上海交通大学-计算机科学与工程系

事件序列的建模与学习

严骏驰

2018年6月4日

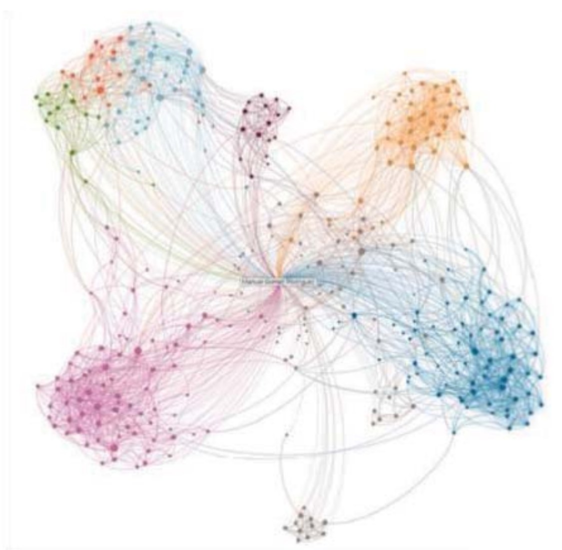
thinklab.sjtu.edu.cn

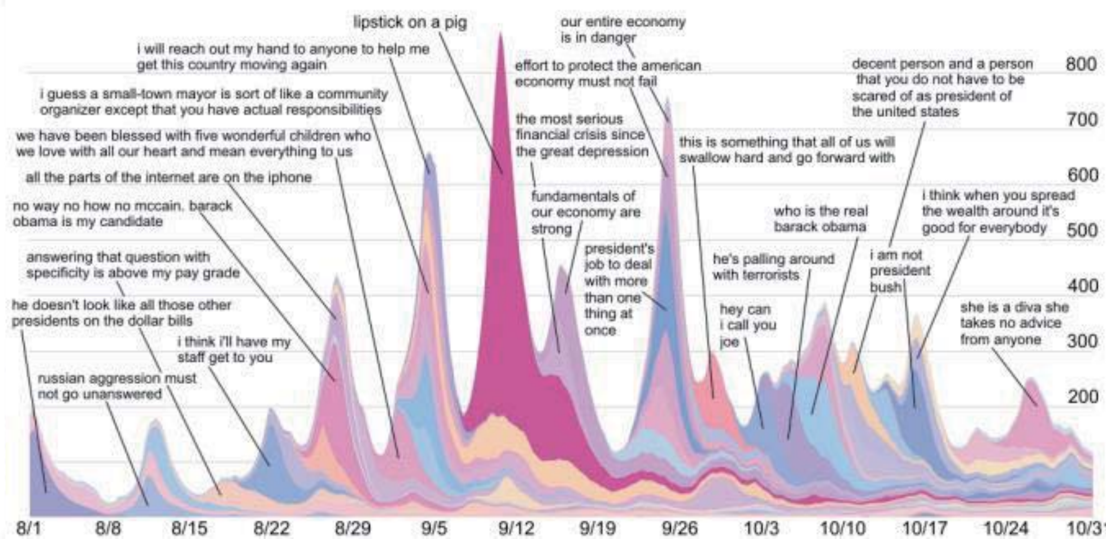
目录与概览

- 事件数据的内涵和例子
- 基于参数点过程的事件序列学习
  - Hawkes process
  - 点过程方法应用案例
    - 预防性维护保养（城市管道等资产密集型行业）
    - 科技文献引用预测（论文专利）
- 基于深度神经网络的事件序列学习
  - 结合时间序列和事件序列的统一深度学习框架
  - MIMIC、MEME、ATM机等问题实证
- 基于GAN框架的事件序列学习
- 总结与展望

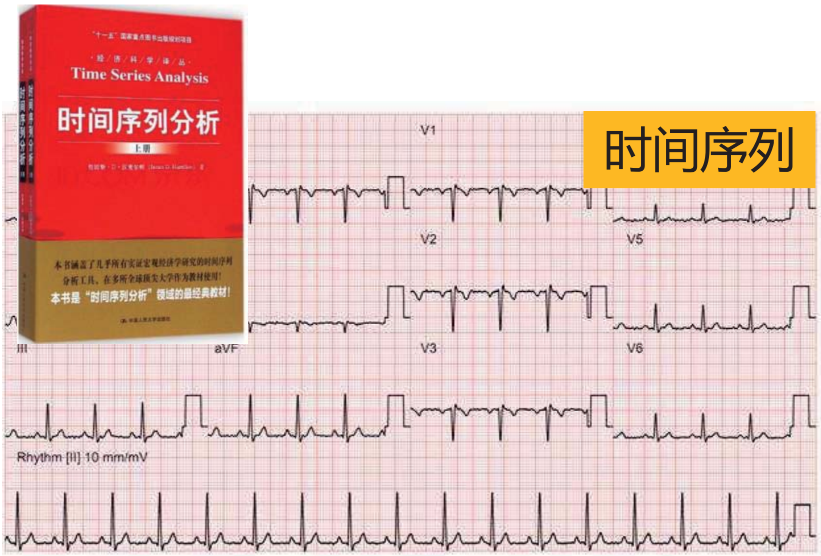
事件序列

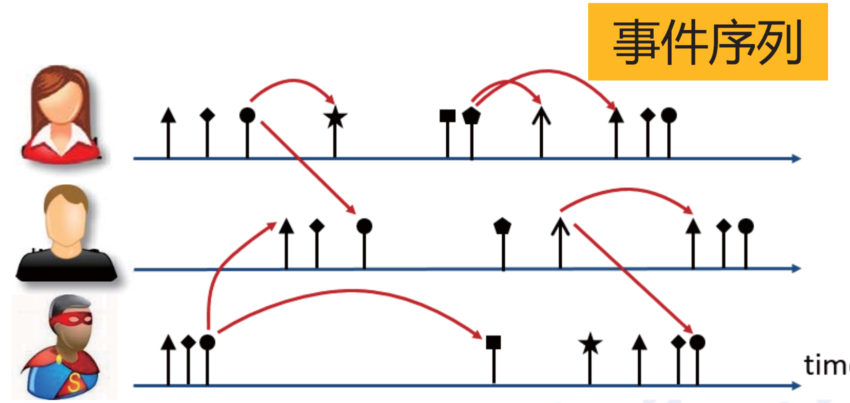
- 事件数据的信息
  - 动态信息：时间戳
  - 多方信息：各个参与方的信息，单方信息：类型、地点等
- 事件数据的特点
  - 时间间隔随机：事件序列（异步）<->时间序列（同步）
  - 不独立：序列间相关、序列内相关
  - 维度高：事件类型、事件主体、事件相关特征

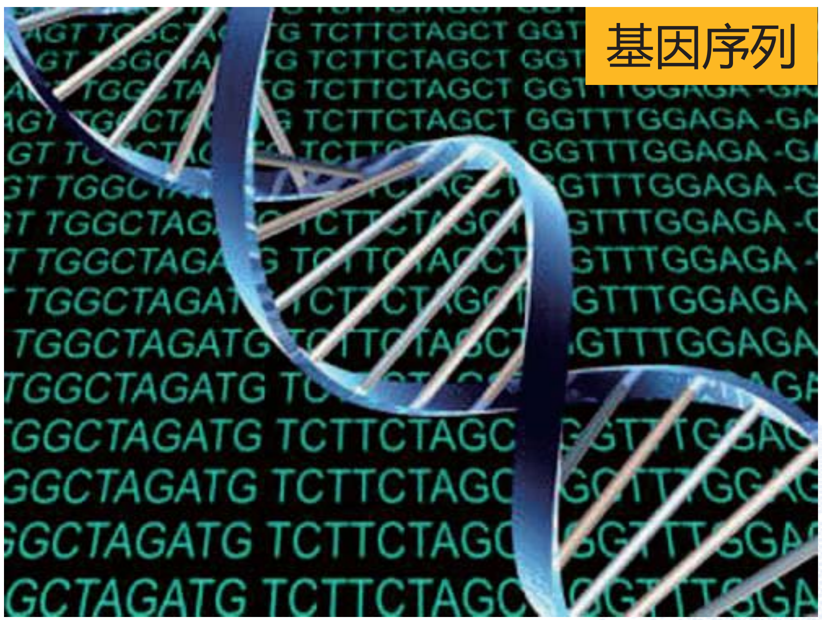




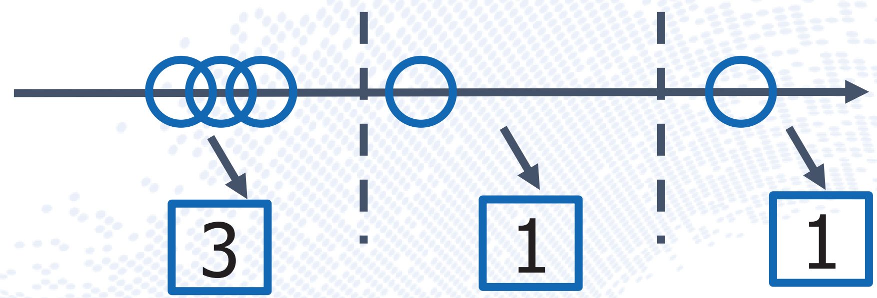
事件序列vs其他序列







离散化



他山之石（基于随机点过程模型）

数据形式	业务场景	应用方式	公开研究文献
交易数据	电商支付	推荐、欺诈检测	NIPS 2015
诊疗记录	询证医疗	精准医疗	Journal of the American Medical Informatics Association 2016
违约数据	担保/借贷	信用分析	SIAM J. Financial Math 2010
故障日志	设备管理	预防性维保	IJCAI 2013, Annals of Applied Statistics 2015
冲突日志	战场日志	反恐、军情分析	PNAS 2010, Security Journal 2011
文献引用	影响力评估	专利定价、学科评估	Science 2013
评论转发	社交网络	传播预测、溯源	ICML 2011, NIPS 2013

Georgia Institute of Technology

Stanford University

Massachusetts Institute of Technology

UCLA

IBM Watson Research Center

点过程模型-条件强度函数

- 条件强度函数

事件数目　　历史

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{E}(\boxed{N(t+\Delta t)} - N(t)\boxed{|\mathcal{H}_t})}{\Delta t} = \frac{\mathbb{E}(dN(t)|\mathcal{H}_t)}{dt}$$

- 常见参数化模型形式（Hawkes process）

背景项成分　　作用项项成分

$$\lambda^*(t) = \boxed{\gamma_0} + \alpha \sum_{t_j < t} \boxed{\gamma(t, t_j),}$$

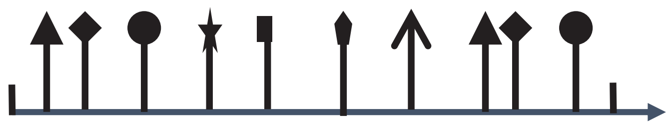
- 最大概率估计（Maximum Likelihood Estimation）

$$\mathcal{L} = \log \frac{\prod_{i=1}^{n} \lambda(t_i)}{\exp(\int_0^T \lambda(t)dt)} = \sum_{i=1}^{n} \log \lambda(t_i) - \int_0^T \lambda(t)dt.$$

典型点过程强度函数形式

Poisson processes:

$$\lambda^*(t) = \lambda$$

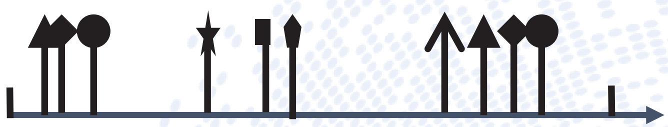


Terminating point processes:

$$\lambda^*(t) = g^*(t)(1 - N(t))$$

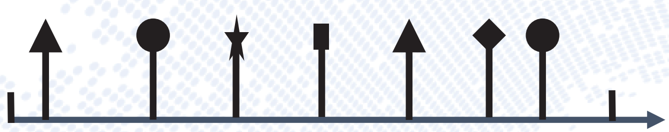
Self-exciting point processes:

$$\lambda^*(t) = \mu + \alpha \sum_{t_i \in \mathcal{H}(t)} \kappa_\omega(t - t_i)$$



Self-correcting processes:

$$\lambda^*(t) = e^{\mu t - \sum_{t_i \in \mathcal{H}(t)} \alpha}$$



基于最大后验概率的模型参数训练算法

$$
\begin{aligned}
\mathcal{L}(\Theta) &= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \sum_{u=1}^{U} \int_0^T \lambda_u(t)dt \\
&= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \sum_{u=1}^{U} \int_0^T (\mu + \sum_{(t',u'):t'<t} a_{uu'} g(t-t'))dt \\
&= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \left(\mu UT + \sum_{u=1}^{U} \int_0^T \sum_{(t',u'):t'<t} a_{uu'} g(t-t')dt \right) \\
&= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \left(\mu UT + \sum_{u=1}^{U} \sum_{i=0}^{n} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} a_{uu_j} g(t-t_j)dt \right) \\
&= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \left(\mu UT + \sum_{u=1}^{U} \sum_{i=0}^{n} \sum_{j=1}^{i} a_{uu_j} \int_{t_i}^{t_{i+1}} g(t-t_j)dt \right) \\
&= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \left(\mu UT + \sum_{u=1}^{U} \sum_{i=0}^{n} \sum_{j=1}^{i} a_{uu_j} \left(G(t_{i+1}-t_j) - G(t_i-t_j)\right) \right) \\
&= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \left(\mu UT + \sum_{u=1}^{U} \sum_{j=1}^{n} a_{uu_j} \sum_{i=j}^{n} \left(G(t_{i+1}-t_j) - G(t_i-t_j)\right) \right) \\
&= \sum_{i=1}^{n} \log \lambda_{u_i}(t_i) - \left(\mu UT + \sum_{u=1}^{U} \sum_{j=1}^{n} a_{uu_j} (G(T-t_j) - G(0)) \right)
\end{aligned}
$$

