一段时间内的传播观测再进行预测，也称为早期预测[1,14]。早期预测相较于事前预测而言，通常在预测性能上有较大幅度的提升，但指导意义不如事前预测。

虽然事前预测和早期预测面向的应用场景和需求有所不同，但其可利用的预测信息或因素具有一定的交集和通用性。因此，本文将这两类流行度预测问题放在一起讨论。

1.1.2 预测任务

在线社交网络中的流行度预测，根据不同场景的目标，主要可以分为分类和回归两类预测任务。分类任务的目标是预测消息在未来的流行度是否会超过某个阈值或是否会翻倍[2-3,15-16]，而回归任务的目标是预测消息在未来某个时刻或最终的流行度这一具体数值[1,5,17-21]。通常认为，回归任务比分类任务更难，也是流行度预测问题中更常采用的一种预测目标。

1.2 问题与挑战

在线社交网络中的流行度预测是一个极具挑战

的问题。其面临的难点主要来自以下几个方面：

（1）**级联传播的不确定性**[13,17]。社交平台上的消息通过用户之间的关注网络或好友网络进行传播，具有级联传播效应（图 1）。也就是说，任何一个中间用户的参与都将改变消息在未来的传播范围，造成了消息未来流行度极大的不确定性。如何刻画消息依赖于网络的这种复杂传播，并减少预测的不确定性，是消息流行度预测研究中的一个关键。

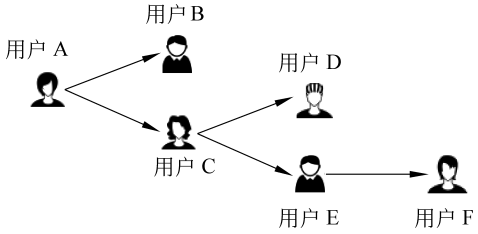


图 1　级联传播示意图

注：边的方向表示消息传播的方向

（2）**消息传播平台的开放性**。由于社交媒体平台的开放性，各种类型的因素都会对消息的流行度造成影响[2-3]。这些因素包括消息本身的内容、发布时间、发布用户等内源性因素，以及线下交流、新闻媒体报道等很难被观测到的外源性因素。如何捕捉并刻画多种因素对消息未来流行度的影响，也是流行度预测研究亟须解决的一个问题。

1.3　传统流行度预测方法

传统的流行度预测方法主要分两类，分别为基于特征提取的流行度预测方法和基于点过程建模的流行度预测方法。

1.3.1　基于特征提取的流行度预测方法

基于特征提取的消息流行度预测方法，通常为特定的平台或数据集，如 Arxiv[22-23]、新浪微博[24-25]、推特[4,15,26]、Digg[20,27]、YouTube[14,20] 等，人工设计各种各样的特征。这些特征主要包括用户特征[4,26,28]、内容特征[28-30]、时序特征[14,20,31-32]和结构特征[2,19,33-34]。其中用户特征包括用户的粉丝数[28]、活跃程度[4]、历史影响力[26]等；内容特征包括文本长度[28]、内容新颖性[29]、情感极性[30]等；时序特征包括观测窗口内的流行度总量[20]、观测窗口内的单位时间流行度[14]等；结构特征包括传播图的连边密度[2]、传播树的深度和广度[19]等。这四类特征在用于消息未来流行度预测的时候，时序特征被发现占据着主导性的作用，而内容特征的预测作用相对最小[3]。值得注意的是，这类基于特征提取的流行度预测方法既可用于事前预测问题，也可用于早期预测问题，其区别主要在于利用的特征不同。事前预测主要利用内容和用户特征[29-30]，而早期预测这四类特征均可利用[2-3]。

基于特征提取的流行度预测方法发现并验证了包括用户、内容、时序及结构等因素的预测有效性，能够为我们预测消息在未来的流行度提供比较初步的认识和理解。但是，这类方法所涉及的特征通常是通过启发式方法提取的，而模型最终的预测性能又非常依赖于这些启发式特征的质量。如何能够自动学习并表示上述这四类因素，是未来的发展趋势和方向，也是基于深度学习的流行度预测方法得以发挥长处的地方。

1.3.2　基于点过程建模的流行度预测方法

基于点过程建模的流行度预测方法，主要是将消息传播过程看成是用户转发行为的一个到达过程。这类方法的核心是对到达过程中的速率函数进行建模[5-6,35-38]，最终的流行度通过对该速率函数进行事件模拟（thinning algorithm）或期望计算得到。根据对速率函数的假设不同，点过程建模方式也有所不同。代表性的点过程建模包括：自增强泊松过程（reinforced Poisson process），在建模速率函数时考虑了消息自身的吸引力、富者愈富机制以及时间衰减效应[5,35]；自激励点过程（self-exciting process）则认为当前速率函数是历史所有转发的累积效应，即历史的每一次转发都对当前速率函数有一次新的激励[6-7,36-37]。

基于点过程建模的流行度预测方法，为建模消息如何获得关注、如何扩散影响提供了一个很好的通用框架。但是，目前的这类方法通常在建模速率函数的时候依赖于某些特定的假设，而在真实情况下，我们并不知道这些假设是否真的成立[39]。这在一定程度上限制了模型的表达能力。如何能从大量的数据中学习并得到一个更灵活自由的点过程建模方法，也是基于深度学习的流行度预测方法希望解决的问题之一。

2　基于深度表示的流行度预测方法

已有的研究证明，消息内容、发布用户、传播时序以及传播结构等因素对预测消息在未来的流行度具有非常重要的指示作用[2-3]。但这些因素本身往往都非常复杂，很难用简单的启发式方法来进行很好的表示。而基于深度表示的流行度预测方法，旨

在利用深度表示学习的技术，对内容、用户、时序、结构等因素进行更好的表示，从而更准确地预测消息未来的流行度。现有基于深度表示的流行度预测方法总结见表 1。

表 1 基于深度学习的流行度预测方法总结

方法	方法名	内容	用户	时序	结构	因素融合	模型融合
Du 等人[39] KDD2016	RMTPP	—	√	√	—	—	—
Li 等人[21] WWW2017	DeepCas		√		√	—	—
Xiao 等人[40] AAAI2017	—	—	√	√	—	—	—
Wang 等人[41] ICDM2017	Topo-LSTM	—	√	—	√	—	—
Wu 等人[42] IJCAI2017	DTCN	—	√	√	—	—	—
Cao 等人[1] CIKM2017	DeepHawkes	—	√	√	√	—	—
Sanjo 等人[8] CIKM2017	—	√	—	—	—	—	—
Xiao 等人[43] NIPS2017	WGANTPP	—	—	√	—	—	—
Chen 等人[44] ISI2017	NPP	√	√	√	—	√	—
Xiao 等人[45] AAAI2018	CWE	—	—	√	—	—	—
Zhang 等人[9] WWW2018	UHAN	√	√	—	—	√	—
Mishra 等人[46] ICWSM2018	RNN-MAS	—	√	√	—	—	—
Wang 等人[47] IJCAI2018	UMAN	√	√	—	—	√	—
Yan 等人[48] IJCAI2018	—	—	—	√	—	—	—
Dou 等人[49] KDD2018	KB-PPN	√	—	√	—	√	—
Wu 等人[50] CIKM2018	PreNets	√	—	√	—	—	√
Li 等人[51] NIPS2018	RLPP	—	—	√	—	—	—
Upadhyay 等人[52] NIPS2018	TPPRL	—	√	√	—	—	—
Xiao 等人[53] IEEE Access2018	ANN	√	—	√	—	—	—
Liao 等人[54] AAAI2019	DFTC	√	—	√	—	√	—
Shao 等人[55] CCIR2019	TCN	—	—	√	—	—	—
Zhao 等人[56] PAKDD2019	KB-PPN	√	—	√	—	√	—
Yang 等人[57] IJCAI2019	FOREST	—	√	√	—	—	—
Chen 等人[58] SIGIR2019	DMT-LIC	—	√	√	√	—	—
Chen 等人[23] ICDE2019	CasCN	—	√	√	√	—	—
Chen 等人[59] Neurocomputing2019	NPP	√	√	√	—	√	—
Cao 等人[60] WSDM2020	CoupledGNN	—	√	—	√	—	—

2.1 基于深度学习的内容表示

在线社交网络中的用户原创内容具有多种不同的模态(modality)，包括文本[8-9,44,47,54,59]、图像[8-9]、商品(item)[49,56]等。这些不同的模态通常包含着非常复杂的语义，因此需要采用不同的深度表示方式来进行内容语义的提取。相关文献分类整理见表 2。

表 2 基于深度学习的内容表示

分类		文献	深度学习模型
文本表示	无序文本	Sanjo 等人[8]	词向量＋池化
	有序短文本	Zhang 等人[9]，Wang 等人[47]	词向量＋循环神经网络或卷积神经网络
	有序长文本	Liao 等人[54]	词向量＋层级注意力
图像表示		Sanjo 等人[8]，Zhang 等人[9]	卷积神经网络

续表

分类	文献	深度学习模型
商品表示	Dou 等人[49]，Zhao 等人[56]	实体向量表示
多模态混合表示	Zhang 等人[9]	注意力机制

2.1.1　内容模态单独表示

对于最常见的文本模态的深度表示，研究者广泛借鉴了自然语言处理中的常用方法，即对文本中的每个词用一个 K 维的向量来表示，使得语义相近的词在向量空间中处于相近的位置，该向量也被称为词向量(word embedding)[61]。在词向量的基础上，需要针对不同的文本类型提出不同的词向量聚合方式。

