Learning Multi Channel Influence in Network

Yingru Li *

Abstract

Many social phenomena, such as the spread of diseases, behaviors, technolies, or products, can naturally be modeled as the diffusion of a contagion across a network. Owing to the potentially high social or economic value of accelerating or inhibiting such diffusions, the goal of understanding the flow of information and predicting information cascades has been an active area of research. In this context, a key task is learning and predicing social influence. Existing methods are based on classical influence models with large amounts of parameters, thus easily overfitted without massive data. Besides, these models rarely consider that diffusion of influence between two entities in network can be caused by multiple reasons: In social network, users can propagate information and influence via multiple communication channels; in disease network, disease can be infected by different ways including air, water, food; in bio-network, e.g. PPI network, proteins connect with each other through multiple pathways. We abstract these phenomena as *multi channel influence*. Since multiple channels are possibly co-existence, in this scenario, the final influence may be a nonlinear combination of the influence from every possible channels, instead of simply a linear combination. In this paper, we mainly study the *noisy-or-like* nonlinear combination of multi-channels influence.

Results show that under the network size of n and the node degree of d with l channels, the lower bound of sample complexity on Network Inference problem of some multi-channel models is $\Omega(d \log nl)$, which is proved with two different methods and is asymptotically equal to the result of $\Omega(d \log n)$ for classical models since l is always a constant. This complexity result indicates that introducing multi channel constraint do not increase the intrinsic complexity of solving network inference problems. Research also shows the influence function class under multi-channel independent cascade model are PAC learnable and the sample complexity is $M = \tilde{O}(\varepsilon^{-2} n^3 l)$, which is not related to the number of the edges, much lower than the sample complexity of that under classical models. we finally give several solutions for network inference problem including frequency statistics, Maximum likelihood estimation and its sparsity version, which shows that introducing prior of multi-channel with *noisy-or* combined effect help reduce the dimensionality of parameter space of influence model and improve the robustness of the algorithms. We also have proposed some potential models on continuous-times multi-channel influence behaviors and then consider modeling online dynamic networks as a future direction.

We concluded with following statements: the *combined effect* can be effectively expressed using the *Noisy or* non-linear model; considering multi channel phenomenon do not increase the intrinsic complexity of network inference problems and introducing the prior of multi channel with noisy-or combination can help reduce the complexity of learning influence functions.

*Huazhong University of Science and Technology, School of Computer Science. szrlee@hust.edu.cn

Abstract

许多社会现象，例如疾病、行为、技术、产品的传播或推广，可以自然的建模成在网络中接触传染的扩散行为。对这样的传播扩散进行加速或者抑制具有潜在的社会价值或经济价值，因此理解信息传播流动以及预测信息传播的路径成为一个热门的研究领域。在这样的背景下，一项重要的任务是学习与预测影响力。解决这一任务现有的方法多基于传统的影响力传播模型。一方面，这些模型含有大量的参数，数据较少时很有可能产生过拟合现象；另一方面，这些模型没有考虑网络内的传播效应可以是由多种因素造成的：例如，在社交网络中，人们可以用不同的通讯渠道来传播信息和影响力；在疾病网络中，疾病也可以通过空气、水、食物等不同的渠道来进行传播；在生物蛋白质相互作用网络中，蛋白质相互作用的方式的也是多种的。

本文将这些现象统称为 *多渠道影响力 (multi channel influence)*。由于不同的渠道可能是共存的，用户间最终的影响强度存在某种*组合效应 (combined effect)*，其可能不是简单地讲不同渠道影响强度进行线性叠加，而是以某种非线性组合的方式叠加。本文主要提出并研究几种*类似或门 (nosiy-or-like)* 的叠加方式，以此构建多渠道影响力的非线性组合模型。