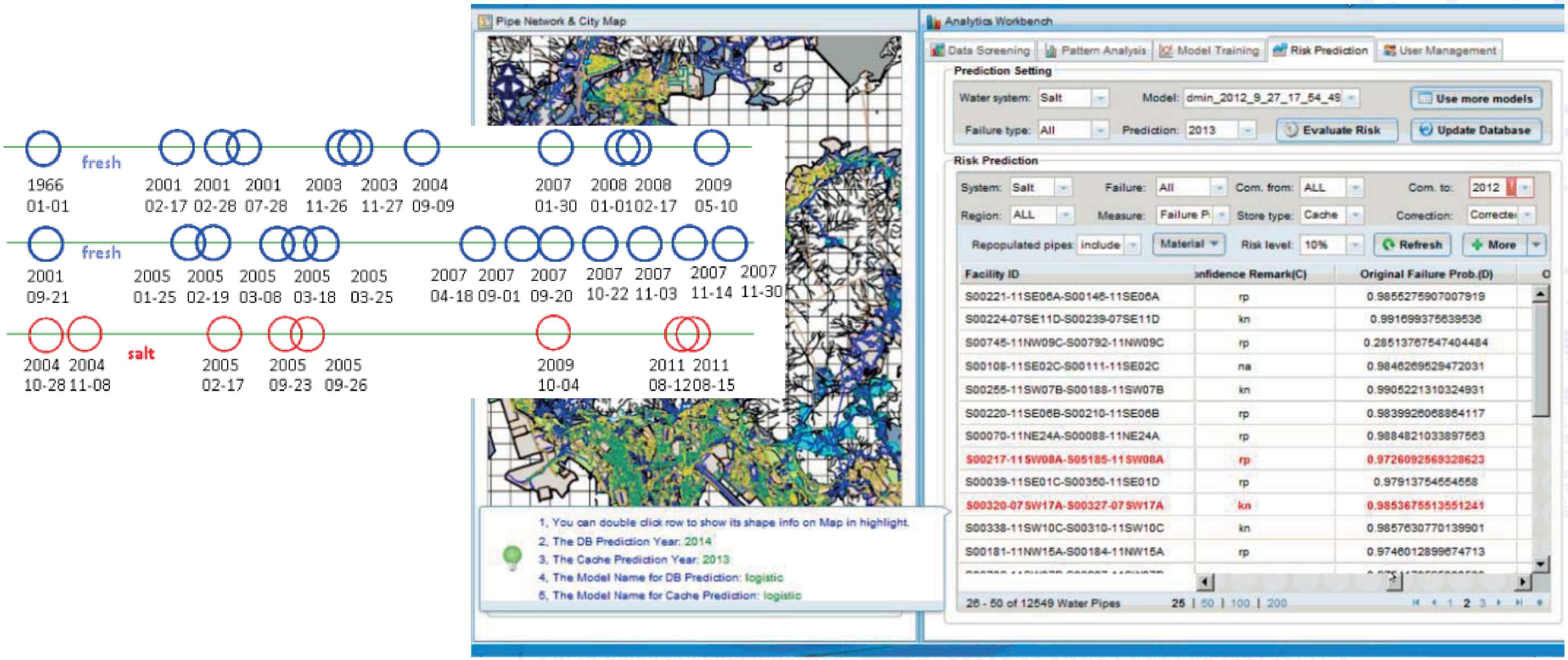
EM algorithm for solving the maximum likelihood problem:

$$
\begin{aligned}
\mathcal{L}(\Theta) &= \sum_{i=1}^{n} \log \left(\mu + \sum_{j=1}^{i-1} a_{u_i u_j} g(t_i - t_j) \right) - \left(\mu UT + \sum_{u=1}^{U} \sum_{j=1}^{n} a_{uu_j} (G(T-t_j) - G(0)) \right) \\
&\geq \sum_{i=1}^{n} \left(p_{ii} \log \frac{\mu}{p_{ii}} + \sum_{j=1}^{i-1} p_{ij} \log \frac{a_{u_i u_j} g(t_i - t_j)}{p_{ij}} \right) - \left(\mu UT + \sum_{u=1}^{U} \sum_{j=1}^{n} a_{uu_j} (G(T-t_j) - G(0)) \right) \\
&= Q(\Theta | \Theta^{(t)}),
\end{aligned}
$$

一种典型的松弛求解算法：
基于Jensen不等式优化下界
Majorization-minimization

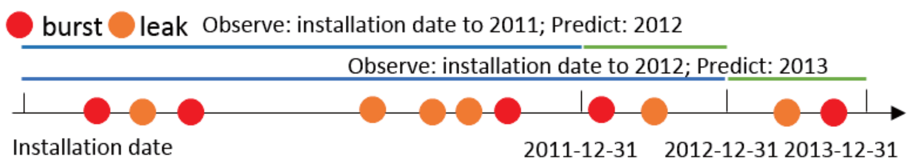
应用案例-某大型城市地下管网故障预警

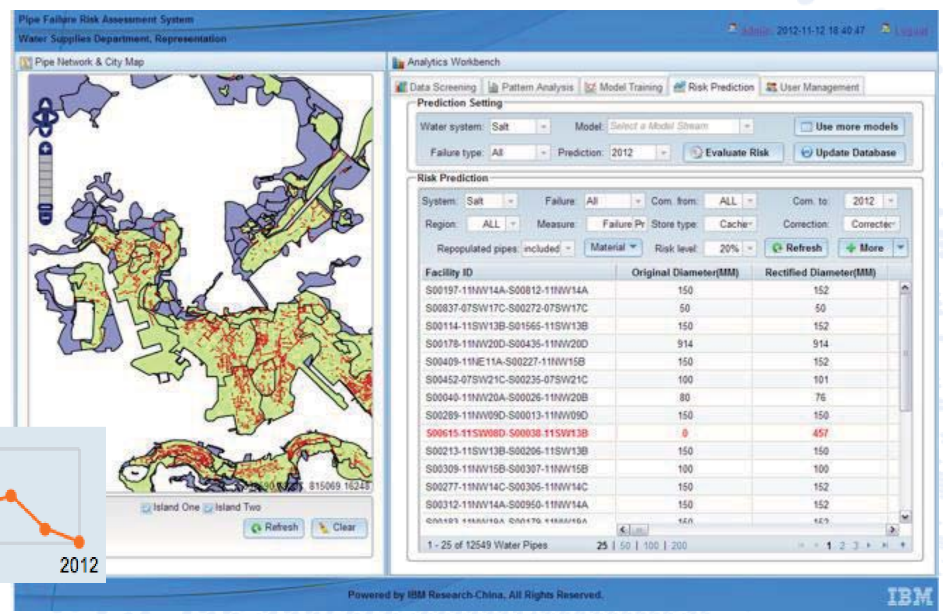
- 问题背景
  - 某大型城市地下管道超过60万，最老管道建于100年前
  - 管道一旦故障，在线维修成本高，负面影响大
  - 需要提前进行预警，提高预防性维护效率

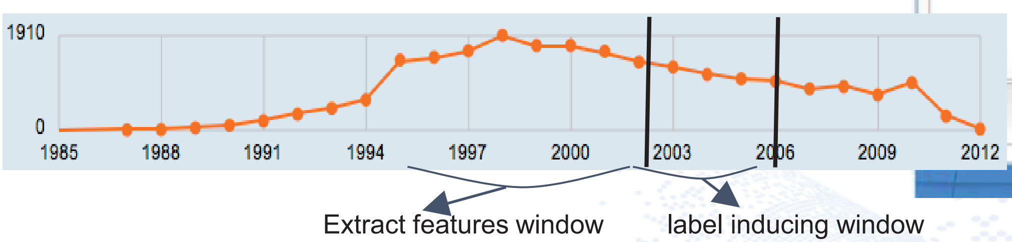


应用案例-某大型城市地下管网故障预警

- 传统预测算法
  - 分类、回归等机器学习模型
  - 训练和预测需要指定固定和统一预测窗口
  - 预测分辨率局限，历史特征信息提取局限（离散化）
- 基于点过程的方法
  - 预测窗口可以在线灵活选择
  - 历史特征细粒度提取





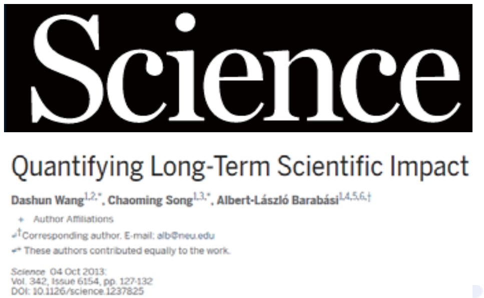


应用案例-文献引用中长期预测

- 问题背景
  - H-index等只能评价当前和过去的影响力
  - 经费分配、人才选拔需要考量未来影响力
  - 专利估值定价与引用次数紧密相关
  - 论文专利都包含自引和他引

平均误差MAPE $$\frac{1}{N}\sum_{d=1}^{N}\left|\frac{c^d(t)-r^d(t)}{r^d(t)}\right|$$

精度 $$\frac{1}{N}\sum_{d=1}^{N}\left|d:\left|\frac{c^d(t)-r^d(t)}{r^d(t)}\right|\le\epsilon\right|$$


Science

Quantifying Long-Term Scientific Impact

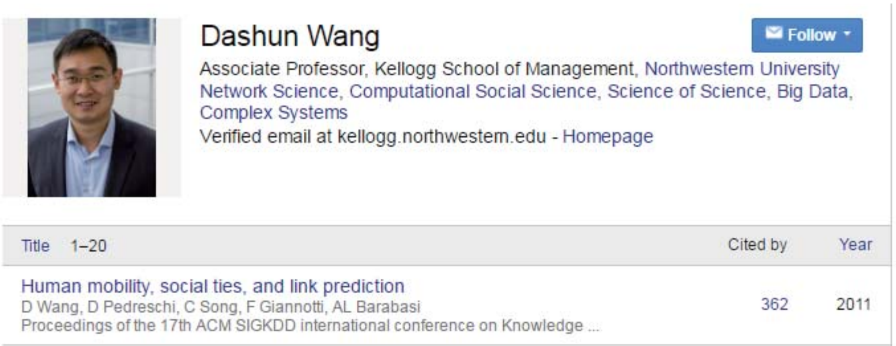
Dashun Wang[1,2,*], Chaoming Song[1,3,*], Albert-László Barabási[1,4,5,6,†]

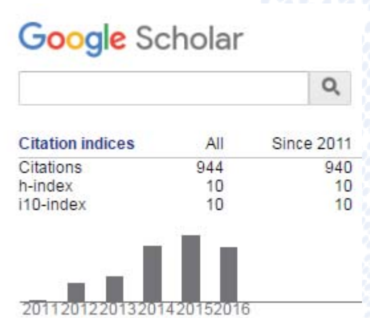
Author Affiliations

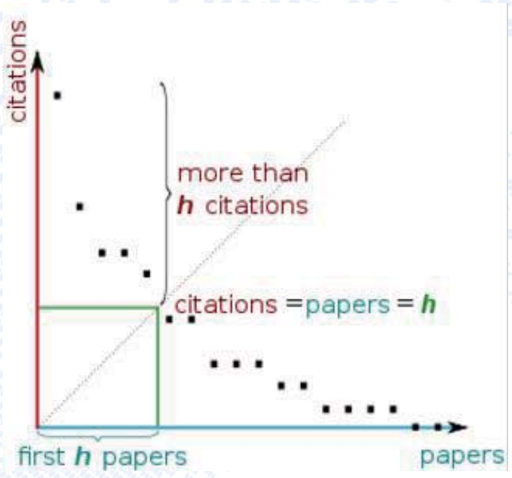
†Corresponding author. E-mail: alb@neu.edu

*These authors contributed equally to the work.

Science 04 Oct 2013:
Vol. 342, Issue 6154, pp. 127-132
DOI: 10.1126/science.1237825

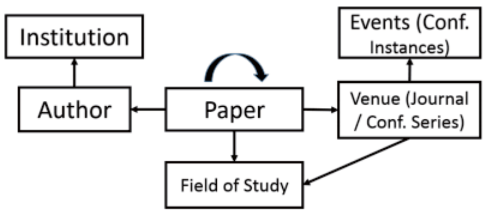


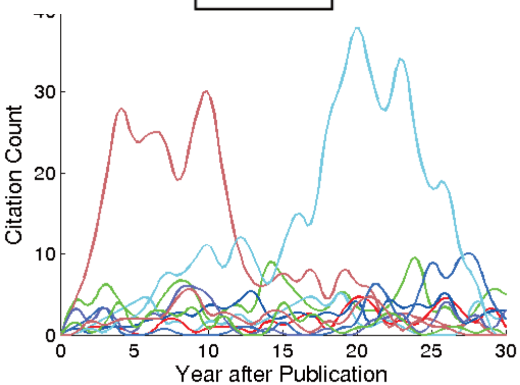


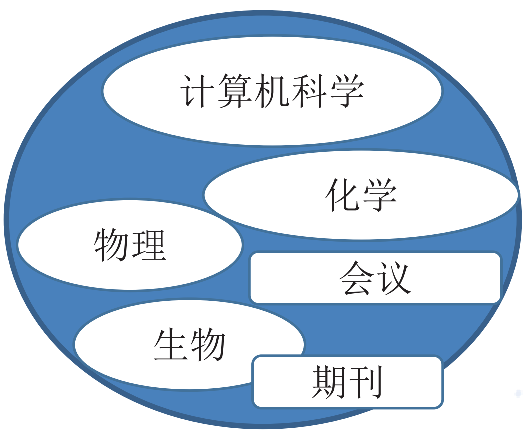


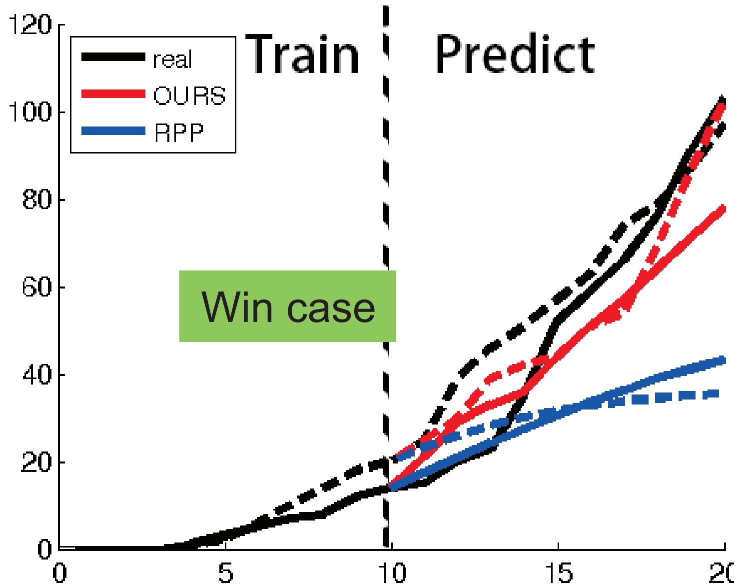
论文引用预测-与Science论文方法对比

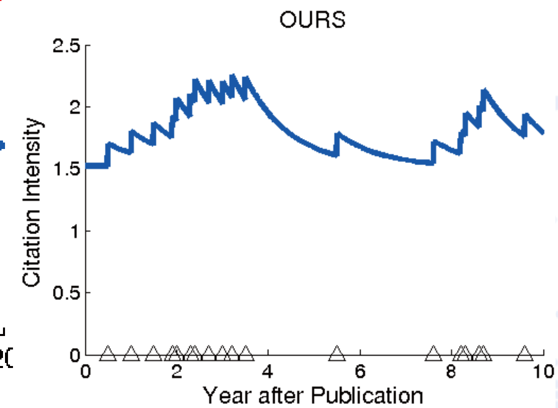
微软学术图谱数据集

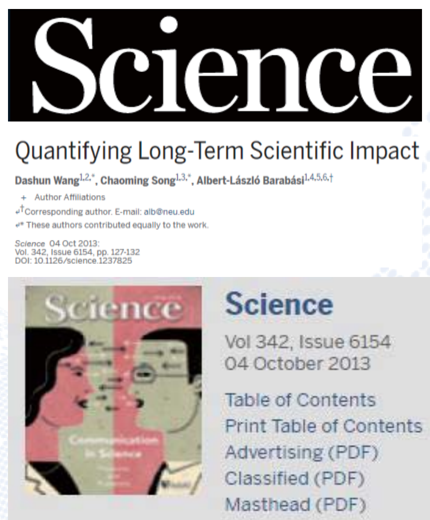











Science

Quantifying Long-Term Scientific Impact

Dashun Wang[1,2,*], Chaoming Song[1,3,*], Albert-László Barabási[1,4,5,6,†]

+ Author Affiliations

†Corresponding author. E-mail: alb@neu.edu

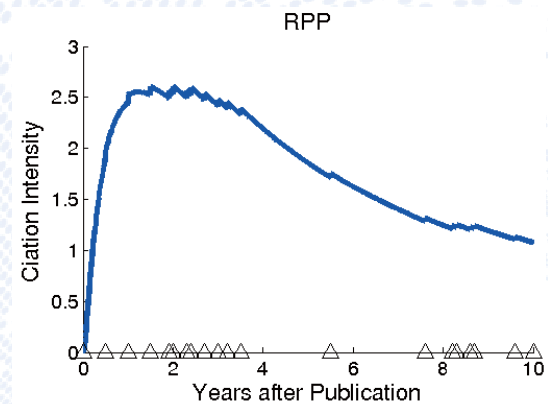
* These authors contributed equally to the work.

