

时空数据论文短期调研报告

袁铭潮

1. 背景

随着机器学习和深度学习的发展，已经能够在很多领域达到甚至超过人类的水准，也有很多领域在积极地融合深度学习，希望借助深度学习实现产业革命。归根结底，深度学习是基于数据的，正是因为大数据时代的来临才有了深度学习的蓬勃发展。

从简单的数字，到文本和图像，机器学习已经能够很好的处理很多类型的数据，并从中进行学习我们所需要的知识。在很多较为复杂的数据上，深度学习目前并不能表现的像简单数据上那么优秀，时空数据就是复杂数据中的一类。

时空数据是包含了时间属性和空间属性的数据，对于时空数据，我们需要学习到其中包含的时间关联和空间关联，借助这两大属性完成很多复杂环境下的任务。

2. 研究现状

传统的对于时空数据的处理是以数据挖掘的方式进行的，主要的方式为：(1) 统计学方法；(2) 关联规则挖掘；(3) 决策树方法；(4) 神经网络方法；(5) 多层次数据汇总归纳等。并在各领域根据自身应用特点提出了不少数据模型以满足不同领域的需求，这是今后研究过程中能够借鉴使用的。

对于时空数据的类型主要分为一下几类：(1) 数值型序列：即传统意义上的狭义的时间序列，其构成元素是数值型的；(2) 事务型序列：其构成元素是事务型的；(3) 事件型序列：其构成元素是事件。

数据挖掘方向的其他技术对于深度学习借鉴意义不大，但数据的预处理是相同的。数据预处理流程与普通数据相仿，常用技术有数据清洗、数据合成、数据变换和数据消减等。

同样，时空数据挖掘所适用的应用同样适用深度

学习来解决，大致分为环境与交通、基于位置的服务以及人体运动时空模式挖掘等等。目前，也有很多团队在使用基于时空数据的深度学习在进行尝试。

近两周，调研了 AAAI2019、ICLR2019、NIPS2019 所接受的论文中时空数据方向的论文，其中 AAAI2019 相关论文数量为 12 篇左右，ICLR2019 接受的论文中并没有找到相关论文，而 NIPS2019 也只找到了 3 篇。可见对于时空数据与深度学习相结合的方向的研究并不多。

3. 时空数据预测交通流的应用

这是根据来自 AAAI2019 的论文，由北京交通大学的团队所发表。他们的研究提出了一种基于注意力的时空图卷积网络，并应用在交通流量的预测上，仅从效果上看是全方位的超越了 LSTM、STGCN 等目前比较热门的网络模型，当然他们所设计的网络模型根据交通数据的特点进行了优化。

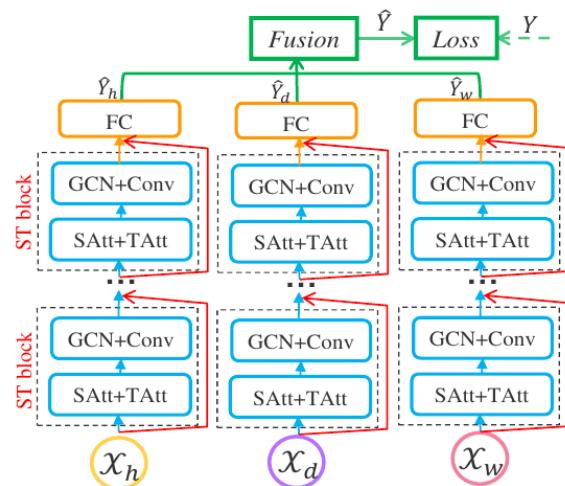


图 1. 基于注意力的时空图卷积网络模型

首先，他们对数据的时间特性和空间特性进行分

开处理，分别定义了两个 attention 对时间特性和空间特性分别进行学习：

$$S = V_s \cdot \sigma((\mathcal{X}_h^{(r-1)} W_1) W_2 (W_3 \mathcal{X}_h^{r-1})^T + b_s)$$

$$E = V_e \cdot \sigma((\mathcal{X}_h^{(r-1)} U_1) U_2 (U_3 \mathcal{X}_h^{r-1})^T + b_e)$$