现有的研究结果表明，在网络规模为 n ，网络中点的度数为 d，并存在 l 种渠道的情况下，解决多渠道模型上的网络推断问题所需样本复杂度下界为 $\Omega(d\log nl)$。渠道个数通常是常数级别的量，因此该下界与经典模型的下界 $\Omega(d\log n)$ 在一个数量级上，由此表明引入多渠道的约束并没有使网络推断问题在本质上变难。研究也表明多渠道独立级联模型下的影响力函数是 *PAC 可学习*的，样本复杂度为 $M=\tilde{O}(\varepsilon^{-2}n^3l)$，与边的个数 m 无关，这一结果与传统的传播的样本复杂度相比在数量级上*大幅下降*。附录中给出了针对多渠道新模型的一些问题的新算法，包括频率统计，极大似然估计及其稀疏性正则化变体以及松弛后的优化问题，将对其进一步的论证和验证。

本文回答了：多渠道的*组合效应*通过 *Noisy-or* 的*非线性生成模型*可以进行*有效*的表达。将模型中引入多渠道的约束没有使得其网络推断问题变得复杂。通过对多渠道影响力*组合效应*的*先验假设*，使得影响力函数 PAC 可学习的样本复杂度*降低*。

Contents

1 绪论

1.1 研究背景及现状

许多社会现象, 例如疾病、行为、技术、产品的传播或推广, 可以自然的建模成在网络中接触传染的扩散行为。对这样的传播扩散进行加速或者抑制具有潜在的很高的社会价值或经济价值, 理解信息传播流动以及预测信息传播的路径成为一个热门的研究领域 [KKT03, CWY09, GRBS11, GBL10, CWW10, ML10, AHK14, RNG16, HXKL16]。

在这样的背景下, 一项重要的任务是重建信息及影响力传播流动的网络, 即*网络推断问题 (Network Inference)*: 通过已有的传播过程数据还原或预测信息传播的路径; 另一项重要的任务是学习并预测人群的影响力, 例如一小部分感染疾病的人群会将传染病传播多大的范围, 一个公司的产品能推广到多少用户等, 形式化的讲, 这便是学习*影响力函数 (influence functions)*: 这样的函数将初始接收者集合映射到在传播扩散过程最后被影响者 (也称之为*激活 (active)*) 的人的集合 [KKT03];

有许多方法用来上述问题 [GBL10, GRBS11, DSYS12, GRLK10, DLBS14, NPS15, NS12, YZ13, ZZS13]。大多数方法是基于数据观察来拟合给定传播扩散模型的参数 [GRLK10, GRBS11, NS12, GBL10, NPS15]。最近, Du 等人 [DLBS14] 提出了一种方法将影响力函数视作覆盖函数 (coverage function) 来进行学习; Narasimhan 等人 [NPS15] 建立了几种被广泛使用的传播模型的影响力函数的严格意义上的 PAC 可学习性框架 (proper PAC learnability)。Rosenfeld 等人 [RNG16] 提出了一种辨别式方法, 通过最优化一种核化的 (kernelized) "线性" 函数来学习影响力函数。He 等人 [HXKL16] 考虑了数据缺失的过程, 特别是对传播过程中激活信息的不完全观测情形。

现存的方法主要关注一些经典传播模型, 例如独立级联模型 (Independent Cascade model) 和线性阈值模型 (Linear Threshold model) 或它们的连续时间化版本如连续时间独立级联模型 (Continues-time Independent Cascade model)。这些传播模型都具有一个严重的局限性, 即当用这些模型刻画现实世界的影响力传播时, 由于现实世界网络非常大, 需要大量的模型参数, 即至少为网络中边的数量。学习如此大量的参数对于高效性和可扩张性的应用需求是一个严峻的问题, 但更严重的问题是, 学习大量参数可能带来过拟合的问题。除此之外, 网络内部信息与影响力传播过程中, 不同的原因可能会产生相同的结果, 这样的现象在传统的模型及其现有的扩展中也没有被仔细考虑。下一节对以上存在的问题进行详细的描述, 并讨论潜在的解决方案。