Science 04 Oct 2013:
Vol. 342, Issue 6154, pp. 127-132
DOI: 10.1126/science.1237825

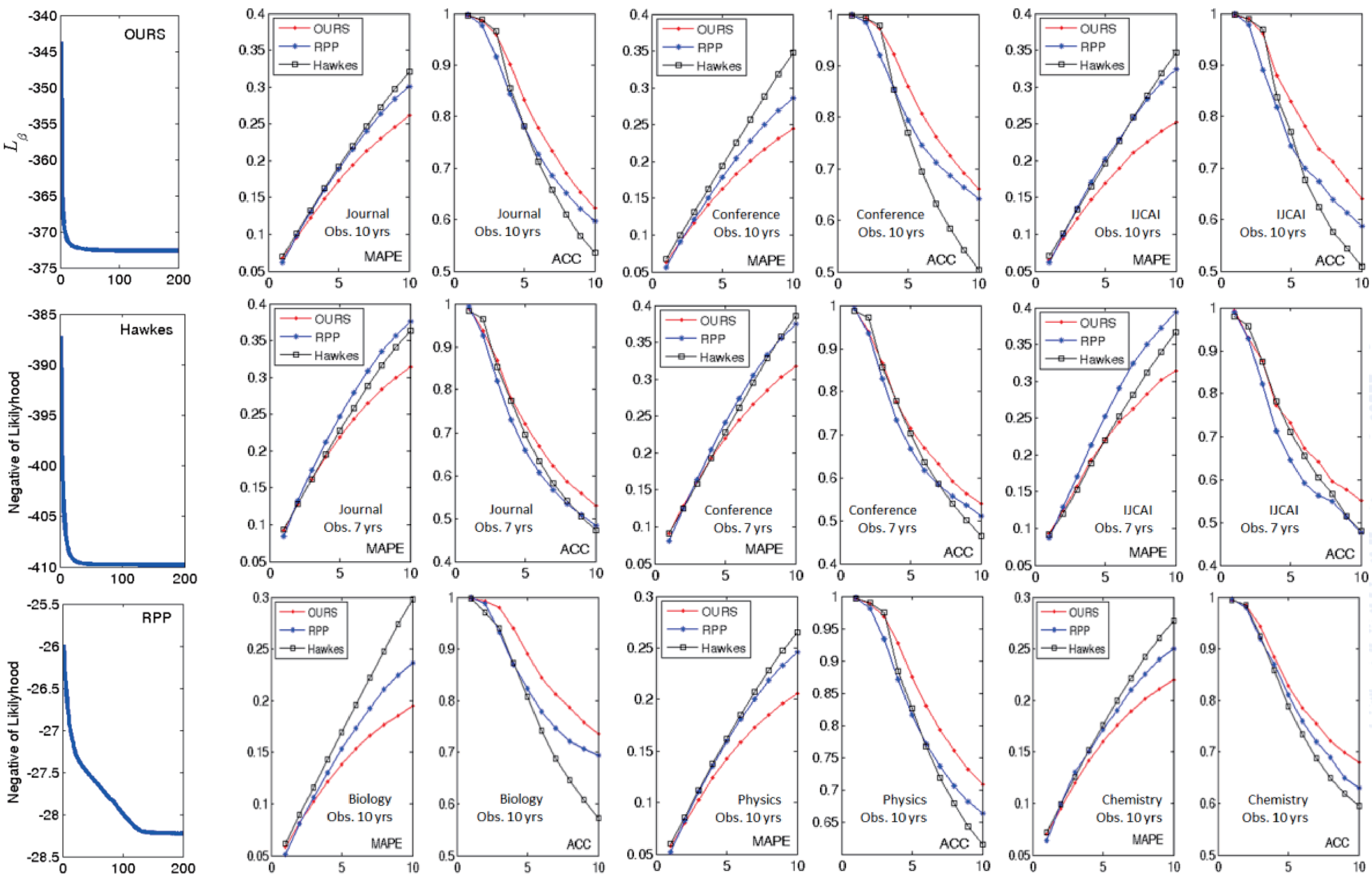
Science

Vol 342, Issue 6154
04 October 2013

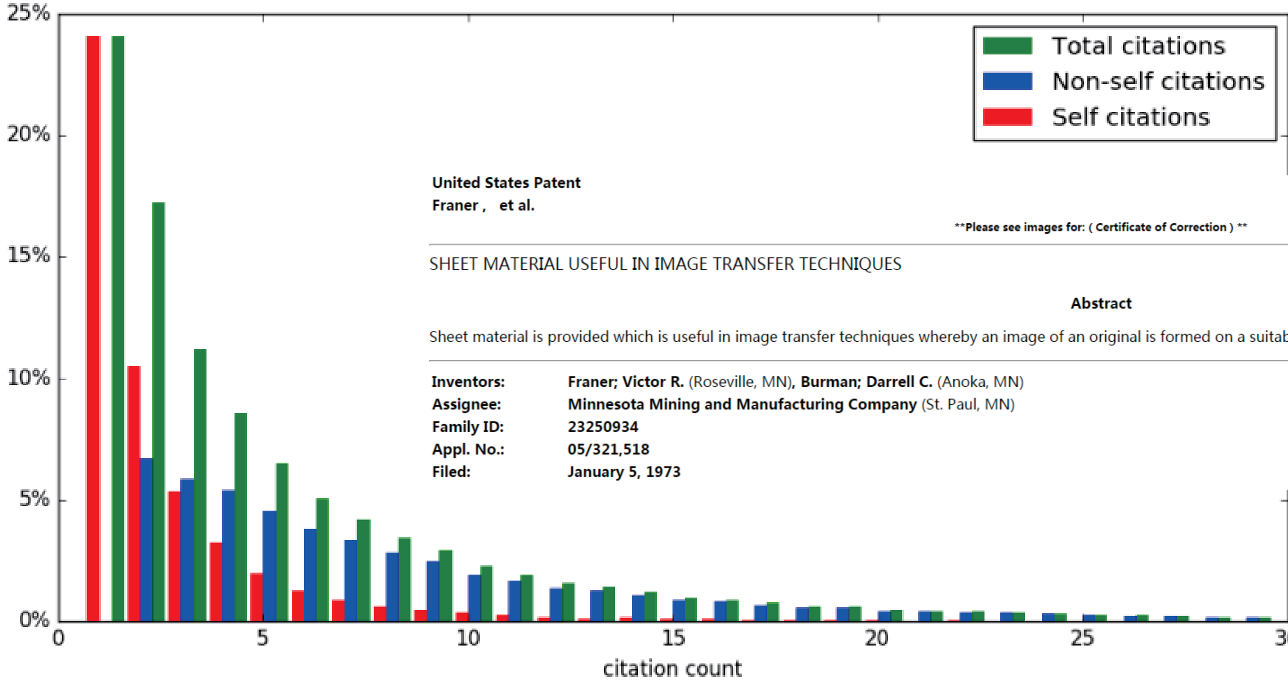
Table of Contents
Print Table of Contents
Advertising (PDF)
Classified (PDF)
Masthead (PDF)



论文引用预测-与Science论文方法对比



专利引用预测-引用类型与次数分布



United States Patent **3,859,094**
Franer , et al. **January 7, 1975**

****Please see images for: (Certificate of Correction) ****

SHEET MATERIAL USEFUL IN IMAGE TRANSFER TECHNIQUES

Abstract

Sheet material is provided which is useful in image transfer techniques whereby an image of an original is formed on a suitable receptor.

Inventors:	**Franer; Victor R.** (Roseville, MN), **Burman; Darrell C.** (Anoka, MN)
Assignee:	**Minnesota Mining and Manufacturing Company** (St. Paul, MN)
Family ID:	**23250934**
Appl. No.:	**05/321,518**
Filed:	**January 5, 1973**

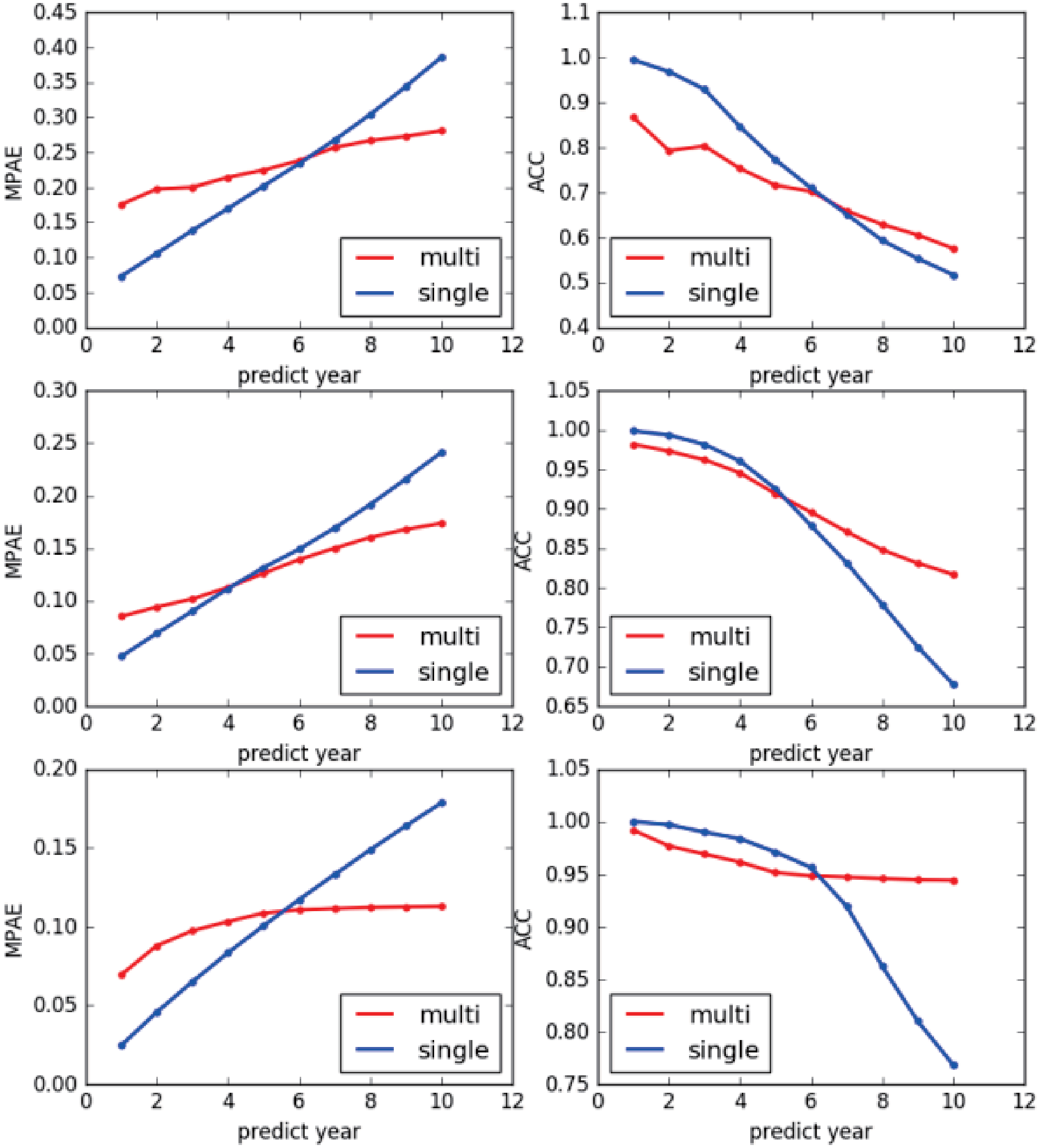
References Cited [Referenced By]

U.S. Patent Documents

3131080	April 1964	Russell
3418468	December 1968	Marx et al.
3457075	July 1969	Morgan et al.
3616015	October 1971	Kingston
3767394	October 1973	Wiese et al.

Primary Examiner: Torchin; Norman G.
Assistant Examiner: Schilling; Richard L.
Attorney, Agent or Firm: Alexander, Sell, Steldt & DeLaHunt

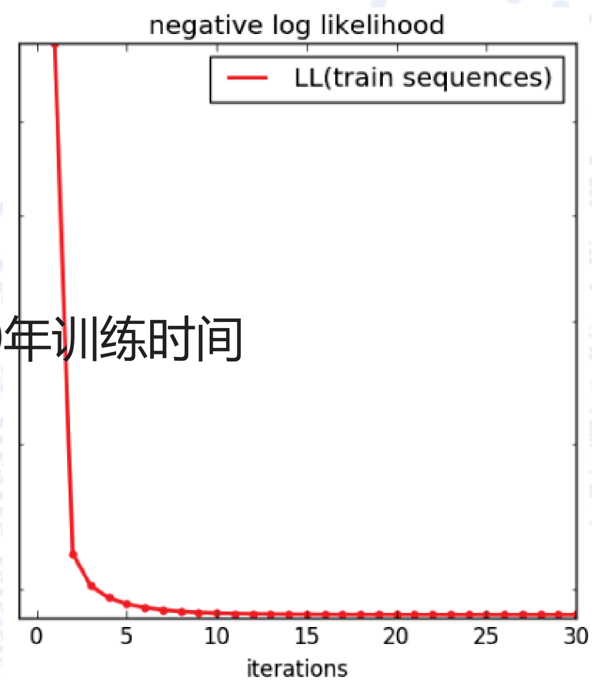
专利引用预测-引用类型敏感预测模型



固定10 年训练时间

固定15 年训练时间

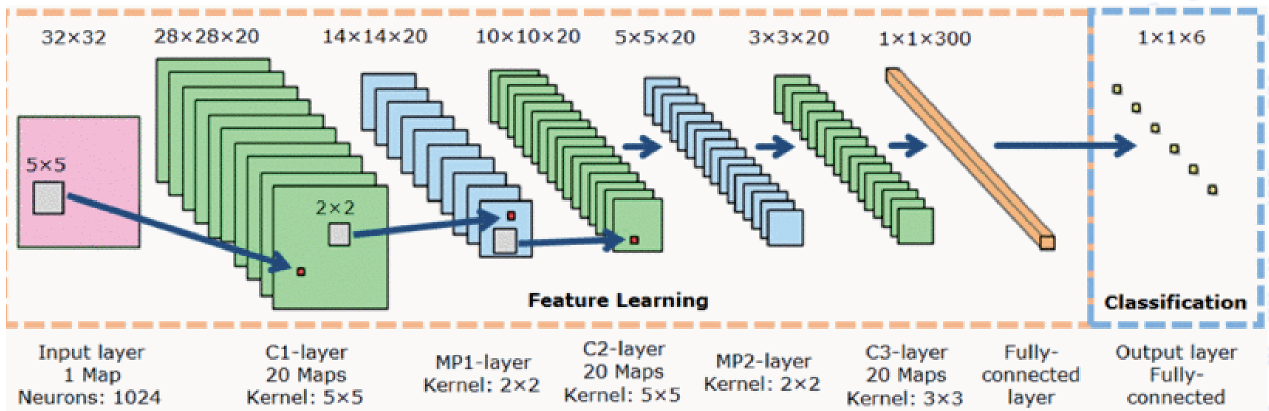
固定20年训练时间



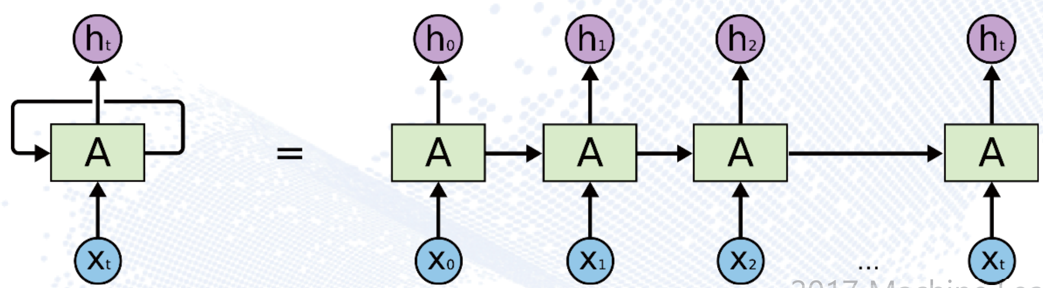
深度学习

- 卷积神经网络 (CNN)、深度置信网络 (DBN)
  - LeNet (Proc. of IEEE 1998), AlexNet (NIPS 2012) etc.
- 循环神经网络模型Recurrent Neural Network (RNN)
  - LSTM (Neural Computation 1997)
  - GRU (Gated recurrent Unit, 2014)