其中 $\mathcal{X}_h^{(r-1)}$ 是第 r 个时空处理块的输入， $W_1, W_2, W_3, U_1, U_2, U_3$ 分别是需要通过训练学习到的参数矩阵，输出是矩阵 S, E ，分别为时间和空间上的注意力矩阵。

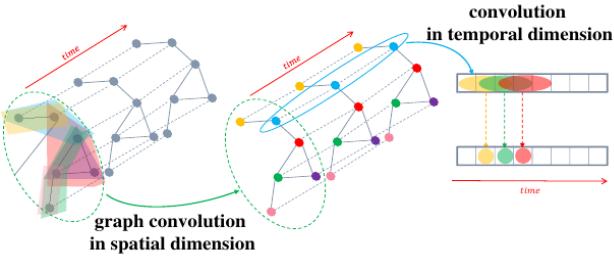


图 2. 图卷积示意图

时空处理块之后接的是一个图卷积层，用的是目前使用较为广泛的效率较高的图卷积方式。也同样是基于空域的图卷积的理论并不是很完备，这里使用的是基于频域的图卷积，以切比雪夫多项式近似计算卷积：

$$\hat{\mathcal{X}}_h^{(r)} = \text{ReLU}(\Phi * (\text{ReLU}(g_\theta *_{\mathcal{G}} \hat{\mathcal{X}}_h^{(r-1)})))$$

网络图中还使用了残差网络加深网络深度，以求达到更好的效果。

数据的处理中，考虑到交通数据的特性，对刚过去的一段时间、昨天该时段、上周这个时段等三段数据进行处理学习，并进行预测。

$$\hat{Y} = W_h \odot \hat{Y}_h + W_d \odot \hat{Y}_d + W_w \odot \hat{Y}_w$$

文中的创新点在于对数据分别结合 attention 得到时间和空间特性的矩阵，在图的处理上依旧是使用的目前比较流行的方法。

在 BERT、T5 等模型中所使用的 self-attention 是否也能很好的对时空属性契合，不知是否值得一试？

4. 时空数据与深度学习结合的应用

时空数据的应用场景目前还是以流的预测为主，譬如交通流的预测，AAAI2019 中，甚至有两篇基于时空

数据用于交通预测的论文，以及类似的由阿里巴巴针对包裹所研究的用于预估包裹送达时间的网络模型。

也有同样基于交通的，对共享单车需求的预测和滴滴出行对乘车需求的预测等等。

以上的预测大多是基于交通时空数据进行预测，目前交通数据确是与时空数据最为吻合且复杂度较低的时空数据。

更为复杂的有基于实时信息辅助城市犯罪分子拦截，对实时性和预测性要求更高，数据更复杂，不限于简单的交通数据，不过这篇论文中所描述的是基于简单的追逃游戏（PEG）的，与现实仍有很大出入。

其余的应用比如基于骨架的动作识别，自 2018 年的 ST-GCN 之后，有了较大的创新。还有团队将时空数据应用于重新识别（ReID）中，也取得了不错的效果。

5. 总结

根据这段时间对与时空数据结合深度学习方向的调研，目前这一方向的研究并不多，主要是几家公司的业务与时空数据较为重叠，比如菜鸟裹裹、滴滴出行等在进行时空数据方向的研究。

目前的成果也并不多，可能有数据集限制的原因在。但总体来说时空数据方向的研究潜力还是很大的，随着其他方向的模型的创新，结合时空特征的效果可能会出乎意料。

对于时空数据的应用还是有很大空间的，能够直观想到的应用比如：

- (1) 基于交通信息构建智慧交通；
- (2) 根据 GPS、GIS 信息提供位置服务；
- (3) 结合 ReID、人脸识别、交通信息等进行路径追踪等等。

时空数据所面临的技术难点还是很多，还有很大的研究空间。