1.2 研究内容

现存的方法主要关注一些经典传播模型, 例如独立级联模型 (Independent Cascade model) 和线性阈值模型 (Linear Threshold model) 或它们的连续时间化版本如连续时间独立级联模型 (Continues-time Independent Cascade model)。这些传播模型都存在一个重要的局限性, 即当用这些模型刻画现实世界的影响力传播时, 由于现实世界网络非常大, 需要大量的模型参数, 即至少为网络中边的数量。学习如此大量的参数对于高效性和可扩张性的应用需求是一个严峻的问题, 但更严重的问题是, 学习大量参数可能带来过拟合的问题。一旦模型参数过拟合, 将导致模型参数的错误估计, 进而导致影响力函数的预测误差变得更大, 这是由于影响力函数对参数的敏感性 (见 3.2.1中的例子) 会放大参数误差所导致的。过去的研究中存在一些方法 [BBM13, DSWZ13] 来降低参数空间的维度。他们主要关注的是影响力与主题建模 (topic modeling) 的关系, 假设用户间传播行为与特定主体模式相关联, 同一个主体 (模型下的隐变量 latent variable) 下的信息传播具有相似性, 模型具体描述见章节 2。然而, 这些隐变量模型是线性的: 具体来说, 这些模型只考虑了可测变量基于隐变量的条件边缘概率与隐变量线性相关的情形, 这是通过主体假设来建模不可避免的局限性, 尤其是对于*因果推断*的任务而言。

除此之外, 网络内部信息与影响力传播过程中, 不同的原因可能会产生相同的结果, 这样的现象在传统的模型及其现有的扩展中也没有被仔细考虑。例如, 在病毒营销的场景下, 公司营销人员根据他们的预算首先选择一部分有影响力的初始用户, 给予他们对所需营销产品的试用机会, 此后这些

初始用户可能会影响他们的亲朋好友，即社交网络中的邻居，来购买产品，被影响的用户又会影响更多的人，这在商业营销中也被称作 口口相传效应 *(word-of-mouth effect)*。值得关注的是，用户影响其“邻居” 的渠道是多样的 [RVB05]，例如直接见面或视频，打电话或语音聊天，发邮件或短信，转发微博朋友圈等等渠道，而且关键的一点是，这些渠道在同一个社交平台内如 facebook、腾讯 QQ 或微信等都是共存的，用户可以使用同一个账户行使所有上述沟通方式，这也为收集潜在的模型所需要的数据提供了现实支持。一方面，通讯渠道可能具有某种固有属性，不同的人使用同一个渠道来对其朋友施加影响可能会产生类似的效果。更重要的是，对于大部分信息通过一些特定的渠道传播，例如视频聊天，会比其它渠道例如文字短信产生一致性的更大的影响，这也是人之常情。另一方面，尽管从像短信这样的 弱渠道 *(weaker channels)* 施加的影响力要小于像视频语音这样的 强渠道 *(dominant channel)*，弱渠道的传播效应是依然重要且有意义的，是不能被视作噪音忽略不计的。换句话说，在利用多种手段，包括 强和弱的渠道，来传递信息与施加影响力后，最终的影响力应当是某种 组合效应 *(combined effect)*。通过多种渠道的疾病传播行为也有类似的问题：对于特定的疾病，存在多种传播渠道如空气、水、食物等，而其中一种途径的传播感染率更高，但最终感染又是不同途径的综合效果。Myers 等人 [MZL12] 主要对内部 (社交网络内) 和外部 (社交网络外，如大众媒体) 的两种产生传播影响力的原因进行了建模和研究。尽管他们也对传播过程中存在多种原因的现象进行了研究，但其考虑的外部影响是网络外的影响，不是网络内点与点的影响，这与本文考虑的问题，即网络内用户之间产生影响有不同的原因，有较大的差别。

因此，在社交网络的语境下，若多种用以传播信息的沟通渠道在网络内部共存，如果只考虑 强渠道而忽略 弱渠道的影响，在解决影响力的学习问题时会产生重大的误差；并且，简单的线性组合对于刻画这种组合效应可能并不太试用。那么本文提出几个重要的基础问题：

1. 对这种组合效应能否通过某种非线性生成模型来进行有效且健壮的表达？
2. 学习这种模型的内在复杂性是什么？
3. 是否能够通过上述引入的先验知识，并通过高效算法，来降低学习网络中影响力函数的复杂度？