卷积神经网络
Convolutional
Neural Networks



循环神经网络
Recurrent Neural
Networks



对现有一些参数化白盒点过程模型的观察

- 不同类型点过程的核心往往在于对条件强度函数进行不同的参数化建模
- 再按给定的强度函数参数化形式计算最大似然估计，估计模型参数值

Model	Background	History event effect
Poisson process	$\lambda(t)$	0
Reinforced poisson process	0	$\sum_{t_i<t} \lambda(t)$
Hawkes process	$\lambda(t)$	$\sum_{t_i<t} \gamma(t, t_i)$
Reactive point process	$\lambda(t)$	$\sum_{t_i<t} \gamma_1(t, t_i) - \sum_{t_j<t} \gamma_2(t, t_i)$
Self-correcting process	0	$\exp(\mu t - \sum_{t_i<t} \alpha)$

端对端深度学习的事件序列建模

- 结合时间序列与事件序列的统一框架

显式参数化模型 → 半显式参数化模型 → 隐参数化模型

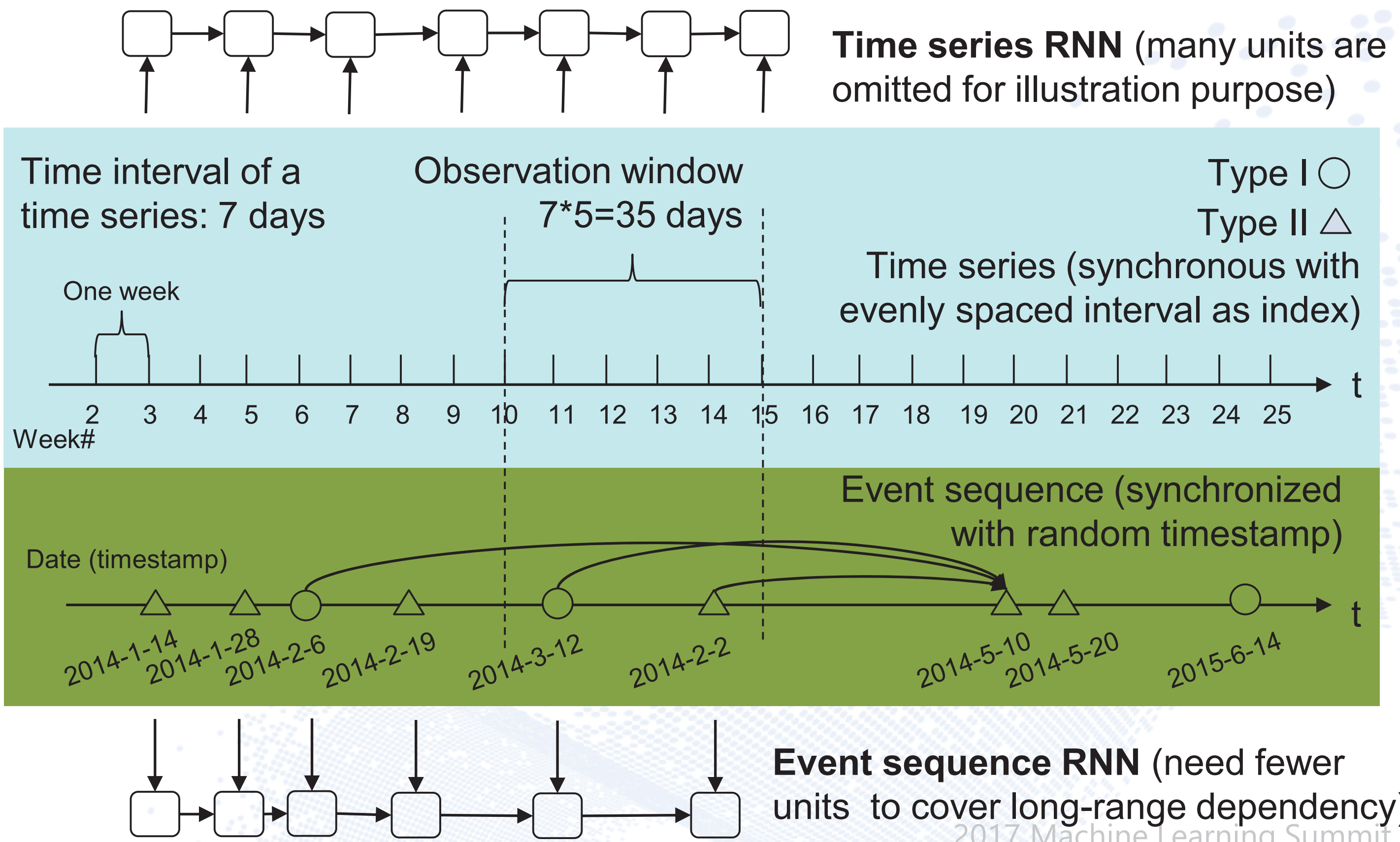
IJCAI-13 Hawkes process

KDD2016 Recurrent Marked Temporal Point Processes

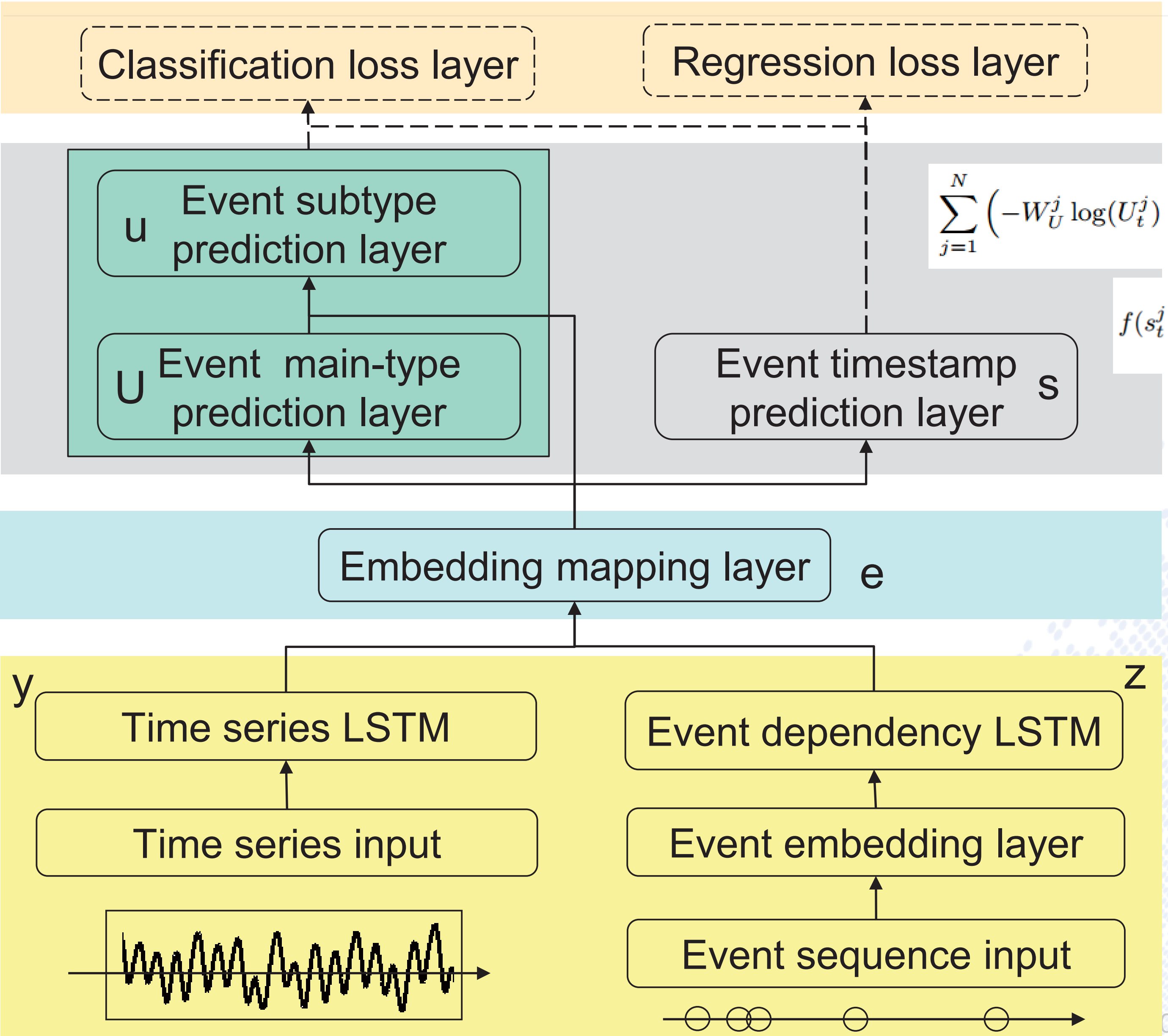
RNN process (AAAI17)

时间序列：适合刻画实时连续更新变量，如气温、CPU占用率、湿度等

事件序列：适合刻画不定期发生事件，如故障、维修等



端对端训练网络结构展示

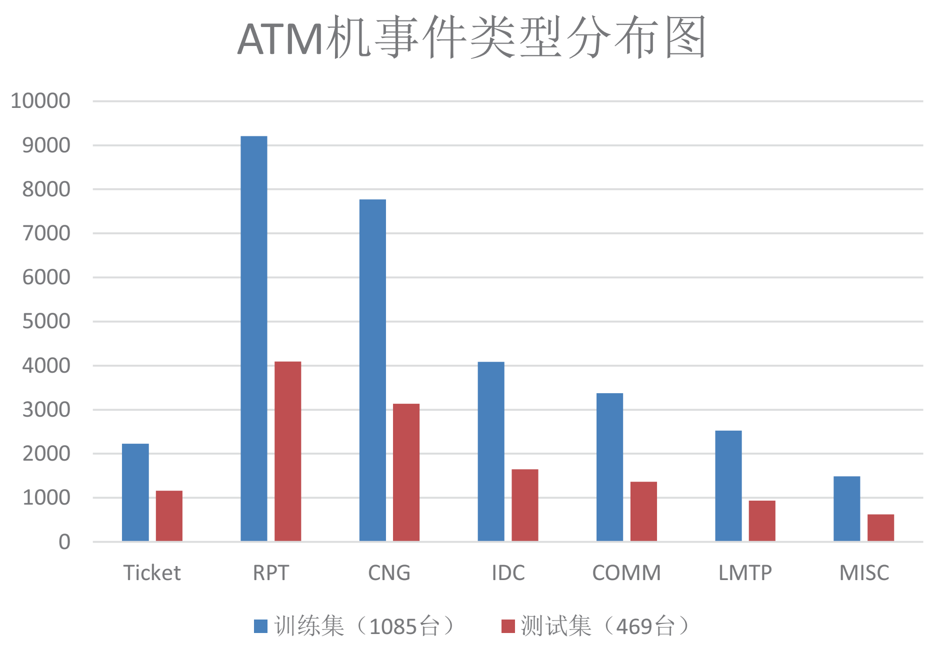


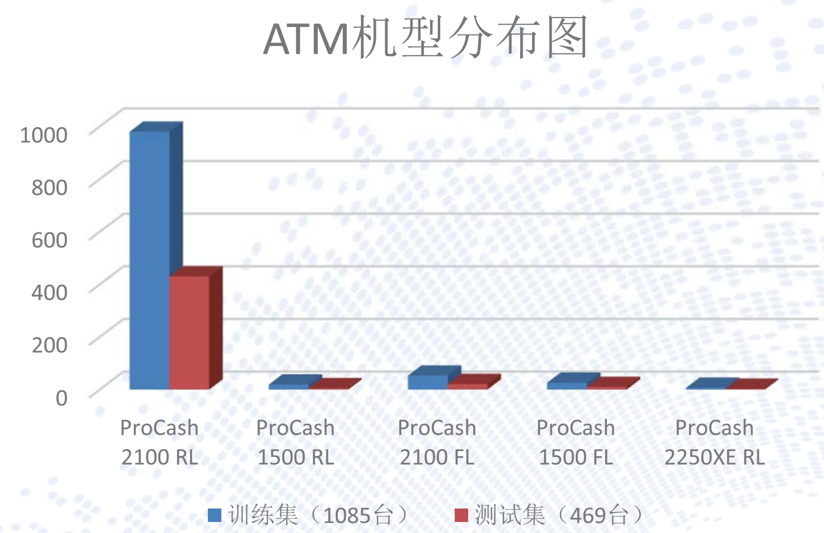
$$\sum_{j=1}^{N}\left(-W_U^j \log(U_t^j) - w_u^j \log(u_t^j) - \log\left(f(s_t^j|h_{t-1}^j)\right)\right)$$

$$f(s_t^j|h_{t-1}^j) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(s_t^j-\tilde{s}_t^j)^2}{2\sigma^2}\right)$$