具体来说，对于标签等无序文本类型，设无序文本的词向量集合为 $W=\{w_1,w_2,\cdots,w_n\}$。由于不同消息中词向量集合的大小 n 不尽相同，Sanjo 等人[8]提出用池化(pooling)的方式来进行词向量的聚合，即文本表示 $c=f(W)$。池化函数 f 可以选择常用的均值函数或最大值函数等。这样的池化聚合方式，不仅能够捕获不同大小文本的特征，还能在捕获过程中忽略词的先后顺序，有效适用于无序文本的场景，如图 2 所示。

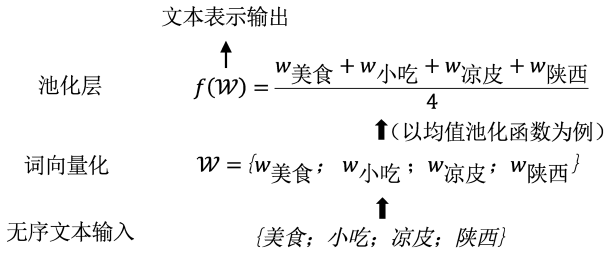


图 2　Sanjo 等人[8]采用的无序文本表示模型

遗憾的是，这样的池化模型很难捕捉序列化文本中的序列特性。因此，对于微博、推特消息等序列短文本，Zhang 等人[9]采用了循环神经网络(recurrent neural network，RNN)来刻画微博文本的序列特性，即将词向量按照文本序列中的出现顺序逐个输入到循环神经网络中，最后一个词的隐向量可以作为该序列的语义表示。而 Wang 等人[47]采用了一维卷积神经网络(convolutional neural network，CNN)来捕捉短文本序列中的局部序列特性，即通过卷积算子提取特定卷积核大小的输入短语特征，并通过池化操作得到固定长度的全局文本语义表示。这两类模型框架如图 3 所示。

面对新闻等长文本时，单一地使用循环神经网络或一维卷积神经网络无法有效地反映长文本的结

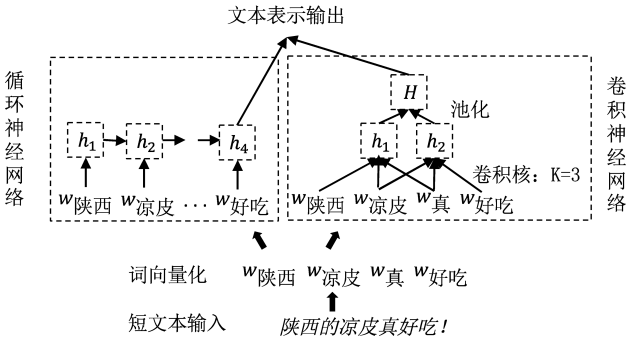


图 3　基于深度学习的短文本表示模型[9,47]

构特性。因此，Liao 等人[54]采用了自然语言处理中常用的层级注意力模型(hierarchical attention networks，HAN)[62]，引入文章结构的先验知识，并利用注意力机制对句子中不同词的权重以及长文本中不同句子的权重进行学习。具体来说，注意力机制通常被定义为一个查询到一组键值对的映射过程[63]，如式(1)所示。

$$c=\sum_{i=1}^{n}\mathrm{softmax}[\mathrm{score}(h_q,h_i)]v_i \tag{1}$$

其中 c 是经注意力机制后得到的文本表示，h_q 表示一个查询，h_i、v_i 表示一组键、值。score(h_q，h_i)函数计算对应信息(值)v_i 的重要程度，而 softmax 函数使该重要程度归一化为 0～1 之间的概率值。在 Liao 等人[54]采用的层级注意力模型中，键和值相同，即在词聚合层和句子聚合层分别为词表达和句子表达。通过这样的注意力机制，对词、句子等不同粒度的文本进行层级加权聚合，有效刻画了长文本的结构及语义特性，如图 4 所示。

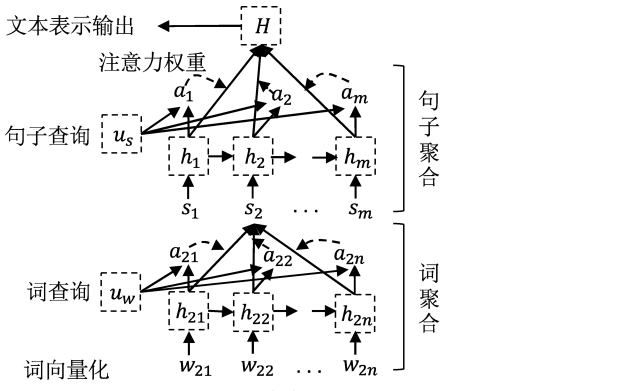


图 4　Liao 等人[54]采用的长文本表示模型

对于图像模态，研究者借鉴了图像特征学习领域的代表性模型，即采用了基于卷积神经网络的图像深度表示模型[8-9]。对于商品(包括音乐、电影、书籍等)的表示，Dou 等人[49]和 Zhao 等人[56]通过将

商品与知识库(knowledge base)中的实体对齐,利用知识库中实体的向量表示作为该商品的上下文信息补充,从而提升了对商品未来流行度的预测准确性。

2.1.2 多模态内容融合表示

除了各个模态的独立表示外,社交网络中的模态常是混合出现且相互影响的。比如新浪微博中,用户会在发布图像的同时配上文字,而这些文字可以让我们更好地去关注图像中的重点部分。为了刻画多模态之间的相互关联和影响,Zhang 等人[9]提出了基于注意力机制的多模态融合表示模型。该模型可以帮助我们根据查询模态,从目标模态中选择出更关键的信息。具体来说,以图像模态和文本模态融合为例。设查询模态为图像,目标模态为文本,则此时注意力机制中查询 h_q 为图像模态表示,键 h_i=值 v_i 为文本模态中对应部分 i(例如文本中某个词)的表示。通过注意力机制,我们可以得到图像模态查询下的文本表示。换言之,通过图像模态的查询,可以将文本模态中对应的部分给予更高的关注度,从而体现模态之间的相互关联和影响。类似地,我们也可以得到文本模态查询下的图像表示。最后,通过拼接文本表示和图像表示,得到基于注意力机制的多模态内容融合表示,模型示意见图 5。

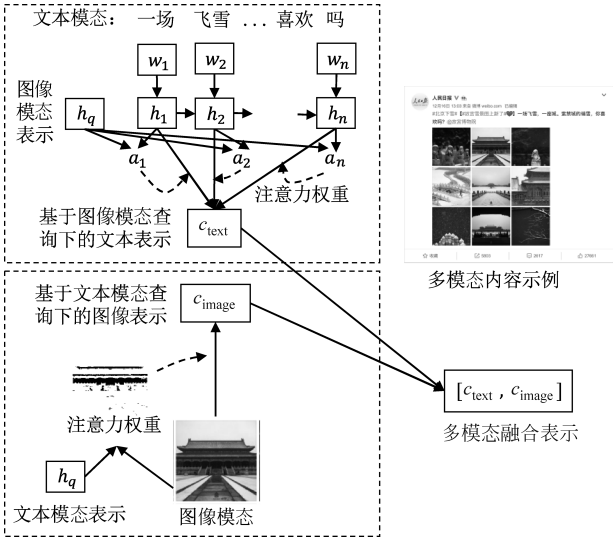


图 5 基于注意力机制的多模态融合示意图[9]

2.2 基于深度学习的用户表示

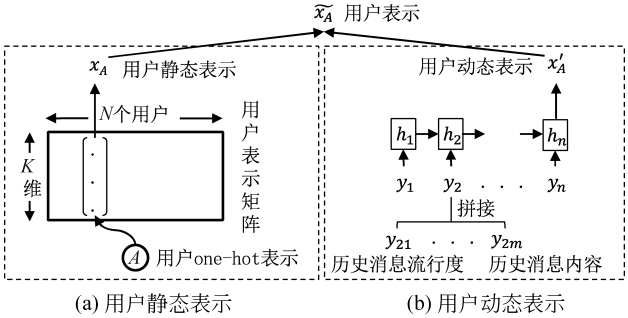
传统的流行度预测方法通常会提取用户的粉丝数、年龄、性别等属性特征以及用户的活跃度、历史发布消息的流行度等历史统计特征来表示发布用户或参与用户的特征,并学习这些特征对消息未来流行度的影响[28]。其中属性特征反映的是用户相对长期稳定的特性,而用户最近的历史统计特征反映的是用户短期的动态特性。仅仅采用启发式的特征提取方法很难适用于所有的场景,且难以发现和表示用户的一些隐藏特性。因此,有研究者分别针对用户长期静态特性和短期动态特性提出了基于深度学习的表示方法。相关文献整理见表 3。

表 3 基于深度学习的用户表示

分类	文献	深度学习模型
用户静态表示	Cao 等人[1],Zhang 等人[9],Li 等人[21],Wang 等人[41],Chen 等人[44,59],Wang 等人[47]	用户表示学习
用户动态表示	Wang 等人[47]	循环神经网络