在本论文中，我们首次提出并研究了考虑多通信渠道下的影响力传播学习问题。多渠道影响力传播是一个复杂的现象，要想有意义并且严格地刻画这一现象，我们至少需要做一些假设来严格描述通过多种渠道传播信息下的影响力传播行为的关键要素。作为第一个尝试，本文首先引入一种类似或门 *(noisy-or like)* 的玩具模型，仅考虑两种渠道作用的独立级联模型 (Two-channel Independent Cascade model (2-IC))：考虑两个通讯渠道，例如视频交流 a 和文字短信 b。通过视频渠道激活成功的概率是 α，短信渠道激活成功的概率是 β，且 $\alpha > \beta$。对任意有向点对 (u, v)，我们说两种渠道的效应组合是类似或门的，即当 u 对 v 使用两种渠道尝试激活的时候，两次尝试的结果与顺序无关，相互独立，有 u 以概率 $1 - (1 - \alpha)(1 - \beta)$ 独立成功地激活 v。本文还建立了的离散时间多渠道独立级联模型 (Multi-channel Independent Cascade model (MIC))，连续时间多渠道独立级联模型 (Multi-channel Continues-time Independent Cascade model (MCIC)) 和广义多渠道独立级联模型 (General Multi-channel Independent Cascade model (GMIC))；用一种相关性衰减假设和简化两层网络分别研究了在以上多渠道独立级联模型下网络推断问题的样本复杂度信息论下界；讨论了多渠道独立级联模型下影响力函数的 PAC 可学习性问题。

1.3 论文组织结构

第2章介绍相关定义、影响力传播模型和数据模型。

第3章介绍学习影响力传播的两个相关问题–网络推断问题和影响力函数学习问题，并介绍两个问题的重要方法和结论，分析了严格意义及非严格意义下的影响力函数的 PAC 可学习性。

第4章引入多渠道社交网络的概念，用以对网络中影响力传播存在多种原因进行形式化的建模。建立了贰渠道、多渠道独立级联模型，广义多渠道独立级联模型，多渠道连续时间独立级联模型等模型。

第5章研究了关于几个多渠道影响力模型的学习理论问题，具体包括给定模型下网络推断问题的信息论下界和给定模型下影响力函数的 PAC 可学习性。其中对信息论下界的研究有两种思路，尽管预先假设不同，但在多渠道离散模型上的两个结论是基本一致的；第二种方法在多渠道连续时间模型上研究结论也与离散模型相似。接下来还分析了多渠道独立级联模型下影响力函数的 PAC 可学习性问题。

第6章对全文进行了总结，总结了全文的结果及创新点和难点，对正在或将来继续进行的工作作为展望进行了描述。附录中包括了一些证明、详细计算过程、算法设计的初步工作和未来工作的设想。

2 影响力传播模型

2.1 传播模型

本文将观念、产品或者行为的传播建模成在社会网络上的传播扩散过程。社会网络可以表达成一种有向图 $G = (V, E)$，在这里 $n = |V|$ 是点的个数，而 $m = |E|$ 是边的个数。任意一条边 $e = (u, v)$ 关联一个参数 w_{uv} 表示用户 u 对 v 影响的强度。本文假设图的结构(边集 E)是在影响力函数学习问题 (Influence function learning) 3.2中是已知的，在网络推断问题 (Network inference) 3.1中是未知的，而所有边上的参数 w_{uv} 需要被学习出来。取决于传播模型，已有的工作有很多方式来表达个人与个人之间的影响力的强度。点可以处在两种状态的其中一种，*闲置的 (inactive)* 或 *激活的 (active)*。我们说一个点被激活当它在传播过程中感染了某种疾病或接纳了某种观点/产品/行为。在这个工作中，本文主要关注的是*递进的 (progressive)* 传播模型，在这类模型中一个点一旦被激活就会一直保持激活状态。

传播过程从一个种子集合(初始接纳者)S 开始，这些种子点在最初激活。这个过程在可以在离散时间或连续时间中进行，取决于一种概率随机过程，在这个过程中，其它的点可能会根据其邻居对其影响力而被激活。令 $N(v)$ 为点 v 的入邻居 (in-neighbors)，并令 A^t 为到时刻 t 为止被激活的点的集合。通常来说，A^t 这个记号只用于离散时间模型，而 t_v 这个记号多用于连续时间模型，也可用于离散时间模型。下文介绍几个广泛使用的经典传播模型和它们现有的扩展模型：