应用案例-某北美银行ATM机故障预测

- 问题描述
  - ATM机频繁故障、类型繁多、历史日志记录丰富
  - 短期预测即将发生的故障时间和类型
- 数据集描述





应用案例-某北美银行ATM机故障预测

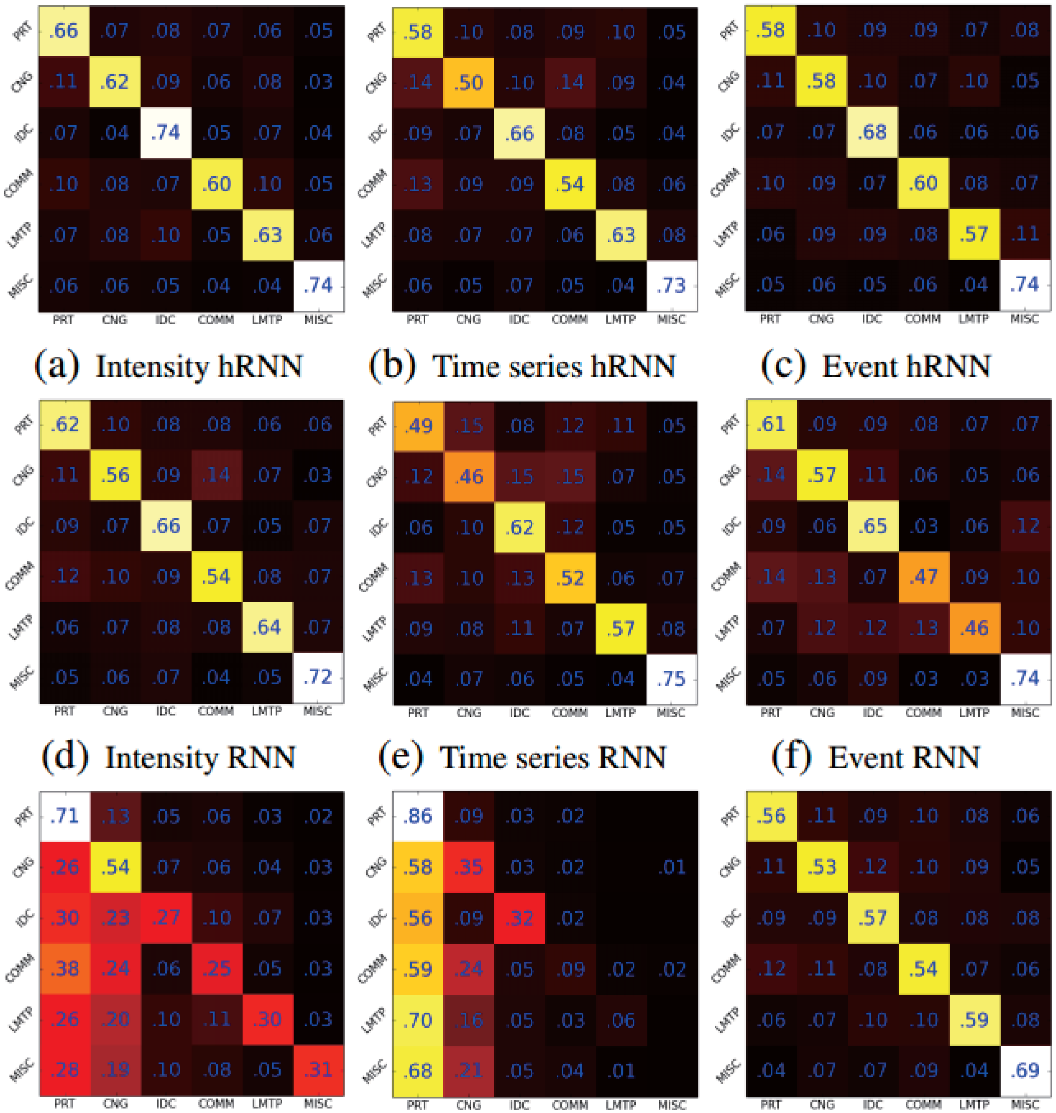
- 实验结果对比
  - 单时间序列输入
  - 单事件序列输入
  - 双序列输入

输入	模型
单时间序列输入	Time series RNN
单事件序列输入	Event sequence RNN
双序列输入	Intensity RNN

	model	main-type	subtype	hierarchical output main-type	hierarchical output subtype
precision	Time series RNN	0.673	0.554	0.582	0.590
	Event sequence RNN	0.671	0.570	0.623	0.614
	Intensity RNN	0.714	**0.620**	**0.642**	**0.664**
	Hawkes process	0.457	0.387	—	—
	Logistic prediction	**0.883**	0.385	—	—
	RMTPP	0.581	0.574	—	—
recall	Time series RNN	0.853	0.522	0.738	0.608
	Event sequence RNN	0.821	0.543	0.770	0.621
	Intensity RNN	**0.905**	**0.614**	**0.805**	**0.661**
	Hawkes process	0.493	0.394	—	—
	Logistic prediction	0.795	0.273	—	—
	RMTPP	0.691	0.583	—	—
F1 score	Time series RNN	0.707	0.533	0.571	0.605
	Event sequence RNN	0.703	0.555	0.651	0.610
	Intensity RNN	0.765	**0.616**	**0.662**	**0.663**
	Hawkes process	0.473	0.386	—	—
	Logistic prediction	**0.832**	0.269	—	—
	RMTPP	0.584	0.572	—	—

	model	main-type	subtype	hierarchical output main-type	hierarchical output subtype
MAE (in days)	Time series RNN	4.37	4.48	4.26	4.41
	Event sequence RNN	4.24	4.42	4.21	4.37
	Intensity RNN	**4.13**	**4.20**	**4.02**	**4.13**
	Hawkes process	5.26	5.46	—	—
	Logistic prediction	4.52	4.61	—	—
	RMTPP	4.28	4.32	—	—
F1 Score+	Time series RNN	0.768	0.547	0.572	0.603
	Event sequence RNN	0.705	0.597	0.639	0.646
	Intensity RNN	**0.825**	**0.661**	**0.684**	**0.708**
	Hawkes process	0.467	0.451	—	—
	Logistic prediction	0.846	0.286	—	—
	RMTPP	0.584	0.619	—	—
MAE+	Time series RNN	4.21	3.78	**4.05**	**3.97**
	Event sequence RNN	4.16	3.84	4.12	4.01
	Intensity RNN	**4.12**	**3.57**	4.21	4.11
	Hawkes process	5.42	3.93	—	—
	Logistic prediction	4.5	4.24	—	—
	RMTPP	4.26	3.99	—	—

应用案例-某北美银行ATM机故障预测

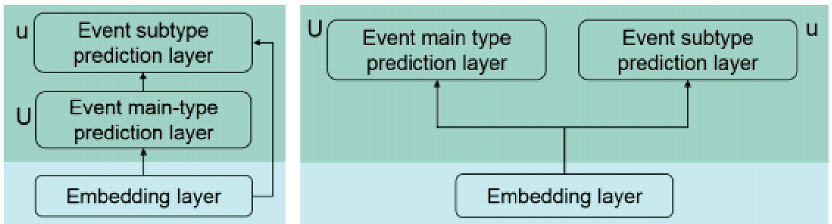


(a) Intensity hRNN (b) Time series hRNN (c) Event hRNN

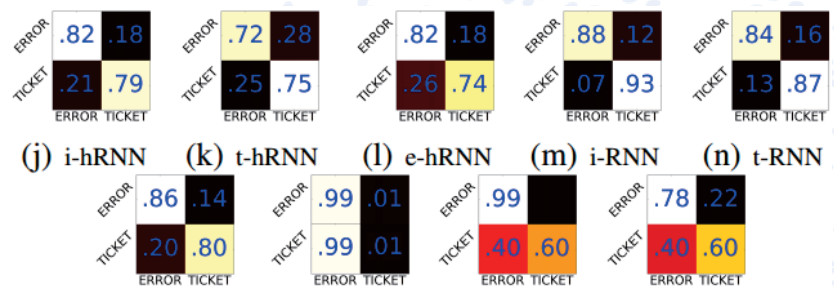
(d) Intensity RNN (e) Time series RNN (f) Event RNN

(g) Hawkes process (h) Logistic (i) RMTPP

子类分类混淆矩阵



(a) Hierarchical layer. (b) Flat independent layer.



(j) i-hRNN (k) t-hRNN (l) e-hRNN (m) i-RNN (n) t-RNN

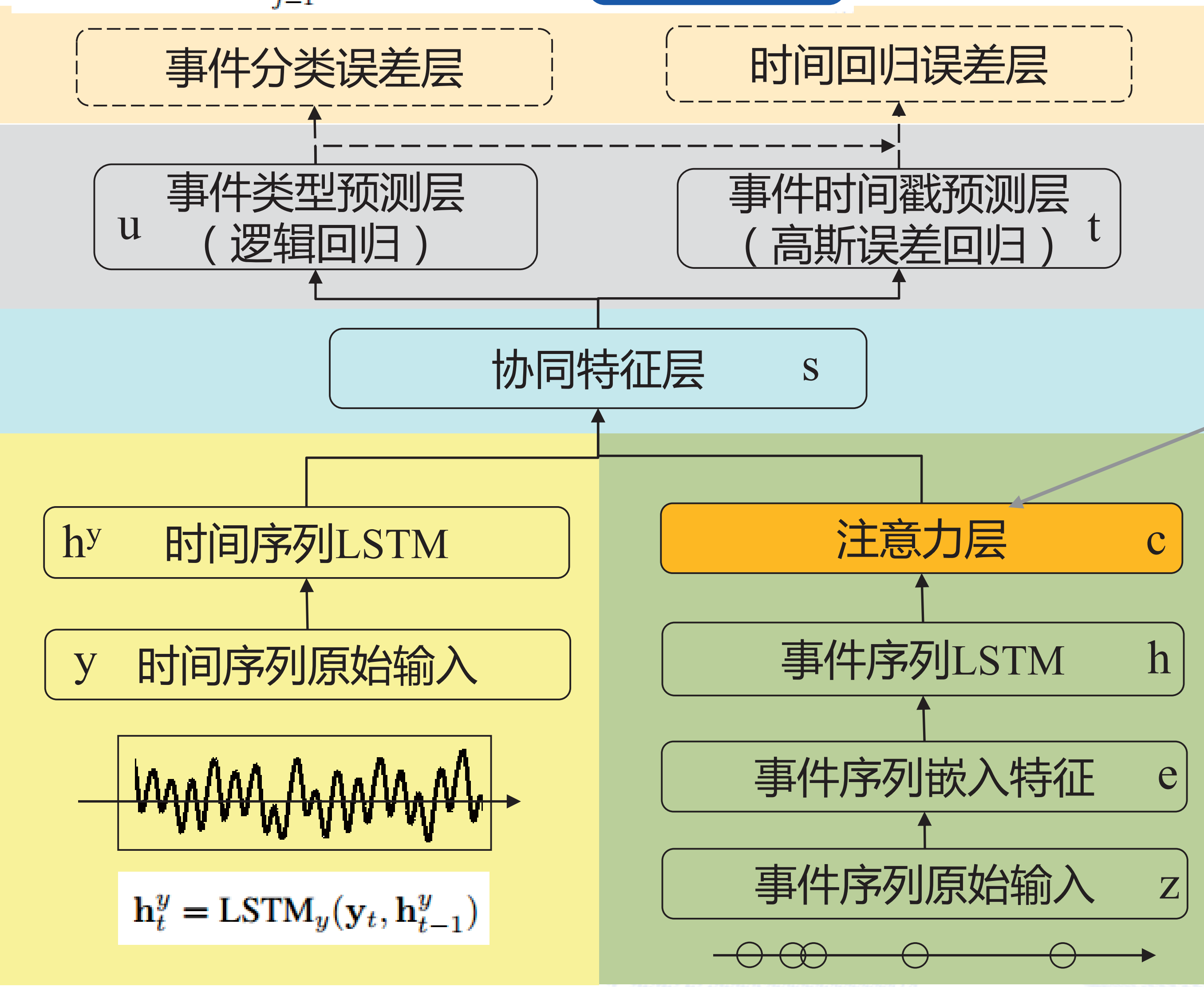
(o) e-RNN (p) Hawkes (q) Logistic (r) RMTPP

子类分类混淆矩阵

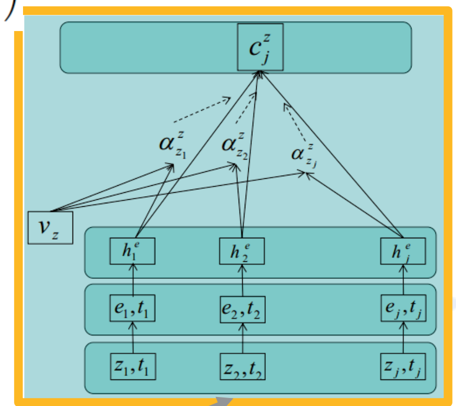
端对端训练网络结构展示-下次事件预测

$$\mathcal{L}(\{z_i,t_i\}_{i=1}^N) = \sum_{j=1}^{N-1}\{\mathbf{b}_{z_{j+1}}\log(\mathbf{u}_{j+1}^{z_{j+1}}) + \boxed{\log\left(f(t_{j+1}|\mathcal{H}_{t_j})\right)}\}$$

$$f(t_{j+1}|\mathcal{H}_{t_j}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(t_{j+1}-t_{j+1}^*)^2}{2\sigma^2}\right)$$



$$\mathbf{h}_t^y = \mathrm{LSTM}_y(\mathbf{y}_t, \mathbf{h}_{t-1}^y)$$



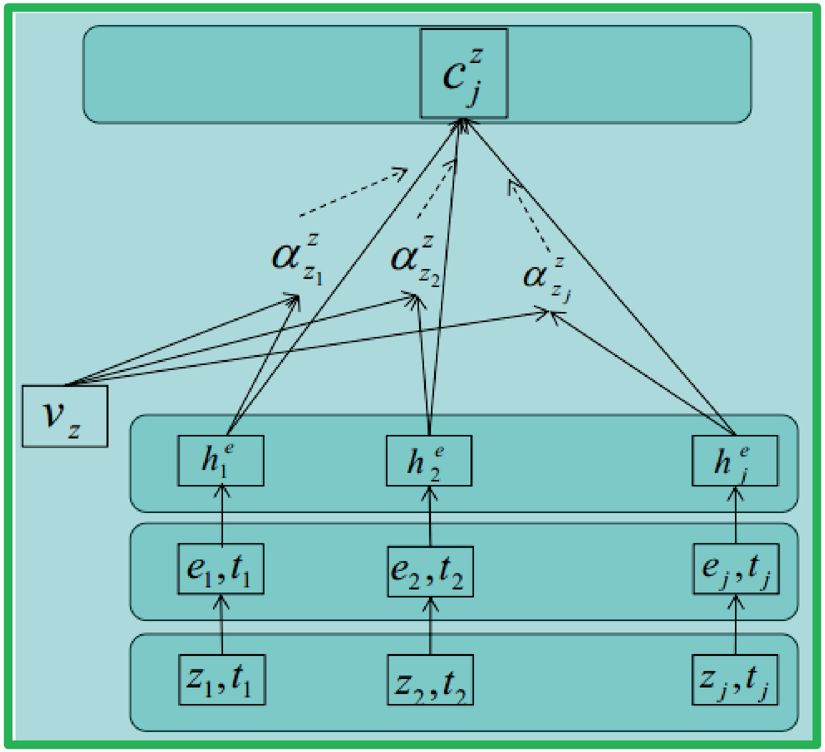
$$\alpha_{z_i}^z = f_{att}(\mathbf{h}_i, \mathbf{v}_z)$$

$$\mathbf{c}_j^z = \sum_{i=1}^{S} \alpha_{z_i}^z \mathbf{h}_i$$

$$\mathbf{e}_i = \mathbf{W}_{em}\mathbf{z}_i,$$

$$\mathbf{h}_i^e = \mathrm{LSTM}_z(\{\mathbf{e}_i, \mathbf{t}_i\}, \mathbf{h}_{i-1}^e),$$

用注意力机制来实现模型可解释性



事件序列: 类型A 类型C 类型D

$$\mathbf{c}_j^z = \sum_{i=1}^{S} \alpha_{z_i}^z \mathbf{h}_i$$

$$\alpha_{z_i}^z = f_{att}(\mathbf{h}_i, \mathbf{v}_z)$$

$$f_{att}(\mathbf{h}_i^e, \mathbf{v}_z) = \begin{cases} 0, & \text{if } |\tanh(\mathbf{h}_i^e * \mathbf{v}_z)| < \epsilon \\ |\tanh(\mathbf{h}_i^e * \mathbf{v}_z)|, & \text{otherwise} \end{cases}$$

- 每个epoch，每个训练序列样本可以计算一次参数A矩阵
- 每个epoch，参数A矩阵更新为所有训练样本更新计算结果之和
- 训练迭代结束后，得到最终A矩阵作为Infectivity Matrix，衡量不同类型事件间的相关度