针对用户的静态特性,结合表示学习近年来的蓬勃发展和成功应用[64-65],Li 等人[21]提出了用户表示学习的方法,即用一个 K 维的向量来表示用户影响,通过将用户映射到低维连续的向量空间中,使得影响力相似的用户具有相似的表示(user embedding)。该用户表示可以通过端到端的方式在流行度预测任务的指导下进行不同场景下的学习,具有更广泛的适用性。后续的诸多工作也都采用了这样的用户表示学习的方式[1,9,41,44,47,59]。

用户静态表示刻画的是用户长期稳定的特性,无法反映用户短期的动态特性。因此,Wang 等人[47]在静态用户表示的基础上,采用循环神经网络对用户近期内发布的 K 条消息的内容和流行度等信息,按照时间先后顺序进行序列化建模表示,并将循环神经网络最后一个节点的输出作为该用户的短期动态特性表示。最终用户表示由用户静态表示和动态表示共同作用得到,从而能同时反映用户的长期稳定特性和短期偏好特性。模型示意图如图 6 所示。



(a) 用户静态表示 (b) 用户动态表示

图 6 基于深度学习的用户表示建模示意图[47]

2.3　基于深度学习的时序表示

由于预测场景的不同，观测到的时序数据也有不同的表现形式。常见的主要有事件序列(event sequence)和时间序列(time series)两种，如图 7 所示。我们对这两种不同类型的时序数据的表示方法进行归纳总结。相关文献整理见表 4。

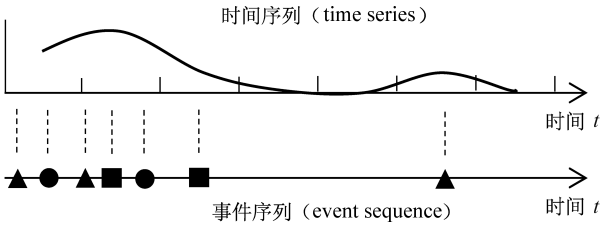


图 7　事件序列和时间序列示意图

表 4　基于深度学习的时序表示

分类			文献	深度学习模型
	序列生成/建模方式	优化方式		
事件序列表示	点过程速率函数特定建模	最大似然优化	—	传统点过程
	点过程速率函数深度建模	—	Cao 等人[1]，Du 等人[39]，Wu 等[42]	循环神经网络
	—	推土机距离优化	Xiao 等人[43,45,53]，Yan 等人[48]	生成对抗网络
	—	奖励函数优化	Li 等人[51]，Upadhyay 等人[52]，Yang 等人[57]	强化学习
时间序列表示	短时间序列：波动特性		Liao 等人[54]	卷积神经网络
	短时间序列：趋势特性		Chen 等人[44,59]，Liao 等人[54]	循环神经网络
	长时间序列		Shao 等人[55]	时序卷积网络
事件序列和时间序列表示融合			Xiao 等人[40]，Mishra 等人[46]	表示拼接

2.3.1　事件序列表示

事件序列不仅记录了事件的类型，同时也记录了事件发生的时间，通常可以表示为 $\{\xi_i=(z_i,t_i)\}_{i=1:N}$。其中 z_i、t_i 分别表示第 i 个事件的类型和时间。需要指出的是，在消息传播的场景下，事件类型和时间一般指参与传播的用户类型和参与传播的时间。点过程模型为事件序列的建模提供了一个通用且强大的框架。研究者根据不同的先验知识，设计了点过程速率函数的不同参数化形式[5-7]。根据参数化后的速率函数，可以通过最大化未来事件的发生似然来进行参数求解。但是，点过程也存在着两个关键的缺陷：①由简单特定的参数化形式构成的速率函数，很难刻画真实场景下的复杂传播，在一定程度上限制了模型的表达能力[39]。②通过最大似然估计(maximum likelihood estimate，MLE)来进行点过程的参数优化求解，常常会使模型陷入严重的模式坍塌等问题[43]。因此，基于深度学习的事件序列表示主要针对这两个问题，提出了不同的解决方案。

为了克服简单的参数化形式对点过程模型能力的限制，Cao 等人[1]在自激励点过程(Hawkes process)的基础上，提出了 DeepHawkes 模型，即在点过程速率函数的建模中引入了用户表示、基于循环神经网络的路径建模以及非参的时间衰减函数，增强了点过程模型的表达能力。但是，DeepHawkes 模型仍然是在自激励点过程的假设下进行的速率函数表示，仍具有模型错误选择的风险。为了放宽速率函数的特定假设，Du 等人[39]率先提出，点过程中的速率函数可以看成历史事件序列的一个非线性映射，并通过循环神经网络来刻画这一非线性映射(见图 8)。通过将历史事件(类型和时间的点对)按时间顺序逐个输入到循环神经网络中，可以刻画历史事件的序列化特性，并通过循环神经网络中节点的隐表示得到该事件序列当前时间步的速率函数值。根据该速率函数值可以预测下一个事件(类型及时间)发生的概率，从而通过最大似然进行模型参数优

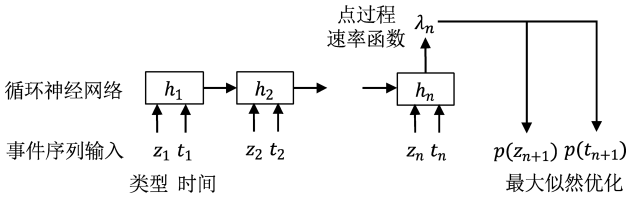


图 8　Du 等人[39]提出的基于循环神经网络的点过程建模

化学习。由于不需要预先假设固定的速率函数参数形式，基于循环神经网络的点过程模型能更好地适应复杂的真实场景，也成为了后续诸多流行度预测方法采用的一个基础模型[42-43,51-52]。

上述模型仍然是通过最大化观测似然来进行优化求解的，面临着模式坍塌等问题。为了避免点过程基于最大似然优化的缺点，Xiao 等人[43,45,53]率先提出了基于生成对抗网络(generative adversarial network，GAN)的深度点过程模型。该模型通过生成器来进行事件序列的生成，然后通过推土机距离(Wasserstein distance)，即判别器，衡量生成序列样本和真实序列样本的分布差异，使得生成的样本序列尽可能贴近真实样本，也就是我们观测到的真实事件序列。此时，则认为该生成器就是对事件序列的一个优秀建模，也是我们最终希望得到的对事件序列的表示模型。生成器通常采用循环神经网络[43,53]、序列到序列模型[45]或点过程模型[48]等，而判别器通常采用循环神经网络[43]、卷积神经网络[45,53]等。整个模型框架如图 9 所示。

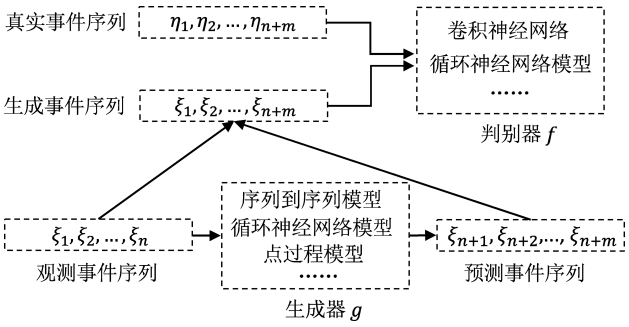


图 9 基于生成对抗网络的事件序列建模[43,45,48,53]

此外，Li 等人[51]和 Utkarsh 等人[52]也提出了基于强化学习(reinforcement learning，RL)的深度点过程模型，从而避免了最大似然优化的缺点。该类模型主要将事件序列的生成看成是随机策略(stochastic policy)下的动作(action)执行，通过执行策略后的奖赏函数(reward function)进行策略的优化学习。其中的策略函数就是我们希望得到的对于事件序列的表达模型，通常也用循环神经网络[51-52]或点过程模型[57]来刻画。整个模型框架如图 10 所示。

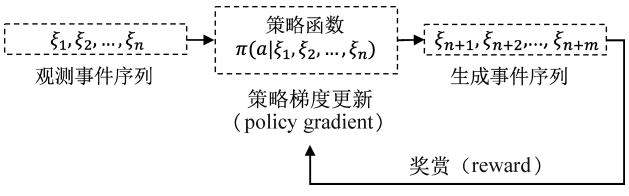


图 10 基于强化学习的事件序列建模[51-52,57]

上述这两类框架，即基于生成对抗网络的深度点过程模型和基于强化学习的深度点过程模型，都在有效学习点过程模型的同时避免了基于最大似然优化的缺点。

2.3.2 时间序列表示

与事件序列不同的是，时间序列$\{y_i\}_{i=1:T}$中的每一个值记录的是在固定且相等时间间隔内事件发生的频次，即固定时间间隔内消息的流行度。消息的流行度在时间序列上具有长期和短期两种不同的趋势特征。为了刻画消息的短期特征，Liao 等人[54]采用了一维卷积神经网络来捕获消息短期内具有平移不动性的波动特征。而为了更好地刻画消息流行度在时间序列上的长期趋势和特征，Chen 等人[44,59]采用了循环神经网络来进行时间序列的表示。但随着观测时序的增长，循环神经网络的性能会大幅下降，存在梯度消失或爆炸，以及串行计算耗时等问题。Shao 等人[55]发现可以通过时间卷积网络(temporal convolutional network，TCN)有效解决观测时序变长的问题，从而更好更快地进行消息的流行度预测。具体来说，时间卷积网络中采用了因果、扩张卷积，即$F^l(t)=\sum_{i=0}^{k-1} f^l(i)\xi^l_{t-d\cdot i}$，$l$为卷积层数，$f^l(i)$为卷积核，$k$为卷积核大小，$d$为扩张系数。因果卷积，即$t-d\cdot i\leqslant t$保证了没有从未来到过去的信息泄露。而扩张卷积，即越到上层d越大，有效加大了感受野，使模型具备处理长序列数据的能力。通常d随着层数的增加呈指数增长，如$d=O(2^l)$。模型示意图如图 11 所示。