- **离散时间独立级联模型 Discrete-time Independent Cascade (IC) model [KKT03]:** 在 IC 模型下，权重 $w_{uv} \in [0, 1]$ 刻画的是激活概率。当一个点 u 在时间 t 变成激活态，它会尝试在时间 $t+1$ 去激活所有目前未激活的邻居。对于其任意邻居 v，这个尝试将以概率 w_{uv} 成功。如果这个尝试成功了，v 会变成激活状态；否则，v 依然处在闲置(非激活)状态。一旦 u 做完了他所有的尝试，它就不能在之后的时间中做进一步的激活尝试。

- **离散时间线性阈值模型 Discrete-time Linear Threshold (LT) model [KKT03]:** LT 模型也是一种离散时间模型。所有点 v 具有一个阈值 θ_v ，阈值 θ_v 从区间 $[0, 1]$ 均匀独立采样。LT 模型下的传播过程在离散时间步长中展开：对网络中任意一个点 v 在时间片刻 t 时被激活，仅当从其邻居的所有的入边权重之和超过它自身的阈值时发生：$\sum_{u \in N(v) \cap A^{t-1}} w_{uv} \geq \theta_v$.

- **主题相关独立级联模型 Topic-aware Independent Cascade (TIC) model [BBM13]:** 在主题相关版本的 IC 模型中，用户间影响概率 w_{uv} 与传播的内容主题相关。对任意有向边 (u, v) 和任意主题 $z \in \{1, \cdots, K\}$，有 w^z_{uv} 表达点 u 对点 v 在主题 z 上施加的影响强度。其次对任意的在网络中传播的*内容 (item)* (产品/观点/文章) i 有一个在不同主题上的分布，即对每个主题 $z \in \{1, \cdots, K\}$，有 $\gamma^z_i = \text{Prob}[Z = z \mid i]$ 其中 $\sum_{i=1}^{K} \gamma^z_i = 1$。在这个模型中，传播过程和 IC 模

型类似: 当一个点 v 首次在传播内容 i 上激活时, 它将有一次影响气未被激活的邻居 u 的尝试机会, 这个尝试独立与过去所有的对 u 的激活尝试。这个尝试将以一个概率成功, 这个概率为以内容 i 在所有主题上的分布对边概率进行线性加权: $w_{uv}^i = \sum_{z=1}^K \gamma_i^z w_{uv}^z$.

- **主题相关线性阈值模型 Topic-aware Linear Threshold (TLT) model [BBM13]:** 对任意有向边 (u, v) 和任意主题 $z \in \{1, \cdots, K\}$, 有一个权值 w_{uv}^z, 且满足 $\sum_{u \in N(v)} w_{uv}^z \leq 1$。每个点 v 从区间 $[0, 1]$ 中独立均匀地采样一个阈值 θ_v 。类似于 LT 模型, 一个点 v 在传播内容 i 于时间片刻 t 激活, 仅当从其邻居的所有的入边权重之和超过它自身的阈值时发生: $\sum_{u \in N(v) \cap A^{t-1}} \sum_{z=1}^K \gamma_i^z w_{uv} \geq \theta_v$.

- 还存在其它一些对于经典离散时间模型的扩展模型 [BKS07, BFO10, CCC+11, HSCJ12, LBGL13, LCL15], 这些共工作考虑了不同传播实体内容, 例如观点、产品等在同一个网络中传播时会发生竞争与合作的自然现象。例如, 苹果公司的 *iPhone* 在社交网络中推广时也会协助苹果公司的 *iWatch* 的推销, 因此它们天生是配套产品; 而对于安卓 *Android* 手机的推广会起到竞争阻碍作用, 因为它们本身就是竞争产品。