✓ 提升下次事件预测精度

✓ 实现潜在关系定量挖掘

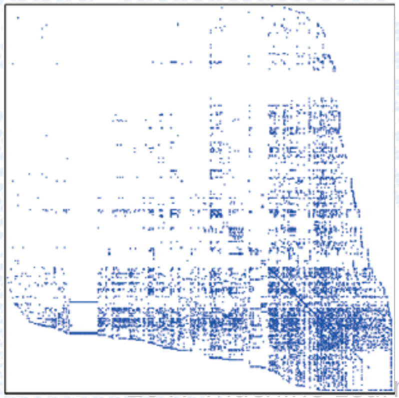
从点过程建模视角回顾黑盒模型的联系

- 时间序列用于刻画基准背景项的瞬时变动
- 事件序列用于刻画历史对当前的影响
- 注意力机制建模不同类型事件间的相互影响

Model	Background	History event effect
Poisson process	$\lambda(t)$	0
Reinforced poisson process	0	$\sum_{t_i<t}\lambda(t)$
Hawkes process	$\lambda(t)$	$\sum_{t_i<t}\gamma(t,t_i)$
Reactive point process	$\lambda(t)$	$\sum_{t_i<t}\gamma_1(t,t_i)-\sum_{t_j<t}\gamma_2(t,t_i)$
Self-correcting process	0	$\exp(\mu t-\sum_{t_i<t}\alpha)$

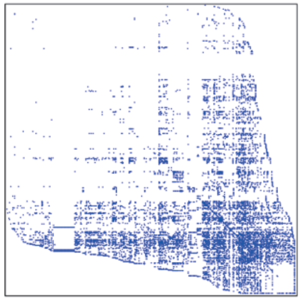
Hawkes过程（事件类型间的激励矩阵A）

$$\lambda_d = \mu_d(t) + \sum_{i:t_i<t} \gamma_{d_i d}(t-t_i)$$
$$= \mu_d(t) + \sum_{i:t_i<t} a_{d_i d}\exp(-w(t-t_i)),$$



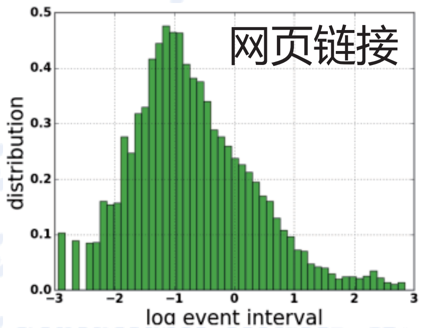
深度事件序列学习应用案例

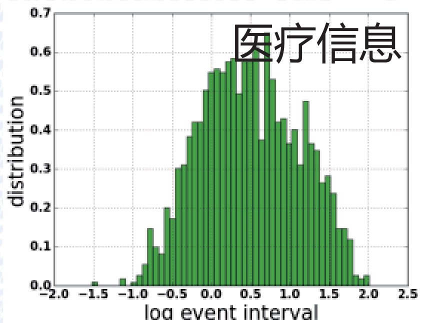
- 仿真测试设置
  - Hawkes模拟出5000个事件序列，包含20类事件
  - 基准项、激励项分别采样自U(0,0.01)、U(0,0.1)均匀分布，其中半数随机置零以满足稀疏性
  - 时间序列则按基准项加上高斯噪声G(0,0.001)



影响矩阵真值

- 预防性维保
  - 问题背景：设备日志分析与故障事件预测
  - 数据来源：某北美银行ATM机两年日志记录
- 网页链接分析
  - 数据来源：MemeTrack
  - http://memetracker.org
- 医疗信息挖掘
  - 数据来源：MIMIC II
  - https://mimic.physionet.org

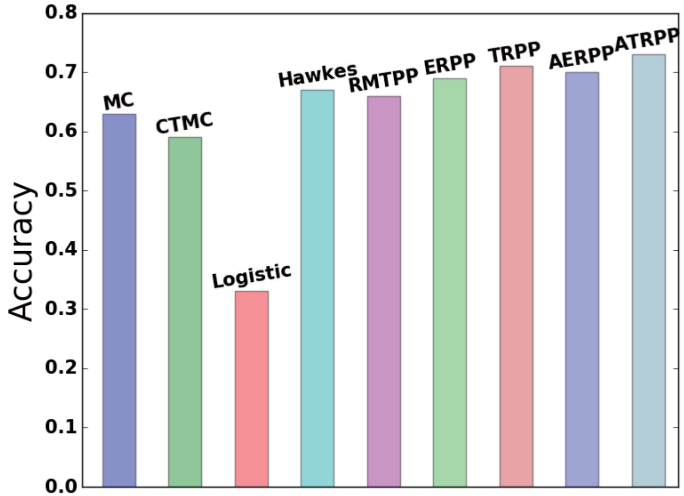




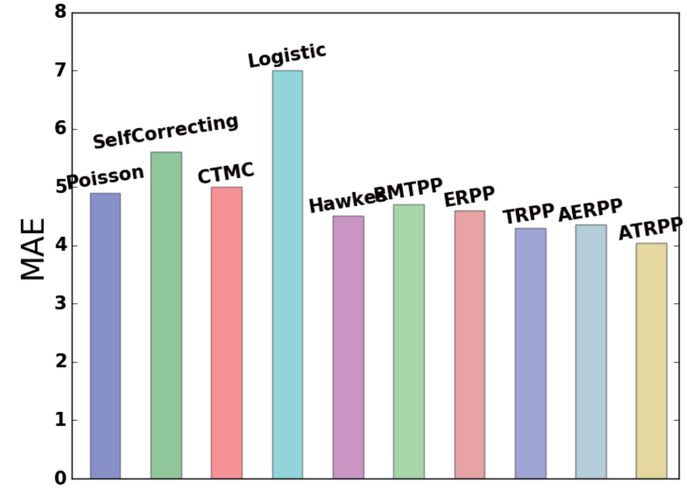
时间间隔分布

仿真数据测试

- ✓ 马尔科夫链
- ✓ LR回归
- ✓ 点过程建模
- ✓ 深度过程学习



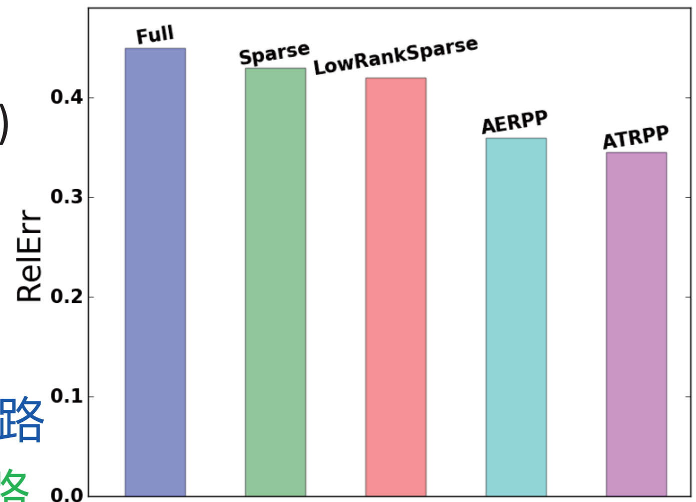
越大越好



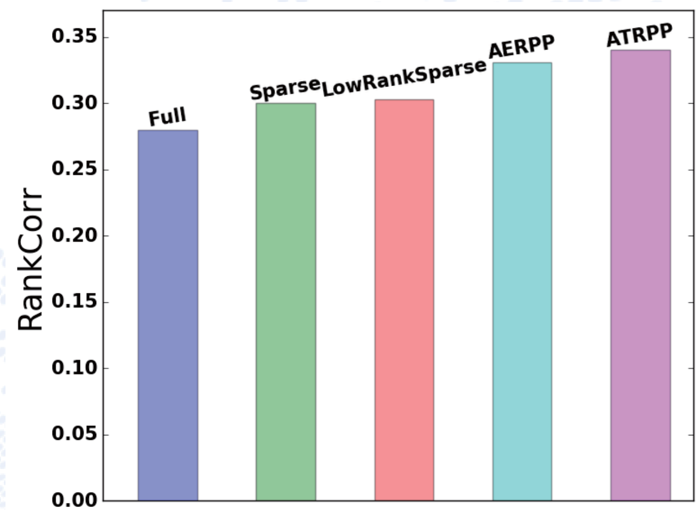
越小越好

事件预测精度

- ✓ Hawkes过程影响矩阵正则化约束
- Full(无正则化约束)
- Sparse
- LowRankSparse
- ✓ 深度过程学习
- AERPP Event单路
- ATRPP Twin双路 Attentional



越小越好

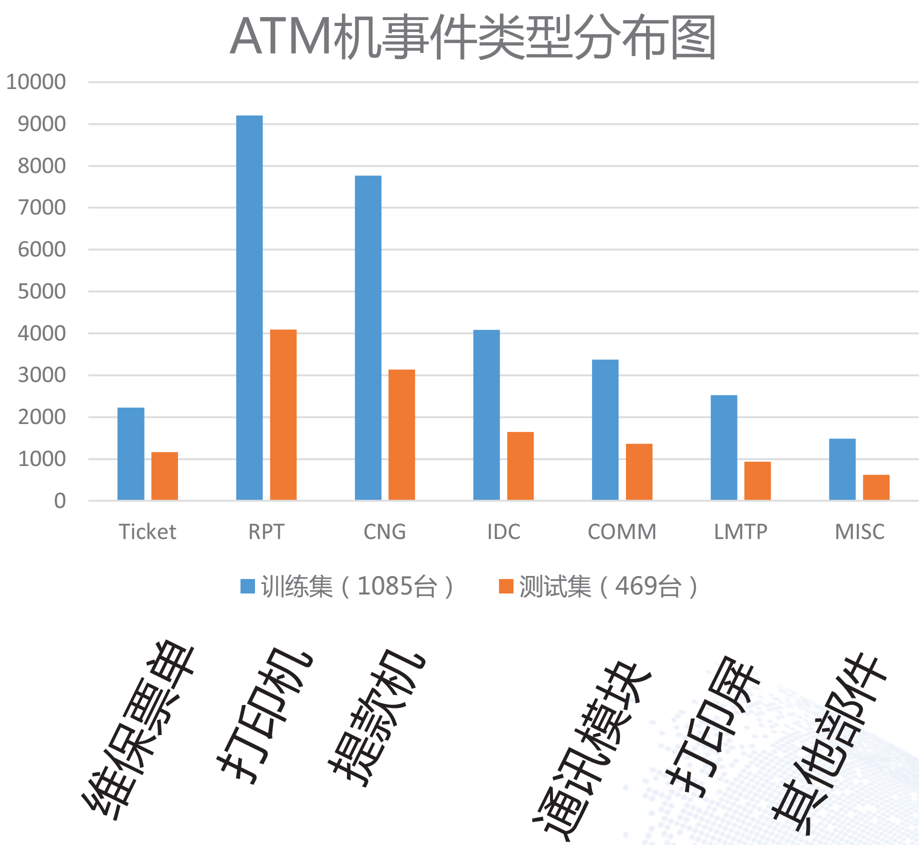


越大越好

挖掘(影响矩阵) 精度

深度事件序列学习应用-预防性维保与故障分析

- ATM机频繁故障、类型繁多、历史日志记录丰富
  - 短期预测即将发生的故障时间和类型
  - 定量挖掘不同类型事件间的关联



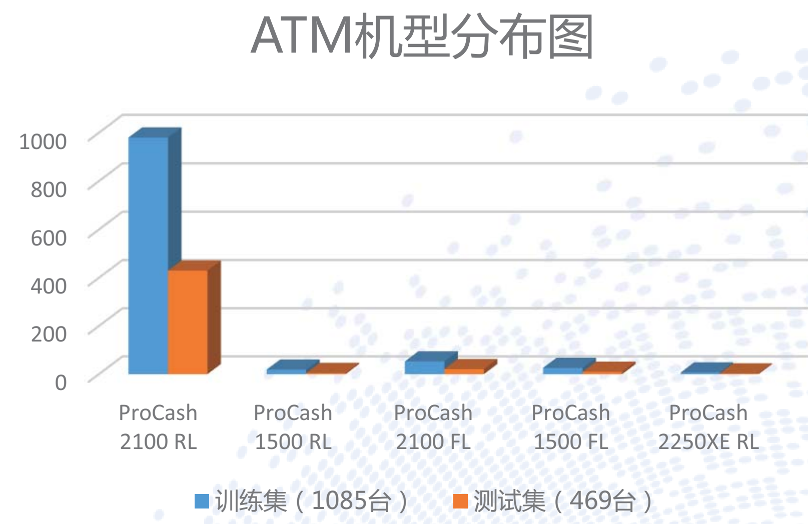


TABLE II: Statistics of event count per ATM, and timestamp interval in days for all ATMs (in brackets).

type	total	max	min	mean	std
TIKT	2226(–)	10(137.04)	0(1.21)	2.09(31.70)	1.85(25.14)
PRT	9204(–)	88(210.13)	0(0.10)	8.64(12.12)	11.37(21.41)
CNG	7767(–)	50(200.07)	0(0.10)	7.29(15.49)	6.59(23.87)
IDC	4082(–)	116(206.61)	0(0.10)	3.83(23.85)	5.84(30.71)
COMM	3371(–)	47(202.79)	0(0.10)	3.16(22.35)	3.90(29.36)
LMTP	2525(–)	81(207.93)	0(0.10)	2.37(22.86)	4.41(34.56)
MISC	1485(–)	32(204.41)	0(0.10)	1.39(24.27)	2.54(34.38)

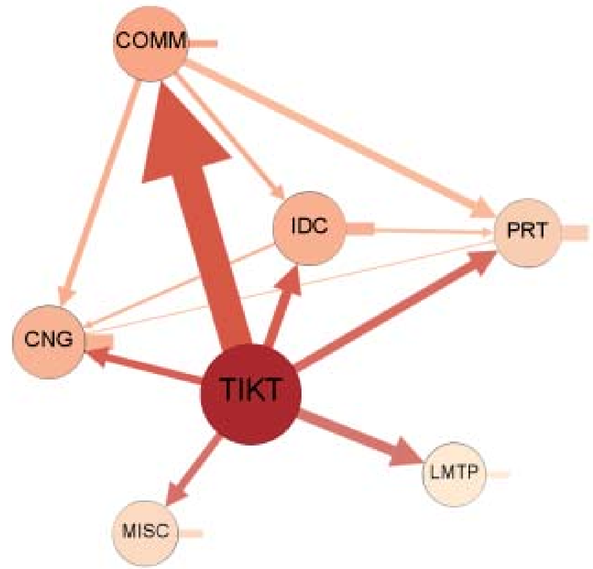
深度事件序列学习应用-预防性维保与故障分析

挖掘结果：

加权图（无定量真值）

在挖掘得到关系图基础上

进行社区检测得到的子图



预测结果

KDD16

Twin双路

Event单路

Attentional

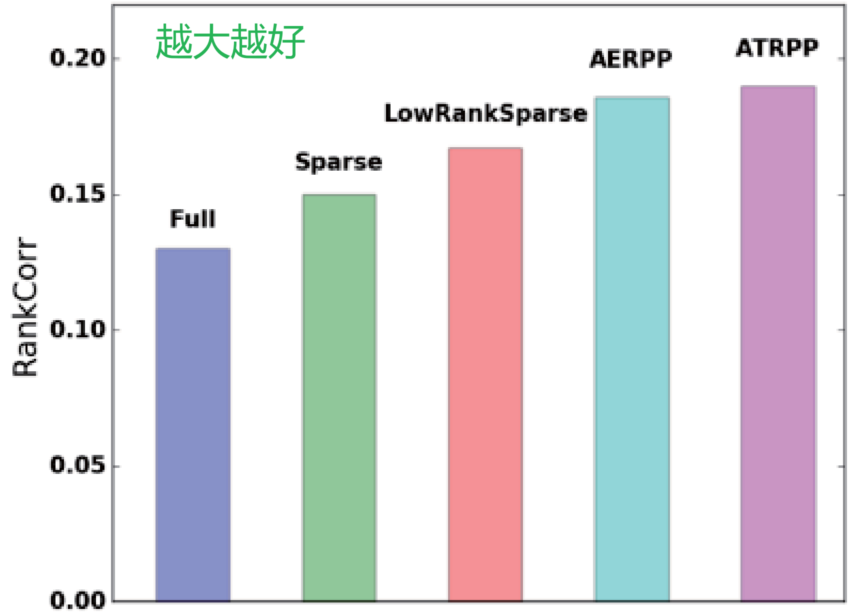
model	precision	recall	F1 score	MAE
Poisson	—	—	—	4.76
SelfCorrecting	—	—	—	4.65
Markov Chain	0.530	0.591	0.545	—
CTMC	0.516	0.554	0.503	5.16
Logistic	0.428	0.375	0.367	4.51
Hawkes	0.459	0.514	0.495	5.43
RMTPP	0.587	0.640	0.607	4.31
TRPP	0.607	0.661	0.626	4.18
ERPP	0.559	0.639	0.599	4.37
ATRPP	**0.615**	**0.688**	**0.634**	**3.92**
AERPP	0.599	0.672	0.617	3.98