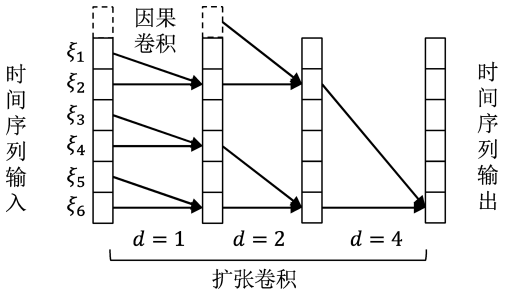


图 11 Shao 等人[55]采用的时间卷积网络示意图

注：卷积核大小 $k=2$

2.3.3 事件序列表示和时间序列表示的融合

当我们既能观测到事件序列，也能观测到时间序列的时候，Xiao 等人[40]认为这两个序列表示之间可以相互补充融合。具体来说，事件序列可以捕获事件驱动的、突然陡峭的信息；而时间序列更易于捕获同步或恒定的长期趋势。通过将这两个不同的序

列表示进行拼接融合，能够得到对时序数据更好的表示。Mishra 等人[46]在其后续的工作中也采用了类似的做法。

2.4　基于深度学习的结构表示

已有研究证明，不同的传播结构对于消息未来的流行度也具有不同的影响[2,19,66]。根据这一基本认识，基于深度学习的结构表示旨在通过端到端的方式对消息的传播结构进行更好的表示。根据表示对象的不同，主要分为对参与用户子图（cascade graph）的表示和对全局传播图（global diffusion graph）的表示（见图 12）。需要注意的是，这里的参与用户子图和全局传播图都是在观测时间窗口内的传播情况。相关文献整理如表 5 所示。

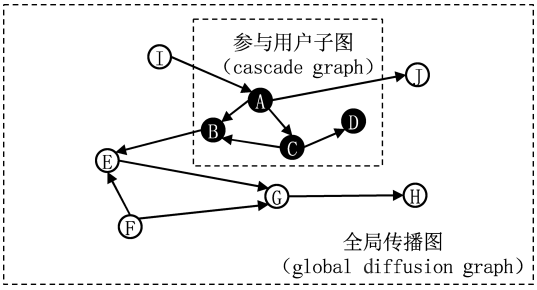


图 12　参与用户子图和全局传播图示意图

表 5　基于深度学习的结构表示

分类		文献	深度学习模型
参与用户子图表示	图序列化	Li 等人[21]	循环神经网络
	有向无环图	Wang 等人[41]	拓扑循环神经网络
	原始子图	Chen 等人[23,58]	图神经网络
全局传播图表示		Cao 等人[60]	耦合图神经网络

对于参与用户子图的表示，Li 等人[21]率先提出 DeepCas 模型，通过随机游走的方式将子图转化为 K 条序列，巧妙地将复杂的图结构转换成了容易处理的序列结构。具体来说，该模型利用循环神经网络对每条游走序列分别建模，最终通过注意力机制对不同游走序列赋以不同的权重进行加权聚合，从而得到参与用户子图的表示。模型框架如图 13 所示。

这样的序列转换方式，不可避免地丢失了一些图结构的信息。因此，后续的研究者开始直接对子图结构进行表示。Wang 等人[41]通过用户参与的时间先后顺序作为约束，只保留 $t_u < t_v$ 的对应连边 (u,v)，从而将参与用户子图转化为有向无环图。

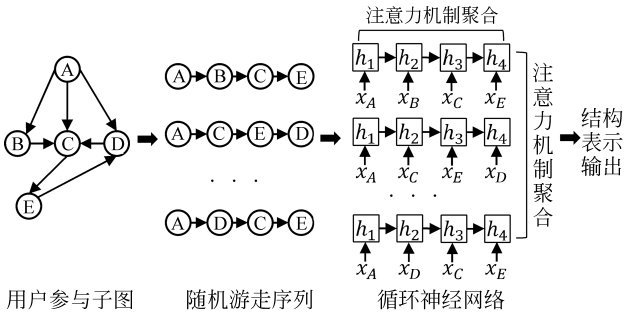


图 13　Li 等人[21]提出的 DeepCas 模型

通过提出的拓扑循环神经网络模型（Topo-LSTM）来有效刻画这类有向无环图的结构特性，即子图中的每个节点受且仅受自己和指向自己的节点所影响。最终子图的表示由所有节点表示经池化函数聚合得到。模型结构如图 14 所示。与此同时，随着图神经网络的蓬勃发展，Chen 等人[23,58]提出直接利用图神经网络模型来建模传播子图的结构特点，即通过图神经网络的邻居聚合函数，有效刻画局部的邻居结构特性，从而在流行度预测任务上获得了一定的性能提升。

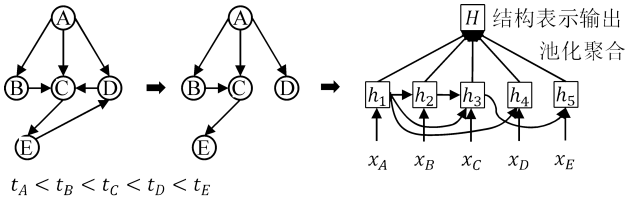


图 14　Wang 等人[41]提出的 Topo-LSTM 模型

上述提到的这些工作，都是对参与用户子图进行刻画表示。但消息流行度预测的其中一大挑战是来自于级联传播的不确定性，而仅仅关注参与用户子图，无法捕捉或刻画这样的级联传播效应。因此，Cao 等人[60]提出对全局传播图进行表示。具体来说，通过提出的耦合图神经网络（CoupledGNN），建模级联传播中关键的用户激活状态和影响力的相互作用和迭代传播，从而有效建模了传播中的级联效应。最终消息的结构表示由所有用户的激活状态表示池化聚合得到。模型结构如图 15 所示。

3　基于深度融合的流行度预测方法

本节我们将综述基于深度融合的流行度预测方法。这里的深度融合根据融合对象的不同，可以分

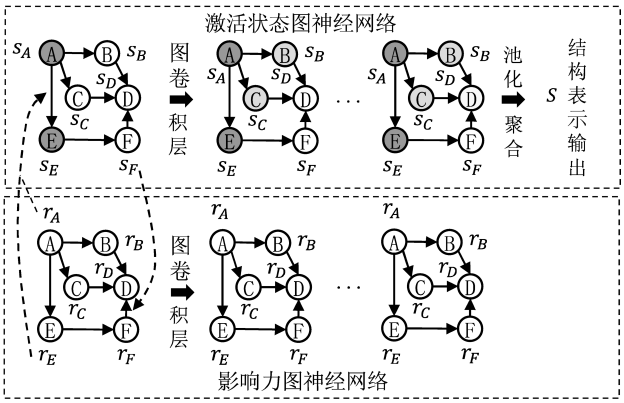


图 15 Cao 等人[60]提出的 CoupledGNN 模型

注：s_* 表示用户激活状态，r_* 表示用户影响力

为多因素融合和多模型融合。现有基于深度融合的流行度预测方法总结见表 1(因素融合和模型融合两列)。

3.1 基于深度学习的多因素融合

在得到内容、用户、时序、结构等因素的深度表示之后，我们还需要对各类因素进行有效融合，从而一起进行消息未来流行度的预测。在研究早期，与传统的基于特征的流行度预测方法类似，研究者通常采用最简单直观的拼接方式将各类因素的表示融合，然后基于拼接后的表示得到最终消息未来流行度的预测模型[44,47]。但 Liao 等人[54]随后发现，对于不同观测阶段、不同发布用户或不同特性的消息而言，不同因素所起的预测作用是不同的。例如，在观测早期，时序因素包含的信息非常少，此时消息本身的内容等因素就会比时序因素更重要；而到了观测后期，时序因素中蕴含了非常丰富的信息，那么此时，时序因素就会比内容因素对消息未来流行度更具有预测性。为了更好地刻画这些不同因素对于消息流行度预测的不同重要程度，Liao 等人[54]提出了基于注意力机制的多因素融合框架。在该框架下，注意力机制中的查询为消息的上下文信息，包括观测时间[54]、消息发布的时间[54]、消息本身的话题[59]或消息发布的用户[9]等；键和值均为某因素的表示。模型框架如图 16 所示。通过该基于注意力机制的多因素融合框架，可以在不同上下文中灵活地捕捉各个因素的重要性并进行有效融合，从而对消息未来流行度进行更准确的预测。