- **连续时间独立级联模型 Continuous-time Independent Cascade (CIC) model [GRLK10, GRBS11, GRLS13b, DSYS12, DSWZ13]:** CIC 模型在连续时间中展开。任意边 $e = (u, v)$ 关联一个延迟分布, 以 w_{uv} 作为分布的参数。当一个点 u 在时间 t 变成新激活的点, 对其所有未激活的邻居 v, 独立地掷一个正面概率为 p 有偏差的色子。如果结果出现是反面, u 将永远不会激活它的邻居 v; 否则, 出现正面的话, 将有一个从 *延迟分布 (delay distribution)* 中采样得到的延迟时间 d_{uv} 。d_{uv} 是指 u 激活 v 所需要的时间, 它可以是无限的 (如果 u 没有成功地激活 v)。特别的, $p = 1$ 永远成立除了在文献 [GRLK10] 中。点在这个过程结束时被视为激活的, 仅当它们是在一个指定的观察窗口 $[0, \tau]$ 内被激活的。如果一个点 v 是被不同邻居都影响了, 只用第一个激活点 v 的邻居才是 *真实父节点 (true parent)*。这样产生的结果是, 尽管这样的社交通信网络可以是任意的有向网络, 但每一次接触传播过程会诱导出一个有向无环图 Directed Acyclic Graph (DAG)。

- 存在不少 CIC 模型的变种, 与原始 CIC 模型的区别主要在于对分配到每条边上的 *延迟分布* 的假设上。Gomez 等人 [GRLK10] 假设不同边上的延迟分布是对指数分布的不同的参数化实例。Gomez 等人在另一篇文献 [GRBS11] 中考虑了三种重要的分布: 指数分布、幂律分布、瑞利分布。Gomez 等人 [GRLS13b] 考虑了分布参数 w_{uv} 随时间进行变化的情况, 用以描述动态变化的传播网络这一社会现象。Du 等人 [DSYS12] 借鉴了 kernel 的无参估计方法来对延迟分布进行真实数据上的估计, 不对延迟分布进行先验假设。和主题相关离散时间模型 [BBM13] 完全一样, Du 等人 [DSWZ13] 改良了在边 $e = (u, v)$ 上的参数 w_{uv} 的定义使之变得主题敏感。Gomez 等人又在文献 [GRLS13a] 中利用 *生存分析 (Survival Analysis)* 的加性风险模型推广了所有此前的相关模型并发展了乘性风险模型使得点既可以增加也可以减少网络中另一个点被激活或感染的可能性或风险。

固定上述定义的一种传播模型和模型参数。对任意的种子集合 S , 令 Δ_S 为以 S 为初始种子集合时最终激活集合的分布。(在 DIC 和 DLT 模型的情况下, 这是当没有新的激活发生时, 激活点的集合; 对与 CIC 模型, 这是在时间 τ 内激活的点的集合。) 对认识点 v, 令 $F_v(S) = \text{Prob}_{A \sim \Delta_S}[v \in A]$ 为 v 以 S 为初始种子, 根据传播模型的动态过程激活的 (边缘) 概率。然后, 定义 *影响力函数 (influence function)* $\boldsymbol{F} : 2^V \to [0, 1]^n$ 将种子集合映射到边缘激活概率向量, 其中 $\boldsymbol{F}(S) = [F_1(S), \ldots, F_n(S)]$。注意到边缘概率并没有捕捉到全部包含在 Δ_S 中关于影响力传播过程的信息 (因为它们并没有反应共同激活 (co-activation) 的模式, 但它们对于许多应用已经足够充分, 例如影响力最大化 (influence maximization) [KKT03] 和影响力估计 (influence estimation) [DSRZ13]。

2.2 级联数据

我们关注从影响力传播产生的级联数据中学习影响力函数及其传播网络的问题。广义上，一个级联 (cascade) $C = (S, A)$ 是随机传播过程的一次实例 (realization)；级联数据可以是*完全观测*的，也可以是只*部分观测*的。Narasimhan 等人 [NPS15] 声称有时我们只能观测到 *哪些点*被激活了，但不知道这些激活 *什么时候*发生的。这个现象被形式化的定义为 *部分 (partial)* 观测的级联 $C = (S, A)$。S 是种子集合，$A \sim \Delta_S, A \supseteq S$ 是在直到随机过程结束时所有被激活的点的集合。这里也有一种对*完全 (full)* 观测级联数据的定义 $C = (S, A^1, \cdots, A^{n-1})$，其中 A^t 是在时间片刻 t 以及此前所有被激活的点的集合，因此有 $A^{t-1} \subseteq A^t$。注意到这里所有被激活的点的完整集合是 $A = A^{n-1}$。在连续时间设定中，作者常用 $\mathbf{t} \equiv \{t_1, \cdots, t_n\}$ 表示 *完全*观测的级联数据记录网络中每个点的激活时间。对每个点 $v, t_v \in [0, \tau] \cup \{\infty\}$。符号 ∞ 表示点没有在观测时间窗口内被激活 – 这并不意味着它们将永远不会被激活。这个记号也可以用在离散时间模型中，只要讲观测时间窗口设为 $\tau = n - 1$，所有的点的激活时间取值为 $0, \cdots, n-1$ 中的一个整数，并且将所有未激活的点 $u \in V \setminus A$ 的激活时间设为无穷 ∞。