深度事件序列学习应用-社交媒体分析

TABLE IV: Prediction evaluation by accuracy and MAE (mean absolute error) on MemeTracker dataset.

model	accuracy@10	accuracy@5	MAE
Poisson	—	—	1.63
SelfCorrecting	—	—	1.70
Markov Chain	0.563	0.472	—
CTMC	0.513	0.453	1.69
Logistic	0.463	0.416	1.72
Hawkes	0.623	0.563	1.68
RMTPP	0.679	0.589	1.55
TRPP	0.681	0.592	1.52
ERPP	0.673	0.586	1.56
ATRPP	**0.694**	**0.598**	**1.43**
AERPP	0.678	0.589	1.45

预测结果对比

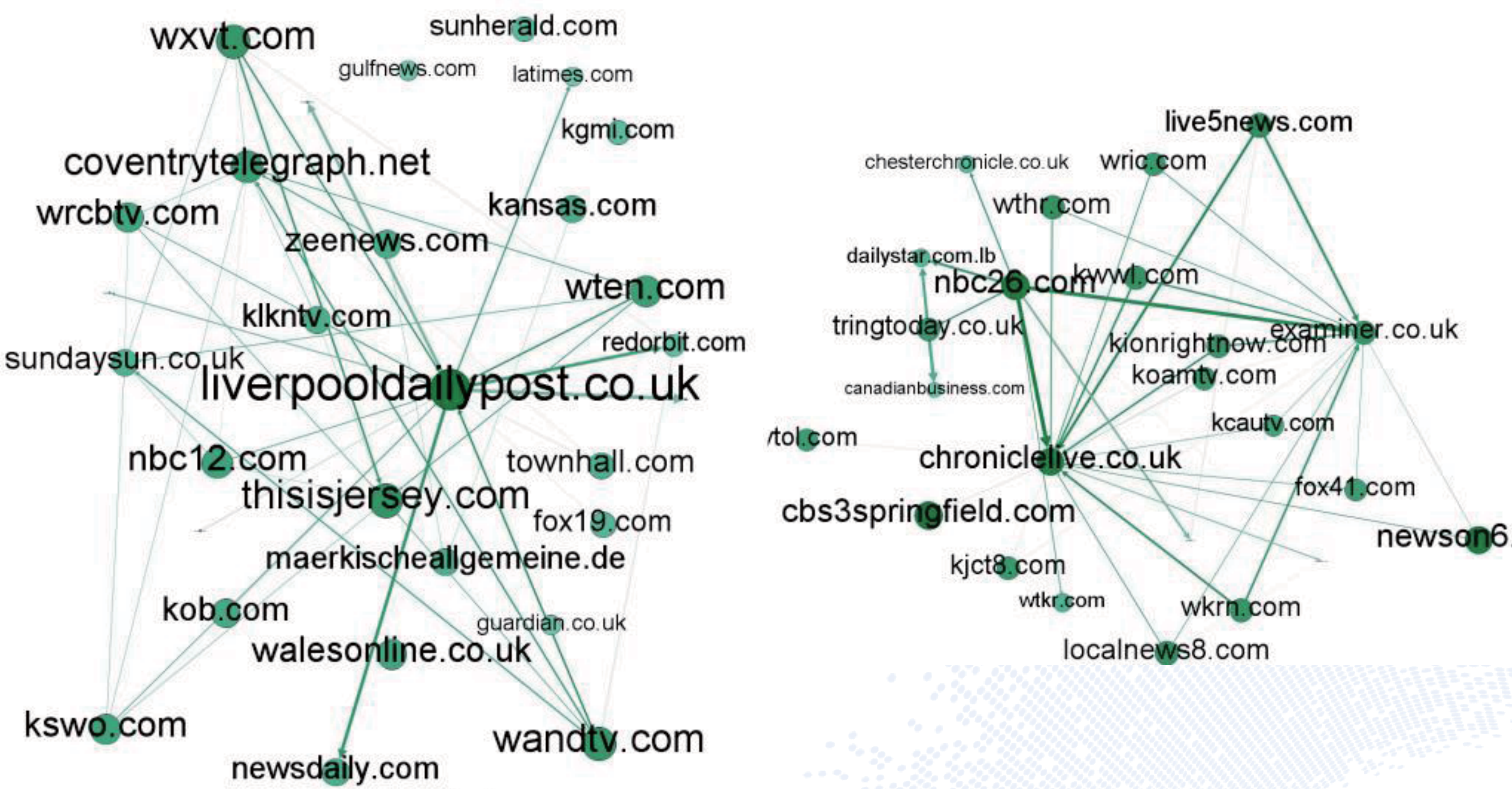


挖掘结果对比（带近似真值，采用文献[1]的真值计算协议）

数据来源	MemeTracker数据集，1.72亿在线新闻或者博客文章，超过1百万消息传播链（meme cascade），以某文章提及某meme记为一次传播事件
实验数据	选择文章数最多的前500个站点（节点），研究其上传播的3100万cascade
事件序列	以某个meme为主题的传播链条cascade
时间序列	该meme在500个站点每小时被提及次数

[1] M. Gomez Rodriguez, J. Leskovec, and A. Krause, Inferring networks of diffusion and influence, SIGKDD 2010

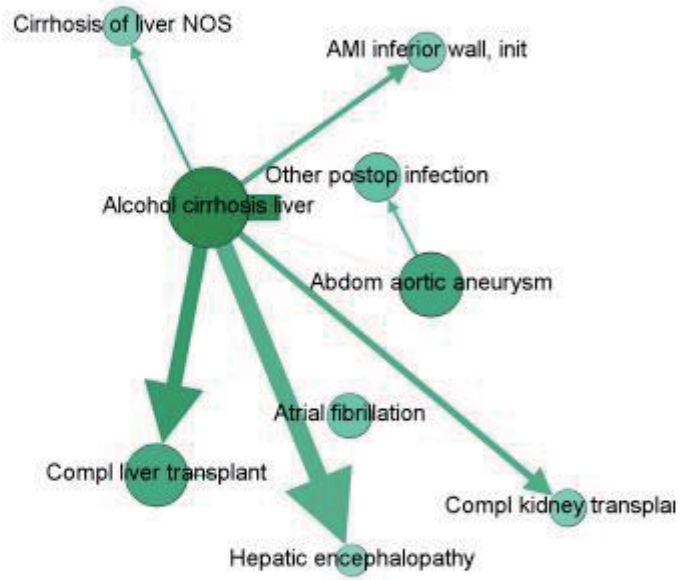
深度事件序列学习应用-社交媒体分析



挖掘得到的加权网络社区图

（在挖掘得到关系图基础上进行社区检测得到的子图）

深度事件序列学习应用-医疗电子记录分析



(a) Liver Diseases

肝病中Diagnose Code网络关系图（无真值）
（在挖掘得到关系图基础上进行社区检测得到的子图）

TABLE V: Prediction evaluation on the MIMIC dataset.

	model	accuracy	MAE
	Poisson	——	0.562
	SelfCorrecting	——	0.579
	Markov Chain	77.53%	——
	CTMC	73.62%	0.583
	Logistic	69.36%	0.643
	Hawkes	78.37%	0.517
KDD16	RMTPP	82.52%	0.546
Twin双路	TRPP	82.26%	0.513
Event单路	ERPP	78.23%	0.521
Attentional	ATRPP	**85.23%**	**0.497**
	AERPP	83.96%	0.503

数据来源	MIMIC-III (Medical Information Mart for Intensive Care III) 40000病人（匿名），跨2001-2012年
实验数据	采用其中门诊做多的937名病人门诊序列，600训练，100验证，237测试
时间序列	病人年龄、体重、心率、血压
事件序列	门诊事件序列（Diagnose Code）

基于对抗学习的事件序列建模-WGANTPP

$$\boxed{\min_{\theta} \max_{w\in\mathcal{W},\|f_w\|_L\le 1}} \frac{1}{L}\sum_{l=1}^{L} \boxed{f_w}(\xi_l) - \sum_{l=1}^{L} f_w(\boxed{g_\theta}(\zeta_l)) - \nu \sum_{l,m=1}^{L} \boxed{\left| \frac{|f_w(\xi_l) - f_w(g_\theta(\zeta_m))|}{\boxed{|\xi_l - g_\theta(\zeta_m)|_\star}} - 1 \right|}$$

min-max对抗　　判别器　　生成器　　正则：WGAN中1-Lipschitz 条件

两个事件序列间的距离

$$\|\xi - \rho\|_\star = \min_\sigma \sum_{i=1}^{n} \|x_i - y_{\sigma(i)}\|_\circ$$

$$\|\xi - \rho\|_\star = \min_{\sigma} \sum_{i=1}^{n} \|x_i - y_{\sigma(i)}\|_\circ + \sum_{i=n+1}^{m} \|s - \tau_{\sigma(i)}\|,$$

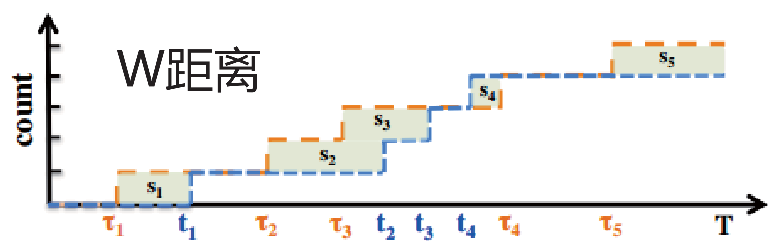
$$\|\xi - \rho\|_\star = \sum_{i=1}^{n} |t_i - \tau_i| + (m-n) \times T - \sum_{i+1}^{m} y_{i+1},$$

产生器 Generator g	
模型形式	RNN
输入	真实事件序列
输出	条件强度函数
后续	采样输出序列

输出的条件强度函数可以通过MLE进行模型训练，而我们则采取GAN框架训练

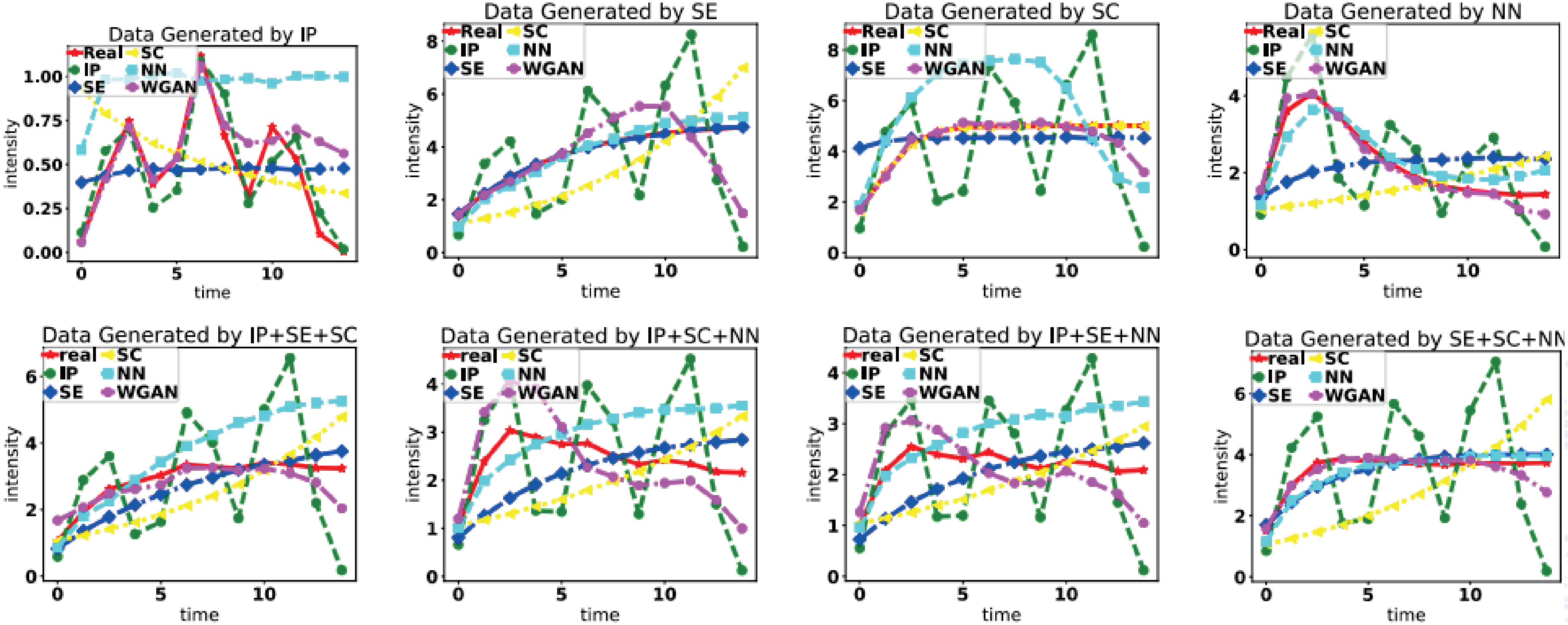
判别器 Critic f	
模型形式	RNN
输入	事件序列
输出	标量代表标记

基于强度函数采样，强度函数为生成器的输出：



b) $\|\cdot\|_\star$ distance between sequences

模型拟合结果(经验概率强度函数)-仿真数据



不同点过程模型生成数据：

- **IP**: Inhomogeneous process
- **SE**: Self-exciting process
- **SC**: Self-correcting process
- **NN**: Recurrent Neural Network process

仿真数据设置：

- 每个点过程类型生成20000个序列
- 固定时间窗口[0,15]
- 四种三类型组合：
- IP+SE+SC, IP+SC+NN, IP+SE+NN, SE+SC+NN

模型拟合结果(经验概率强度函数)-真实数据

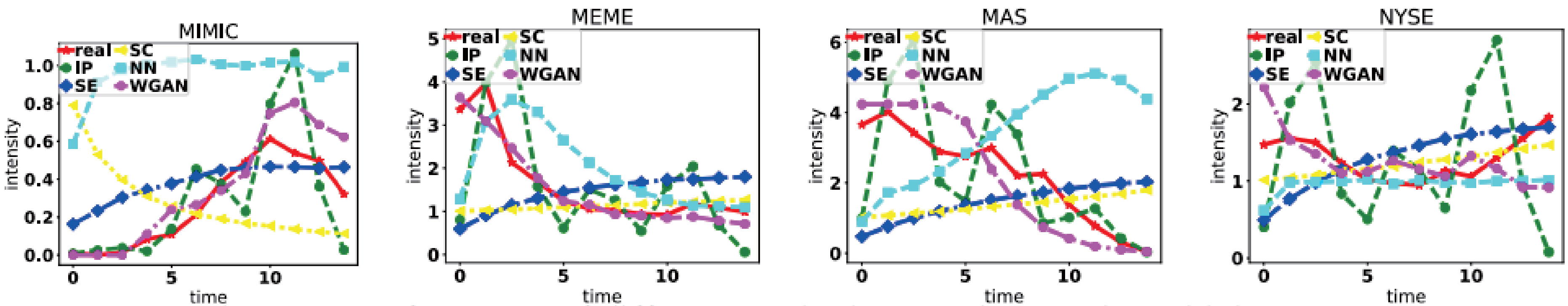


Figure 3: Performance of different methods on various real-world datasets.