3.2 基于深度学习的多模型融合

基于特征提取和基于点过程建模的流行度方法

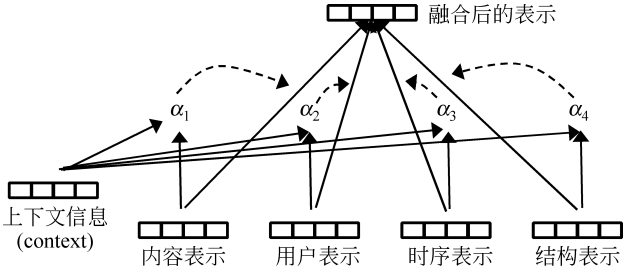


图 16 Liao 等人[54]提出的基于注意力机制的多因素融合框架

是当前用来解决社交网络中消息流行度预测问题的两类基本范式。但是，这两类方法都有各自的缺陷。基于特征提取的流行度预测方法需要人工设计各类特征，一些特征在特定的场景下可能会有缺失。而基于点过程建模的流行度预测方法需要足够长时间的观测才能保证模型的有效学习，且已有的点过程模型面临着模型错误选择的困境。为了弥补这两类方法各自的不足，充分发挥其各自的优点，Wu 等人[50]提出了基于生成对抗学习的多模型融合方法。一方面，基于点过程建模的方法就像一个“生成器”，可以用于未来流行度的生成；而另一方面，基于特征提取的方法可以作为一个“判别器”，用于区分(特征，真实的流行度)所组成的点对和(特征，生成的流行度)组成的点对。基于这样的双方博弈，能够使得该模型融合框架充分发挥基于特征提取和基于点过程建模的流行度预测方法的预测优势。模型框架如图 17 所示。

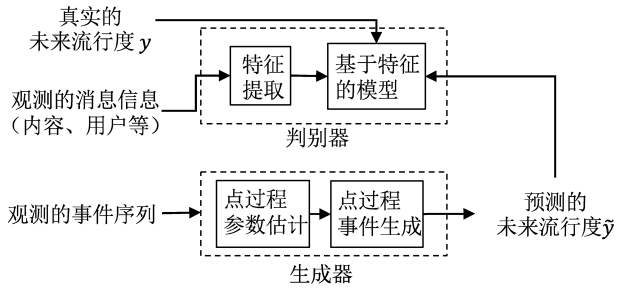


图 17 Wu 等人[50]提出的基于生成对抗学习的多模型融合框架

4 基于深度学习的流行度预测方法评价

消息流行度预测方法的性能需要有一个合理的评价体系来衡量。好的评价体系能指引流行度预测领域的前进方向，并使其健康发展。本节将从评价数据集、评价指标两个角度出发，总结现有的基于深度学习的流行度预测方法的评价，总体情况见表 6。

表 6　基于深度学习的流行度预测方法的评价总结

方法	数据集	数据集描述	来源	公开	评级指标
Cao 等人[1]	新浪微博-1	2016 年 6 月 1 日新浪微博数据	—	是	LMSE
	APS	APS 论文引用数据	Shen 等人[5]	否	
Sanjo 等人[8]	Cookpad	日本菜谱分享网站 Cookpad 中的菜谱数据	—	否	MSE、MAE
Li 等人[21]	推特-1	2016 年 6 月的推特数据	—	否	LMSE
	AMINER	ACM、DBLP、MAG 论文引用数据集	Tang 等人[67]	是	
Chen 等人[23]	新浪微博-1	2016 年 6 月 1 日新浪微博数据	Cao 等人[1]	是	LMSE
	HEP-PH	Arxiv 中高能物理论文引用数据	Leskovec 等人[22]	是	
Wu 等人[42]	TPIC17	Flickr 平台的图片流行度数据	—	是	SRC、MAE
Zhang 等人[9]	TPIC17	Flickr 平台的图片流行度数据	Wu 等人[42]	是	MSE、MAE
Chen 等人[44,59]	推特-2	2016 年 8 月 9 日到 12 月 10 日的推特数据	—	否	*C*-Accuracy
Mishra 等人[46]	推特-3	2011 年 10 月 7 日到 11 月 7 日的推特数据	Zhao 等人[6]	是	APE
	YouTube	2014 年 5 月 29 日到 12 月 26 日的 YouTube 视频数据	Rizoiu 等人[7]	是	
Wang 等人[47]	豆瓣-1	豆瓣平台上上海、北京等地举办的事件流行度数据	—	否	MSE、MAE
	推特-4	2016 年 5 月的推特采样数据	—	否	
Yan 等人[48]	MAG	Microsoft Academic Graph 中的论文引用数据	—	否	MAPE、*R*-Accuracy
	NYSE	NYSE 的股票交易数据	Du 等人[39]	否	
Dou 等人[49]、Zhao 等人[56]	LFM-1b	Last.fm 平台搜集的用户听音乐记录数据	Schedl 等人[68]	是	MAPE、MRSE、*R*-Accuracy
	MovieLens	MovieLens 网站搜集的电影评分数据	Harper 等人[69]	是	
Wu 等人[50]	推特-3	2011 年 10 月 7 日到 11 月 7 日的推特数据	Zhao 等人[6]	是	MAPE、Kendall、Coveage@k
Xiao 等人[53]	MAG	Microsoft Academic Graph 中的论文引用数据	—	否	MAPE、*R*-Accuracy
Liao 等人[54]	微信数据	微信公众号的文章数据	—	否	*C*-Accuracy、F_1 值
Shao 等人[55]	新浪微博-1	2016 年 6 月 1 日新浪微博数据	Cao 等人[1]	是	LMSE
Yang 等人[57]	Memetracker	Memetracker 平台收集的流行于各大博客的短语数据	Leskovec 等人[70]	是	LMSE
	推特-5	2011 年 10 月的推特数据	Hodas 等人[71]	是	
	豆瓣-2	豆瓣平台的用户书籍阅读数据	Zhong 等人[72]	是	
Chen 等人[58]	新浪微博-1	2016 年 6 月 1 日新浪微博数据	Cao 等人[1]	是	MSE
	APS	APS 论文引用数据	Shen 等人[5]	否	
Cao 等人[60]	新浪微博-2	新浪微博数据，包含用户关注网络和 30 万微博消息	Zhang 等人[24]	是	MRSE、MAPE、*R*-Accuracy

注：来源中“—”表示该数据集由本文作者自己搜集得到

4.1　评价数据集

由于预测场景和预测对象的不同，在线社交网络的流行度预测研究中，采用了多种不同的评价数据集，包括推特、新浪微博等社交网站数据集[1,6-7,24,39,70-71]、论文引用数据集[22,67]以及音乐、电

影等商品时序数据集[68-69,72]。其中,只有部分数据集被公开提供,因此接下来主要介绍这些常用的公开数据集。

4.1.1 社交网站数据集

• **新浪微博-1数据集**①由Cao等人[1]提供,包含了2016年6月1日发布且转发数大于10的全量微博消息(119 313条),以及这些消息在发布后24小时内的转发情况。该转发情况除了包括用户的转发时间外,还提供了用户的转发路径。由于同时包含时序、用户、结构(路径)等信息,后续诸多基于深度学习的流行度预测方法[1,23,55,58]均采用了这一数据集作评价。

• **推特-3数据集**②由Zhao等人[6]提供,包含了2011年10月7日到11月7日发布的166 076条消息中,每个用户的参与时间和对应粉丝数。由于该数据集主要提供了时序信息,后续主要被用于时序深度表示的流行度预测方法[46,50]的评价。

• **YouTube数据集**③由Rizoiu等人[7]提供,包含了2014年5月29日至12月26日在推特中被提到过的视频,共计8 000万左右,以及这些视频在Youtube和对应的推特上的浏览、转发等时序数据。后续也被用于时序深度表示的流行度预测方法[46]的评价。

• **新浪微博-2数据集**④由Zhang等人[24]提供。该数据集从100个随机用户种子出发,通过关注关系搜集更多的用户。最终,数据集包括170万用户和这些用户最新的1000条微博(包括转发用户和转发时间),以及这些用户间的关注关系网络。由于该数据集包含了全局的社交网络,Cao等人[60]提出的基于全局传播图建模的预测方法使用了这一数据集作评价。

• **TPIC17数据集**⑤由Wu等人[42]提供,包含了Flicker平台3年的图像浏览分享数据。具体包括图像等内容信息、用户信息以及浏览、分享等时序信息。该数据集主要在图像流行度预测方法[9,42]的评价所采用。

• **Memetracker数据集**⑥由Leskovec等人[70]提供,包含了约9 600万在各大新闻平台、博客等流行的文化模因(Meme)。通过将Meme作为预测对象,博客或新闻平台作为用户,可获得对应消息下的用户、时序等信息。该数据集也在基于时序深度表示的Yang等人[57]的工作中作为评价数据集使用。

• **推特-5数据集**⑦由Hodas等人[71]提供,包含了2011年10月推特上发布的66 059个URL链接(消息)和对应的参与用户、参与时间等信息。同时,该数据集还提供了用户间的关注关系网络这一结构信息。基于时序深度表示的Yang等人[57]的工作中使用了这一数据集作评价。

4.1.2 论文引用数据集

• **HEP-PH数据集**⑧由Leskovec等人[22]提供,包含了1993年到2003年发布在Arxiv上的约3.5万高能物理相关论文,以及这些论文间的引用关系。此外,该数据集也包括论文的用户、时序以及结构信息。Chen等人[23]在其提出的基于传播子图表示的工作中采用了该数据集作为评价数据集。

• **AMINER数据集**⑨由Tang等人[67]提供,包括了从ACM、DBLP、MAG(Microsoft Academic Graph)以及其他来源获取到的约63万论文及相关引用数据。将论文作为预测对象,引用作为流行度,作者间的引用关系作为网络结构,得到对应的用户、时序以及结构信息。Li等人[21]提出的基于传播子图表示的工作采用了该数据集作为评价数据集。