2.2.1 不完全观测.

类似于 Narasimhan 等人 [NPS15]，He 等人 [HXKL16] 关注 *是否激活 (activation-only)* 观察 [1] 为了刻画一些点的激活可能未被观察到的事实，He 等人使用以下关于独立随机数据缺失的模型：对于每个 (已激活的) 点 $v \in A \setminus S$，v 的激活信息独立地以概率 r 是被实际*观测到 (observed)*。以概率 $1 - r$，点的激活信息未被实际观测到。对于种子点 $v \in S$，激活信息永远不会缺失。形式化地，定义 $\tilde{A}$ 如下：每个 $v \in S$ 都确定性的在 $\tilde{A}$ 中，并且每个 $v \in A \setminus S$ 都独立地以概率 r 在 $\tilde{A}$ 中。那么，不完全级联数据 (the incomplete cascade) 记作 $\tilde{C} = (S, \tilde{A})$.

2.3 本章小结

本章介绍了一些常用的记号和概念，几个广泛使用的经典传播模型和它们现有的扩展模型，以及对于数据生成的不同假设与建模。

3 网络推断与影响力函数可学习性

3.1 网络推断

影响力或信息传播的网络常常很难获取，是*未知 (unknown)* 的或*未观测到的 (unobserved)*。通常来说，我们最多只能观察到一些特定的网络中的点被影响激活的*时间*，但我们不知道*谁*影响了他们。一个信息传播的例子如下，当博客主发现了新信息后，他们会将其写入博客但可能不会显示地引用来源信息尽管我们也许可能找到一些待定的来源。类似的，在疾病传播中，我们能够观测到病人得病了，但不知道谁影响了他们。此外，在病毒营销的设定中，我们能够观测到客户购买了某个产品或接纳了某种特定的行为，但并不能够明显地了解谁是影响施加者，谁导致了这次接纳或购买行为。因此，问题是重构这个未知的或未被观测的用以传播扩散的社交网络是否可行？这个网络的结构是什么？这便是网络推断 (Network inference) 问题。形式化地，只提供 M 个 *完全*观测级联数据的信息。$\mathcal{C} = \{C_1 = (S_1, A_1^1, \cdots, A_1^{n-1}), \cdots, C_M = (S_M, A_M^1, \cdots, A_M^{n-1})\}$ 或在连续时间设定中 $\mathcal{C} = \{\mathbf{t}^1, \cdots, \mathbf{t}^i, \cdots, \mathbf{t}^M\}$，我们能否推断出潜在的边集 E？

[1] Narasimhan 等人 [NPS15] 将之称作 *部分观测 partial observations*；He 等人 [HXKL16] 将这个术语稍作改变来避免与“不完全观测 incomplete observations” 的混淆。

这个问题常被归结为极大似然估计问题 Maximum Likelihood Estimation (**MLE**) 问题 [ML10, GRBS11, NS12] or 带正则化的 **MLE** 问题 [PH15, GRSDS16]。基本上，求解步骤为首先假定一种传播生成模型，再估计使得级联数据集的似然最大化的模型参数。

3.1.1 离散时间独立级联模型

级联数据的似然. 令 $A^{-1} = \emptyset$ 且 $A^0 = S$. 假设传播模型为 IC 模型，一个 完全观测级联数据 C 的对数似然是：

$$\mathcal{L}(C \mid \boldsymbol{w}) = \sum_{t=1}^{n-1} \sum_{v \in A^t} \left(\log \left(1 - \prod_{u \in A^{t-1}} (1 - w_{uv}) \right) + \sum_{u \in A^{t-2}} \log(1 - w_{uv}) \right) + \sum_{v \in V \setminus A} \sum_{u \in A} \log(1 - w_{uv}) \tag{3.1}$$