Table 2: Deviation of empirical intensity for real-world data.

Data	Estimator				
	MLE-IP	MLE-SE	MLE-SC	MLE-NN	WGAN
MIMIC	0.150	0.160	0.339	0.686	0.122
Meme	0.839	1.008	0.701	0.920	0.351
MAS	1.089	1.693	1.592	2.712	0.849
NYSE	0.799	0.426	0.361	0.347	0.303

真实数据集：

MIMIC：医疗数据集

MEME：社交网络数据集

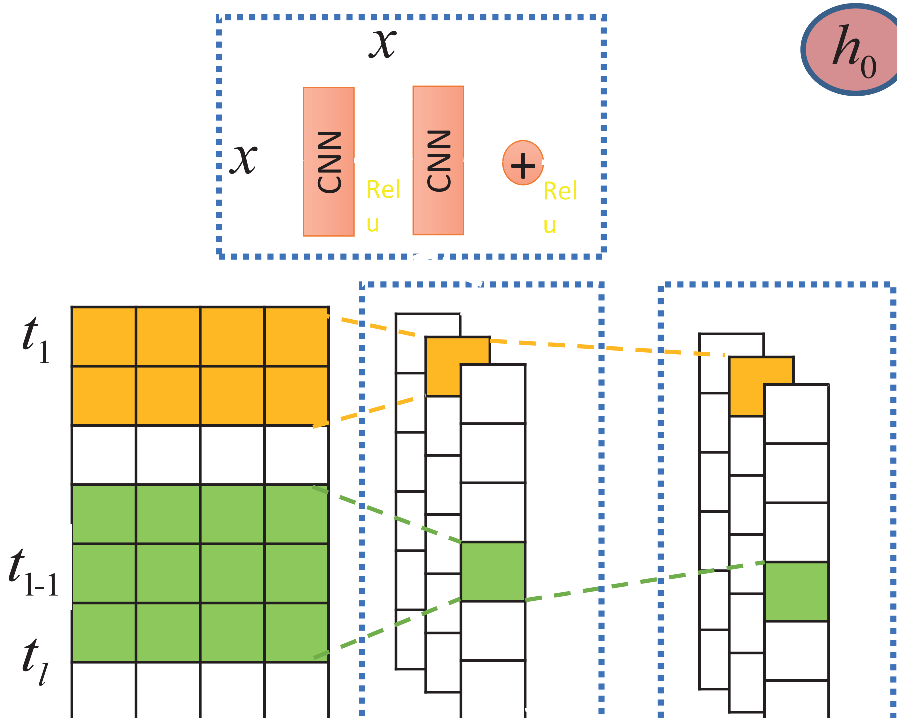
MAS：文献引用数据集

NYSE ：股票市场数据集

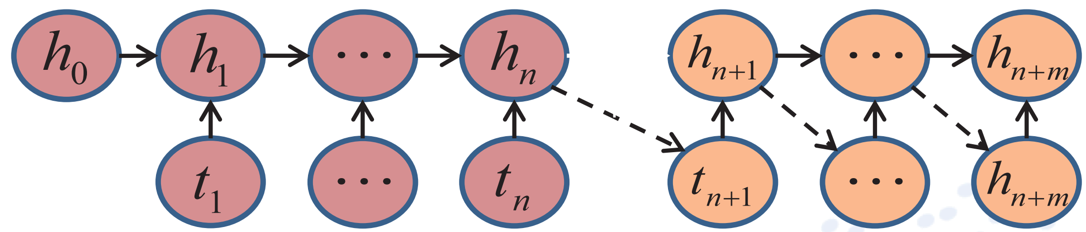
ARXIV论文：

S. Xiao, M. Farajtabar, X. Ye, **J. Yan***, L. Song, X. Yang, H. Zha: Wasserstein Learning of Deep Generative Point Process Models, arXiv:1705.08051, 2017

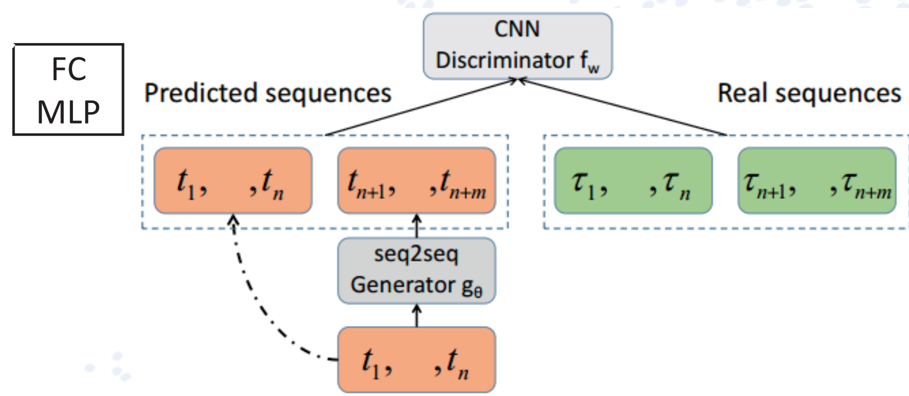
条件GAN:判别器f_w和生成器g_θ(AAAI18)



判别器f_w



条件生成器g_θ



事件预测实验设置(AAAI18)

仿真实验点过程形式设置

Inhomogeneous Poisson process (IP)

$$\lambda(t) = \sum_{i=1}^{k} \alpha_i \left(2\pi\sigma_i^2\right)^{-1/2} \exp\left(-(t-c_i)^2/\sigma_i^2\right)$$

Self-exciting process (SE) [2]

$$\lambda(t) = \mu + \beta \sum_{t_i<t} g(t-t_i)$$

Self-correcting process (SC) [3]

$$\lambda(t) = \exp(\eta t - \sum_{t_i<t} \gamma).$$

线性叠加混合模型：IP+SE+SC

$$Pa.Dev. = \frac{||\hat{\theta} - \theta^\star||_2}{||\theta^\star||_2}.$$

事件预测误差

$$Pr.Dev. = \|\xi - \eta\|_\star$$
$$= \sum\nolimits_{i=1}^{n} |t_i - \tau_i| + \sum\nolimits_{i+1}^{m} T - \tau_{i+1},$$

数据集	链接
MIMIC	https://mimic.physionet.org
LinkedIn	https://github.com/HongtengXu/HawkesProcess-Toolkit/blob/master/Data/LinkedinData.mat
IPTV	https://github.com/HongtengXu/HawkesProcess-Toolkit/blob/master/Data/IPTVData.mat
NYSE	https://github.com/dunan/NeuralPointProcess/tree/master/data/real/book order

[1] John Frank Charles Kingman. Poisson processes. Wiley Online Library, 1993.
[2] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 1971
[3] Valerie Isham and Mark Westcott. A self-correcting point process. *Stochastic Processes and Their* 345 *Applications*, 8(3):335–347, 1979.

事件预测实验结果(AAAI18)

Table 1: Deviation of parameters and prediction for ground-truth and learned model by applying different methods on the synthetic data generated by different point processes.

	Model	Estimator					Conditional Wassestein Estimator	
		MLE-IP	MLE-SE	MLE-SC	MLE-NN	Seq2Seq	SS	CWE
Pa. Dev.	IP	0.03 (3.0e-5)	0.45 (5.0e-4)	0.67 (3.6e-4)	0.36 (3.5e-2)	0.31 (2.6e-3)	0.21 (5.6e-2)	0.09 (5.2e-3)
	SE	0.31 (4.6e-5)	0.02 (3.3e-4)	0.29 (1.5e-5)	0.24 (7.8e-3)	0.19 (2.3e-3)	0.15 (3.9e-2)	**0.02 (4.2e-3)**
	SC	0.94 (7.4e-4)	0.82 (7.4e-4)	0.04 (8.8e-5)	0.10 (2.6e-3)	0.12 (3.3e-3)	0.09 (3.5e-2)	0.07 (6.4e-3)
Pr. Dev.	IP	0.48 (1.3e-4)	0.79 (8.9e-5)	0.93 (3.4e-5)	0.72 (5.8e-2)	0.68 (6.6e-3)	0.64 (3.4e-2)	**0.45 (5.2e-3)**
	SE	1.55 (6.7e-5)	0.94 (1.9e-5)	1.52 (3.7e-4)	1.29 (4.5e-2)	1.27 (6.2e-3)	1.24 (8.2e-2)	**0.96 (9.1e-2)**
	SC	0.58 (7.3e-4)	0.76 (3.1e-5)	0.33 (9.9e-4)	0.44 (3.4e-3)	0.47 (5.2e-3)	0.40 (6.2e-3)	0.36 (6.3e-3)
Pa. Dev.	IP+SE	0.48 (6.2e-5)	0.36 (3.6e-5)	0.32 (6.7e-4)	0.23 (2.3e-2)	0.21 (3.4e-3)	0.18 (2.7e-2)	**0.08 (8.3e-3)**
	IP+SC	0.76 (5.3e-5)	0.88 (3.6e-4)	0.87 (6.2e-5)	0.28 (1.6e-2)	0.29 (7.5e-3)	0.23 (6.2e-3)	**0.11 (6.7e-3)**
	SC+SE	0.51 (7.2e-4)	0.69 (2.6e-4)	0.55 (6.3e-4)	0.32 (6.3e-2)	0.35 (7.5e-3)	0.29 (4.6e-2)	**0.15 (1.2e-3)**
Pr. Dev.	IP+SE	1.65 (5.4e-5)	1.41 (2.3e-5)	1.83 (5.3e-4)	1.03 (5.9e-2)	0.93 (3.1e-3)	0.89 (7.5e-2)	**0.76 (6.8e-3)**
	IP+SC	1.03 (3.0e-4)	0.98 (3.2e-4)	0.95 (0.9e-5)	0.43 (3.9e-3)	0.48 (6.2e-3)	0.40 (4.9e-3)	**0.31 (3.8e-3)**
	SC+SE	1.62 (4.5e-4)	1.43 (2.3e-5)	1.28 (6.7e-4)	0.89 (8.2e-2)	0.92 (4.6e-3)	0.85 (3.1e-2)	**0.63 (2.7e-3)**

Table 2: Deviation of prediction for real-world data.

Data	Estimator						
	MLE-IP	MLE-SE	MLE-SC	MLE-NN	Seq2Seq	SS	CWE
MIMIC	0.25 (2.5e-5)	0.15 (5.3e-4)	0.26 (7.3e-5)	0.19 (2.3e-2)	0.17 (5.3e-3)	0.16 (4.1e-3)	**0.10 (2.5e-3)**
LinkedIn	0.24 (3.1e-4)	0.19 (4.8e-4)	0.17 (9.3e-4)	0.14 (9.1e-3)	0.14 (4.1e-3)	0.12 (8.9e-2)	**0.11 (9.4e-2)**
IPTV	1.46 (3.4e-5)	1.24 (2.8e-5)	1.52 (8.1e-5)	1.21 (2.8e-3)	1.19 (4.2e-2)	1.13 (8.4e-3)	**0.95 (4.9e-3)**
NYSE	2.25 (4.1e-5)	1.96 (6.5e-4)	2.34 (7.3e-5)	1.57 (4.8e-2)	1.55 (2.9e-3)	1.47 (7.3e-3)	**1.23 (2.8e-3)**

总结与展望

- 参数化点过程模型
  - 多用于统计领域，近年来引起机器学习领域广泛关注
- 深度循环网络学习模型
  - 多用于自然语言处理、语音识别、视频分析
  - 近来用于事件序列建模
- 通过注意力机制提高深度学习模型的可解释性
- 对抗学习结合似然估计模型

单维点过程建模
IJCAI'13

多维点过程建模
AAAI'15, IJCAI'16, AAAI'17, KDD'18

神经网络点过程
AAAI'17

对抗学习
NIPS'17 AAAI'18 IJCAI'18

特点	白盒点过程模型	黑盒深度学习模型
可解释性		
融入先验		
求解灵活性		
对先验依赖		
预测精度		

相关论文

主要合作者：肖帅(SJTU), 吴伟昌(SJTU),杨小康教授(SJTU), 刘鑫(ECNU), 查宏远教授(Gatech, ECNU)

Published papers:

1. X. Liu, **J. Yan***, X. Wang, H. Zha, S. Chu: On Predictive Patent Valuation: Forecasting Patent Citations and Their Types. AAAI 2017
2. S. Xiao, **J. Yan***, X. Yang, H. Zha, S. Chu: Modeling the Intensity Function of Point Process Via Recurrent Neural Networks. AAAI 2017: 1597-1603
3. S. Xiao, **J. Yan***, C. Li, B. Jin, X. Wang, X. Yang, S. Chu, H. Zha: On Modeling and Predicting Individual Paper Citation Count over Time. IJCAI 2016: 2676-2682.
4. **J. Yan**, S. Xiao, C. Li, B. Jin, X. Wang, B. Ke, X. Yang, H. Zha: Modeling Contagious Merger and Acquisition via Point Processes with a Profile Regression Prior. IJCAI 2016: 2690-2696.
5. **J. Yan**, C. Zhang, H. Zha, M. Gong, C. Sun, J. Huang, S. M. Chu, X. Yang: On Machine Learning towards Predictive Sales Pipeline Analytics. AAAI 2015: 1945-1951
6. **J. Yan**, Y. Wang, K. Zhou, J. Huang, C. Tian, H. Zha and W. Dong. Towards Effective Prioritizing Water Pipe Replacement and Rehabilitation. IJCAI 2013.
7. S. Xiao, M. Farajtabar, X. Ye, **J. Yan**, L. Song, H. Zha: Wasserstein Learning of Deep Generative Point Process Models, NIPS 2017
8. W. Wu, **J. Yan***, X. Yang, H. Zha: Decoupled and Reformulated Learning of Multi-Dimensional Marked Point Processes, KDD 2018
9. **J. Yan**, X. Liu, L. Shi, C. Li, H. Zha: Discriminative and Adversarial Learning of Differentiable Temporal Point Processes, IJCAI 2018
10. S. Xiao, H. Xu, **J. Yan***, M. Farajtabar, X. Yang, L. Song, H. Zha. Learning Conditional Generative Models for Temporal Point Processes, AAAI 2018.
11. S. Xiao, **J. Yan***, M. Farajtabar, L. Song, X. Yang, H. Zha: Joint Modeling of Event Sequence and Time Series with Attentional Twin Recurrent Neural Networks. arXiv:1703.08524, submitted to TNNLS 2017