4.1.3 时序数据商品集

• **LFM-1b数据集**⑩由Schedl等人[68]提供,包含了从Last.fm平台搜集到的用户听音乐数据集。将音乐作为预测目标,聆听总量作为流行度,可获得对应的用户、时序等信息。Dou等人[49]和Zhao等人[56]通过将音乐与知识图谱中的实体对应,采用该数据集评价了他们提出的基于商品内容(实体表达)及时序的流行度预测方法。

• **MovieLens数据集**⑪由Harper等人[69]提供,包含了从MovieLens网站搜集到的电影评分数据。类似LFM-1b数据集,将电影作为预测目标,评分总量作为流行度,可获得用户、时序信息。Dou等人[49]和Zhao等人[56]也将该数据集作为他们方法的评价数据集。

• **豆瓣-2数据集**⑫由Zhong等人[72]提供,包含

① http://github.com/CaoQi92/DeepHawkes
② http://snap.stanford.edu/seismic/
③ https://github.com/andrei-rizoiu/hip-popularity
④ http://www.aminer.cn/influencelocality
⑤ http://github.com/social-media-prediction/TPIC2017
⑥ http://snap.stanford.edu/data/memetracker9.html
⑦ http://www.isi.edu/~lerman/downloads/twitter/twitter2010.html
⑧ http://snap.stanford.edu/data/cit-HepPh.html
⑨ http://www.aminer.cn/citation
⑩ http://www.cp.jku.at/datasets/LFM-1b/
⑪ http://grouplens.org/datasets/movielens/
⑫ http://sites.google.com/site/erhengzhong/datasets

了豆瓣平台中用户对书籍的阅读记录。将书籍作为预测对象，书籍的阅读次数作为流行度，也可获得对应的用户及时序信息。该数据集在基于时序深度表示的 Yang 等人[57]的工作中被作为评价数据集使用。

4.2 评价指标

在基于深度学习的流行度预测方法评价中，根据预测任务和目标的不同，可分为分类、回归以及排序这三大类评价指标。为了从各方面全面地衡量预测方法的性能，通常会采用多个或多类指标来进行方法的评价。

4.2.1 分类评价指标

• **分类准确率（*C*-Accuracy）**[44,54,59]衡量流行度被准确分类的消息占比，如式(2)所示。

$$C-\text{Accuracy}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(l(y_i)=l(\hat{y}_i)) \tag{2}$$

其中 y_i 和 $\hat{y}_i$ 分别表示消息 i 的真实流行度和预测流行度，$l(*)$表示流行度的对应类别，如热门、冷门等类别。

• F_1[54]是精确率（Precision）和召回率（Recall）的调和平均数，用于衡量样本类别不均匀时的二分类性能：

$$F_1=\frac{2\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{3}$$

其中 Precision＝方法返回的正确正例数/方法返回的所有正例数，Recall＝方法返回的正确正例数/样本中的所有正例数。

4.2.2 回归评价指标

• **平均绝对误差（mean absolute error，MAE）**[8-9,42,47]用于衡量消息的真实流行度和预测流行度之间的差异绝对值，如式(4)所示。

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{4}$$

• **均方误差（mean square error，MSE）**[8-9,47,58]衡量消息的真实流行度和预测流行度之间的差异平方，如式(5)所示。

$$\text{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{5}$$

• **Log 均方误差（log mean square error，LMSE）**[1,21,23,55,57]先将消息的流行度进行对数（log）量级转换，从而衡量消息的真实流行度量级和预测流行度量级之间的差异平方，如式(6)所示。

$$\text{LMSE}=\frac{1}{n}\sum_{i=1}^{n}(\log(y_i)-\log(\hat{y}_i))^2 \tag{6}$$

• **平均绝对百分比误差（mean absolute percentage error，MAPE）**[48-50,53,56,60]衡量消息的真实流行度和预测流行度之间的差异占消息真实流行度的平均百分比，如式(7)所示。

$$\text{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{7}$$

• **平均相对均方误差（mean relative squared error，MRSE）**[49,56,60]和 MAPE 指标类似，衡量真实流行度和预测流行度之间的差异占消息真实流行度的百分比均方值，如式(8)所示。

$$\text{MRSE}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2 \tag{8}$$

由于避免了不可导的绝对值操作，该评价指标更易优化。

• **回归准确率（*R*-Accuracy）**[48-49,53,56,60]衡量在误差接受范围内，消息流行度被正确预测的占比，如式(9)所示。

$$R\text{-Accuracy}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\left|\frac{y_i-\hat{y}_i}{y_i}\leqslant\epsilon\right| \tag{9}$$

其中，ϵ 为接受的预测误差。

• **绝对百分位误差（absolute percentile error，APE）**[46]先将消息的流行度进行排序百分位转换，然后衡量消息的真实流行度百分位和预测流行度百分位之间的差异绝对值，如式(10)所示。

$$\text{APE}=\frac{1}{n}\sum_{i=1}^{n}|p(y_i)-p(\hat{y}_i)| \tag{10}$$

4.2.3 排序评价指标

• **斯皮尔曼排序相关性（Spearman ranking correlation，SRC）**[42]用于衡量所有消息的真实流行度和预测流行度之间的排序相关性，如式(11)所示。

$$\text{SRC}=\frac{1}{n-1}\sum_{\text{rank}=1}^{n}\left(\frac{y_{\text{rank}}-\bar{y}}{\sigma_y}\right)\left(\frac{\hat{y}_{\text{rank}}-\bar{\hat{y}}}{\sigma_{\hat{y}}}\right) \tag{11}$$

其中，$\bar{y}$ 和 σ_y 分别表示所有消息的真实流行度的平均值和方差，$\bar{\hat{y}}$ 和 $\sigma_{\hat{y}}$ 分别表示所有消息的预测流行度的平均值和方差。SRC 值越大，表示真实流行度和预测流行度的排序越相关。

• **肯德尔排序相关系数（Kendall rank coefficient，Kendall）**[50]也是用于衡量排序相关性的一种常用方式，如式(12)所示。

$$\text{Kendall}=\frac{2(C(y,\hat{y})-D(y,\hat{y}))}{n(n-1)} \tag{12}$$

具体来说，将消息 i 的真实流行度和预测流行度形成一个元素，即 $E_i=(y_i,\hat{y}_i)$。对于任意配对

的 E_i 和 E_j，当且仅当 $y_i \leqslant y_j \wedge \hat{y}_i \leqslant \hat{y}_j$ 或 $y_i \geqslant y_j \wedge \hat{y}_i \geqslant \hat{y}_j$ 时，这两个元素的配对被称为是一致的。$C(y,\hat{y})$表示一致的配对数，$D(y,\hat{y})$表示不一致的配对数。Kendall 指标的取值范围为[−1,1]，越接近于 1，表示预测的流行度就越接近真实的流行度。

• **覆盖率(coverage@k)**[50] 计算前 k 个预测流行度最大的消息，其中真实流行度也位于前 k 的占比。该评价指标可以用于衡量预测方法对于热门消息的预测准确程度。

4.3 方法评价

基于深度学习的流行度预测方法由于所面向的预测场景、使用的信息或因素不同，其使用的评价数据集和评价方法也不尽相同，很难进行横向的统一比较。本文中，我们仅对其中采用了同一个公开数据集的部分方法，在同一评价指标上进行性能的横向对比，结果见表 7。从表中可以看到，在新浪微博-1 公开数据集上，在对比的基于深度学习的流行度预测方法中，Chen 等人[23] 提出的图卷积神经网络 CasCN，既可以有效捕捉传播子图的结构，也可以捕获时序特性，获得了最佳的流行度预测性能。

表 7 基于新浪微博-1 数据集[1]的方法性能对比（观测 1 小时）

方法	方法名	LMSE
Li 等人[21] WWW2017	DeepCas	2.958
Wang 等人[41] ICDM2017	Topo-LSTM	2.772
Cao 等人[1] CIKM2017	DeepHawkes	2.448
Chen 等人[23] ICDE2019	CasCN	2.242

5 发展现状分析和趋势展望

基于深度学习的流行度预测方法，通过对内容、用户、时序、结构等因素的深度表示，以及多因素、多模型的深度融合，实现了对消息未来流行度预测性能的大幅提升。但目前仍有以下一些开放问题亟待解决。

5.1 可解释性

基于深度表示的流行度预测方法，可以有效地表示各类复杂的因素。但如何在提升表示能力的同时兼顾因素的可解释性，这仍是一个需要解决的问题。特别是对事前预测的场景，相较于预测性能而言，人们同样也十分关注模型为后续其他消息的发布所提供的指导意义。到底什么样的消息内容更吸引人？该选择什么样的用户来进行消息的发布或传播？如何使基于深度学习的流行度预测方法具备一定的可解释性，是未来可以探寻的一个方向。可能的解决方案是从建模前和建模后两方面出发。在建模前，考虑建立具备可解释性的模型，如 attention 机制就具有解释因素重要性的能力。在建模后，考虑对学习到的深度学习黑箱参数，如用户表示等，进行一定的可视化及探究，从而提供一定的可解释性。

5.2 用户稀疏性

基于“词向量”的用户表示方法，可以灵活地刻画每个用户的特性，并通过端到端的方式进行训练学习。但随着社交网络的蓬勃发展，社交网络中的用户规模也不断扩大。每个用户参与的消息呈现极大的差异性，而参与的消息总数也呈幂律分布（如图 18 所示）。即大部分用户参与的消息总数都很少，只有少部分活跃用户参与了大量的消息。在用户参与消息如此稀疏的场景下，我们很难通过端到端的方式为每个用户学习到一个合适的表示。如何能在大规模的用户空间下，为每个非活跃用户也学到有效合适的表达，对于进一步提升消息流行度预测方法的可用性和准确性至关重要。考虑结合自监督学习等技术(self-supervised learning)[73]，通过辅助任务的定义和学习，来提升学习到的用户表示质量，克服用户稀疏性问题，是未来的一个可能方向。

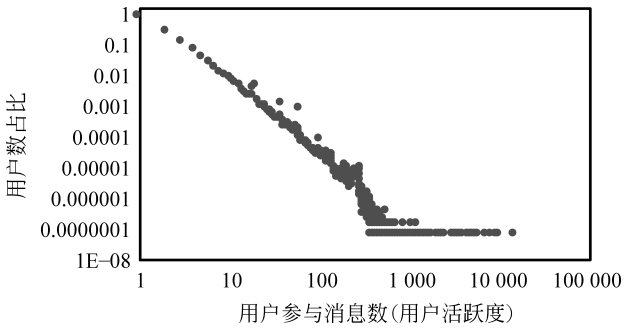


图 18 微博 2016 年 6 月 1 日用户活跃度分布

5.3 结构复杂性

对于传播结构的表示，现有方法或者面向参与用户子图或者面向全局传播图。但这两种表示对象都有各自的缺点。参与用户子图一般规模较小，能够在有限时间内进行高效表示，但无法刻画消息传

播中关键的级联传播效应。而在全局传播图中，我们能够更好地刻画参与用户对潜在用户形成的级联传播效应，从而更好地预测未来消息流行度。但这类方法的时间复杂度会随着网络规模的增大而增大，非常耗时，难以进行即时的消息流行度预测。目前基于深度学习的结构表示，还存在着预测精度和预测性能之间的两难困境。借鉴图摘要（graph summarization）领域的技术[74]，通过抽取简洁的“超图”来描述原本复杂的网络，并尽可能地减少图信息的损失，也许是未来解决这一困境的关键。

5.4　评价体系规范化

现有的基于深度学习的流行度预测研究中，由于面向的预测场景、预测任务、利用信息或因素的不同，导致性能评价体系中采用的评价数据集和评价指标差异很大。此外，很多研究所采用的数据集都是涉及隐私的非公开数据集，对于后续研究的可持续发展也很不利。为了该领域的健康发展，我们需要建立一套规范化的评价体系。具体来说，一方面需要对领域所需的评价指标达成共识，在同一套评价指标下（可以多个）对不同的方法进行统一评价。另一方面，也需要建立一个包含各类因素，即内容、用户、时序、结构的公开数据集，使得无论面向什么预测场景，利用哪类因素的方法，都可以在这个公开数据集上进行评价，从而使各方法能进行横向对比。从目前统计的公开数据集来看，我们仍缺少这样一个全面的基准数据集。规范化的评价体系的建立，对于流行度预测领域的可持续健康发展将会起关键作用。

6　结束语

在线社交网络中的消息流行度预测，对平台服务方和使用者都具有至关重要的作用。而深度学习的发展和社交网络中消息传播数据的积累，为基于深度学习的消息流行度预测研究提供了坚实的发展基础。本文对近年来基于深度学习的流行度预测研究进行了归纳梳理，将其分为基于深度表示的流行度预测方法和基于深度融合的流行度预测方法。其中基于深度表示的流行度预测研究包括内容、用户、时序、结构等因素的深度表示；而基于深度融合的流行度预测研究根据融合对象不同，分为多因素融合和多模型融合。本文主要围绕近年来基于深度学习的流行度预测研究展开，归纳梳理已有的研究成果，指出仍待解决的一些开放问题，尝试为研究人员建立一个较完整的研究视图，希望能为推进该领域进一步的研究提供一定的帮助。

参考文献

[1] Cao Q, Shen H, Cen K, et al. DeepHawkes: Bridging the gap between prediction and understanding of information cascades[C]//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. USA: ACM, 2017: 1149-1158.

[2] Shulman B, Sharma A, Cosley D. Predictability of popularity: Gaps between prediction and understanding [C]//Proceedings of the 10th International AAAI Conference on Web and Social Media. USA: AAAI, 2016: 348-357.

[3] Cheng J, Adamic L, Dow P A, et al. Can cascades be predicted? [C]//Proceedings of the 23rd International Conference on World Wide Web. USA: ACM, 2014: 925-936.

[4] Gao X, Cao Z, Li S, et al. Taxonomy and evaluation for microblog popularity prediction[J]. ACM Transactions on Knowledge Discovery from Data, 2019, 13 (2): 15-40.

[5] Shen H, Wang D, Song C, et al. Modeling and predicting popularity dynamics via reinforced poisson processes[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. USA: AAAI, 2014: 291-297.

[6] Zhao Q, Erdogdu M A, He H Y, et al. Seismic: A self-exciting point process model for predicting Tweet popularity[C]//Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2015: 1513-1522.

[7] Rizoiu M A, Xie L, Sanner S, et al. Expecting to be hip: Hawkes intensity processes for social media popularity[C]//Proceedings of the 26th International Conference on World Wide Web. Switzerland: International World Wide Web Conferences Steering Committee, 2017: 735-744.

[8] Sanjo S, Katsurai M. Recipe popularity prediction with deep visual-semantic fusion[C]//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. USA: ACM, 2017: 2279-2282.

[9] Zhang W, Wang W, Wang J, et al. User-guided hierarchical attention network for multi-modal social image popularity prediction[C]//Proceedings of the 27th International Conference on World Wide Web. Switzerland: International World Wide Web Conferences

Steering Committee,2018: 1277-1286.
[10] 李洋，陈毅恒，刘挺.微博信息传播预测研究综述[J].软件学报,2016,27(2): 247-263.
[11] 胡颖，胡长军，傅树深，等. 流行度演化分析与预测综述[J].电子与信息学报,2017,39(4): 805-816.
[12] 胡长军，许文文，胡颖，等. 在线社交网络信息传播研究综述[J].电子与信息学报,2017,39(4): 794-804.
[13] Hofman J M, Sharma A, Watts D J. Prediction and explanation in social systems[J]. Science, 2017, 355 (6324): 486-488.
[14] Pinto H, Almeida J M, Gonçalves M A. Using early view patterns to predict the popularity of youtube videos[C]//Proceedings of the 6th ACM International Conference on Web Search and Data Mining. USA: ACM, 2013: 365-374.
[15] Romero DM, Tan C, Ugander J. On the interplay between social and topical structure[C]//Proceedings of the 7th International AAAI Conference on Weblogs and Social Media. USA: AAAI, 2013: 516-525.
[16] Weng L, Menczer F, Ahn Y Y. Predicting successful memes using network and community structure[C]// Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. USA: AAAI, 2014: 525-544.
[17] Martin T, Hofman J M, Sharma A,et al. Exploring limits to prediction in complex social systems[C]// Proceedings of the 25th International Conference on World Wide Web. Switzerland: International World Wide Web Conferences Steering Committee, 2016: 683-694.
[18] Cao Q, Shen H, Gao H,et al. Predicting the popularity of online content with group-specific models[C]// Proceedings of the 26th International Conference on World Wide Web. Switzerland: International World Wide Web Conferences Steering Committee, 2017: 765-766.
[19] Bao P, Shen H W, Huang J,et al. Popularity prediction in microblogging network: A case study on Sina Weibo[C]//Proceedings of the 22nd International Conference on World Wide Web. USA: ACM, 2013: 177-178.
[20] Szabo G, Huberman B A. Predicting the popularity of online content[J]. Communications of the ACM, 2010,53(8): 80-88.
[21] Li C, Ma J, Guo X,et al. Deepcas: An end-to-end predictor of information cascades[C]//Proceedings of the 26th International Conference on World Wide Web. Switzerland: International World Wide Web Conferences Steering Committee,2017: 577-586.
[22] Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: Densification laws, shrinking diameters and possible explanations[C]//Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. USA: ACM, 2005: 177-187.
[23] Chen X, Zhou F, Zhang K,et al. Information diffusion prediction via recurrent cascades convolution [C]//Proceedings of the 35th International Conference on Data Engineering. USA: IEEE, 2019: 770-781.
[24] Zhang J, Liu B, Tang J,et al. Social influence locality for modeling retweeting behaviors[C]//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. USA: AAAI, 2013: 2761-2767.
[25] Qiu J, Tang J, Ma H,et al. DeepInf: Social influence prediction with deep learning[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2018: 2110-2119.
[26] Bakshy E, Hofman J M, Mason W A, et al. Everyone's an influencer: Quantifying influence on Twitter[C]//Proceedings of the 4th ACM International Conference on Web Search and Data Mining. USA: ACM, 2011: 65-74.
[27] Ahmed M, Spagna S, Huici F,et al. A peek into the future: Predicting the evolution of popularity in user generated content[C]//Proceedings of the 6th ACM International Conference on Web Search and Data Mining. USA: ACM, 2013: 607-616.
[28] Tsur O, Rappoport A. What's in a hashtag: Content based prediction of the spread of ideas in microblogging communities[C]//Proceedings of the 5th ACM International Conference on Web Search and Data Mining. USA: ACM, 2012: 643-652.
[29] Petrovic S, Osborne M, Lavrenko V. Rt to win! Predicting message propagation in Twitter[C]// Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. USA: AAAI, 2011: 586-589.
[30] Jenders M, Kasneci G, Naumann F. Analyzing and predicting viral tweets[C]//Proceedings of the 22nd International Conference on World Wide Web. USA: ACM, 2013: 657-664.
[31] 朱海龙，云晓春，韩志帅. 基于传播加速度的微博流行度预测方法[J]. 计算机研究与发展，2018，55(6): 1282-1293.
[32] 高金华，沈华伟，程学旗，等. 基于相似消息的流行度预测方法[J]. 中文信息学报，2018，32(11): 79-

85.

[33] Gao S, Ma J, Chen Z. Effective and effortless features for popularity prediction in microblogging network[C]//Proceedings of the 23rd International Conference on World Wide Web. USA: ACM, 2014: 269-270.

[34] 曹玖新，吴江林，石伟，等. 新浪微博网信息传播分析与预测[J]. 计算机学报，2014，37(4)：779-790.

[35] Gao S, Ma J, Chen Z. Modeling and predicting retweeting dynamics on microblogging platforms [C]//Proceedings of the 8th ACM International Conference on Web Search and Data Mining. USA: ACM, 2015: 107-116.

[36] Bao P, Shen H W, Jin X,et al. Modeling and predicting popularity dynamics of microblogs using self-excited hawkes processes[C]//Proceedings of the 24th International Conference on World Wide Web. USA: ACM, 2015: 9-10.

[37] Mishra S, Rizoiu M A, Xie L. Feature driven and point process approaches for popularity prediction [C]//Proceedings of the 25th ACM International Conference on Information and Knowledge Management. USA: ACM, 2016: 1069-1078.

[38] Yu L, Cui P, Wang F,et al. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics[C]//Proceedings of the 2015 IEEE International Conference on Data Mining. USA: IEEE, 2015: 559-568.

[39] Du N, Dai H, Trivedi R,et al. Recurrent marked temporal point processes: Embedding event history to vector[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2016: 1555-1564.

[40] Xiao S, Yan J, Yang X, et al. Modeling the intensity function of point process via recurrent neural networks[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. USA: AAAI, 2017: 1597-1603.

[41] Wang J, Zheng V W, Liu Z, et al. Topological recurrent neural network for diffusion pre-diction[C]// Proceedings of the 17th IEEE International Conference on Data Mining. USA: IEEE, 2017: 475-484.

[42] Wu B, Cheng W H, Zhang Y, et al. Sequential prediction of social media popularity with deep temporal context networks[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. USA: AAAI, 2017: 3062-3068.

[43] Xiao S, Farajtabar M, Ye X, et al. Wasserstein learning of deep generative point process models [C]//Proceedings of the 31st Conference on Neural Information Processing Systems. USA: Curran Associates, 2017: 3247-3257.

[44] Chen G, Kong Q, Mao W. An attention-based neural popularity prediction model for social media events [C]//Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics. USA: IEEE, 2017: 161-163.

[45] Xiao S, Xu H, Yan J, et al. Learning conditional generative models for temporal point processes[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. USA: AAAI, 2018: 6302-6309.

[46] Mishra S, Rizoiu M A, Xie L. Modeling popularity in asynchronous social media streams with recurrent neural networks[C]//Proceedings of the 20th International AAAI Conference on Web and Social Media. USA: AAAI, 2018: 201-210.

[47] Wang W, Zhang W, Wang J, et al. Learning sequential correlation for user generated textual content popularity prediction[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. USA: AAAI, 2018: 1625-1631

[48] Yan J, Liu X, Shi L, et al. Improving maximum likelihood estimation of temporal point process via discriminative and adversarial learning[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. USA: AAAI, 2018: 2948-2954.

[49] Dou H, Zhao W X, Zhao Y, et al. Predicting the popularity of online content with knowledge-enhanced neural networks[C]//Proceedings of the 24th KDD Deep Learning Day. 2018.

[50] Wu Q, Yang C, Zhang H, et al. Adversarial training model unifying feature driven and point process perspectives for event popularity prediction[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. USA: ACM, 2018: 517-526.

[51] Li S, Xiao S, Zhu S, et al. Learning temporal point processes via reinforcement learning[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. USA: Curran Associates, 2018: 10781-10791.

[52] Upadhyay U, De A, Rodriguez M G. Deep reinforcement learning of marked temporal point processes [C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. USA: Curran Associates, 2018: 3168-3178.

[53] Xiao S, Yan J, Yang X, et al. Publication popularity modeling via adversarial learning of profile-specific

dynamic process[J]. IEEE Access, 2018, 6: 19984-19992.

[54] Liao D, Xu J, Li G, et al. Popularity prediction on online articles with deep fusion of temporal process and content features[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. USA: AAAI, 2019: 200-207.

[55] Shao J, Shen H, Cao Q, et al. Temporal convolutional networks for popularity prediction of messages on social medias[C]//Proceedings of the 25th China Conference on Information Retrieval. Germany: Springer, 2019: 135-147.

[56] Zhao W X, Dou H, Zhao Y, et al. Neural network based popularity prediction by linking online content with knowledge bases[C]//Proceedings of the 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 2019: 16-28.

[57] Yang C, Tang J, Sun M, et al. Multi-scale information diffusion prediction with reinforced recurrent networks[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. USA: AAAI, 2019: 4033-4039.

[58] Chen X, Zhang K, Zhou F, et al. Information cascades modeling via deep multi-task learning[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. USA: ACM, 2019: 885-888.

[59] Chen G, Kong Q, Xu N, et al. NPP: A neural popularity prediction model for social media content[J]. Neurocomputing, 2019, 14(333): 221-230.

[60] Cao Q, Shen H, Gao J, et al. Popularity prediction on social platforms with coupled graph neural networks [C]//Proceedings of the 13th ACM International Conference on Web Search and Data Mining. USA: ACM, 2020: 70-78.

[61] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of 27th International Conference on Neural Information Processing Systems. USA: Curran Associates, 2013: 3111-3119.

[62] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. USA: Association for Computational Linguistics, 2016: 1480-1489.

[63] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. USA: Curran Associates, 2017: 5998-6008.

[64] Perozzi B, Al-Rfou R, Skiena. S. Deepwalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2014: 701-710.

[65] Tang J, Qu M, Wang M, et al. Line: Large-scale information network embedding[C]//Proceedings of the 24th International Conference on World Wide Web. Switzerland: International World Wide Web Conferences Steering Committee, 2015: 1067-1077.

[66] Zhang J, Tang J, Zhong Y, et al. Structinf: Mining structural influence from social streams[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. USA: AAAI, 2017: 73-79.

[67] Tang J, Zhang J, Yao L, et al. Arnetminer: Extraction and mining of academic social networks[C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2008: 990-998.

[68] Schedl M. The LFM-1b dataset for music retrieval and recommendation[C]//Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval. USA: ACM, 2016: 103-110.

[69] Harper F M, Konstan J A. The movielens datasets: History and context[J]. ACM Transactions on Interactive Intelligent Systems, 2016, 5(4): 1-19.

[70] Leskovec J, Backstrom L, Kleinberg J. Meme-tracking and the dynamics of the news cycle[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2009: 497-506.

[71] Hodas N O, Lerman K. The simple rules of social contagion[J]. Scientific Reports, 2014, 4(4343): 1-17.

[72] Zhong E, Fan W, Wang J, et al. Comsoc: Adaptive transfer of user behaviors over composite social network[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2012: 696-704.

[73] Kim D, Cho D, Kweon I S. Self-supervised video representation learning with space-time cubic puzzles [C]//Proceedings of the AAAI Conference on Artificial Intelligence. USA: AAAI, 2019: 8545-8552.

[74] Liu Y, Safavi T, Dighe A, et al. Graph summarization methods and applications: A survey[J]. ACM Computing Surveys, 2018, 51(3): 1-34.

(下转第 32 页)

states from social media[C]//Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA: ACM, 2019: 79-88.

[69] Du J, Zhang Y, Luo J, et al. Extracting psychiatric stressors for suicide from social media using deep learning[J]. BMC Medical Informatics and Decision Making, 2018, 18(2): 77-86.

李静(1983—),博士研究生,讲师,主要研究领域为社会网络用户行为。
E-mail: lijing8388@126.com

万常选(1962—),博士,教授,博士生导师,主要研究领域为数据挖掘、情感分析、信息检索等。
E-mail: wanchangxuan@263.net

刘德喜(1975—),通信作者,博士,教授,主要研究领域为社会媒体处理、自然语言处理、计算心理学。
E-mail: dexi.liu@163.com

(上接第 18 页)

曹婧(1992—),博士研究生,主要研究领域为社交网络分析和消息传播预测。
E-mail: caoqi@ict.ac.cn

高金华(1989—),博士,助理研究员,主要研究领域为公众观点挖掘。
E-mail: gaojinhua@ict.ac.cn

沈华伟(1982—),通信作者,博士,研究员,主要研究领域为社会计算和网络数据挖掘。
E-mail: shenhuawei@ict.ac.cn