这主要包括成功激活的点的似然及未被激活的点的似然两项，成功激活的点根据模型假设只能被在其前一个时间片刻刚激活的点激活；而对于未被激活的点，所有进行激活它的尝试都失败了。

子问题：点的似然最大化. 根据上述解释，一个级联数据的似然可以分解为该级联数据中所有点的激活时间的似然。换用 $\mathbf{t}$ 的记号，基于每个点的对数似然项是：

$$\begin{aligned} \mathcal{L}_v(C \mid \boldsymbol{w}) &= \mathcal{L}_v(\mathbf{t} \mid \boldsymbol{w}) \\ &= \sum_{u : t_u < t_v - 1} \log(1 - w_{uv}) + \log \left(1 - \prod_{u : t_u = t_v - 1} (1 - w_{uv}) \right) \end{aligned} \tag{3.2}$$

注意到这一项对于未被激活的点也是成立的因为这些点的激活时间为设定为无穷。

接下来，给定级联数据集包含 M 个级联数据，基于此分解，可以通过结果 n 独立的子问题来并行化给定 M 个级联数据总的 **MLE** 问题：

$$\boldsymbol{w}_v = \arg\max_{\boldsymbol{w}_v} \mathcal{L}_v(\mathcal{C} \mid \boldsymbol{w}_v), \tag{3.3}$$

其中 $\boldsymbol{w}_v$ 是对于点 v 的入边的 非负边权的集合，且

$$\begin{aligned} \mathcal{L}_v(\mathcal{C} \mid \boldsymbol{w}_v) &= \sum_{i=1}^{M} \mathcal{L}_v(\mathbf{t}^i \mid \boldsymbol{w}_v) \\ &= \sum_{i=1}^{M} \sum_{u : t_u^i < t_v^i - 1} \log(1 - w_{uv}) + \log \left(1 - \prod_{u : t_u^i = t_v^i - 1} (1 - w_{uv}) \right) \end{aligned} \tag{3.4}$$

通过选取一个阈值(通常为 0)对于学到的权值，一条边是否存在将被简单地确当下来只要这条边的权值超过了阈值。

3.1.2 连续时间独立级联模型

对于 CIC 模型，本章节介绍一个广泛使用的似然的形式化方式。首先回顾一下必要的标准记号。对于每条边 $e = (u, v)$，随机延迟时间的似然是 $d_{uv} = t_v - t_u$ is $f(t_v \mid t_u; w_{uv}) = f(t_v - t_u; w_{uv})$，这也

被成为这对边的 *传播似然 transmission likelihood*。累计密度函数，记为 $F(t_v \mid t_u; w_{uv})$，是用传播似然函数的积分计算得来。给定点 u，它在时间点 t_u 被激活，边 $e = (u, v)$ 的 *生存函数 survival function* 是点 v 在时间 t_v 前仍 未被点 u 激活的概率：

$$S(t_v \mid t_u; w_{uv}) = 1 - F(t_v \mid t_u; w_{uv})$$

而定义在边 (u, v) 上的 *危险函数 hazard function*，或称瞬时激活率，是一个比例：

$$H(t_v \mid t_u; w_{uv}) = \frac{f(t_v \mid t_u; w_{uv})}{S(t_v \mid t_u; w_{uv})}$$

给定级联数据下的生存概率. 本节计算这样的概率：一个点一直保持未被激活的状态直到时间 τ，在它一些父节点已经被激活的情形下。考虑一个级联数据 $\mathbf{t}$ 和一个点 v 在整个观测时间窗口内都 没有被激活，即 $t_v > \tau$。由于每个已激活的点 z 都可能独立地影响点 v，网络中没有一个点在时间 τ 内激活点 v 的概率是所有激活时间小于 τ 并指向 v 的边的生存函数的连乘，

$$\prod_{t_z < \tau} S(\tau | t_z; w_{zv}) \tag{3.5}$$

级联数据的似然. 计算所有点的激活时间的似然，以及考率未被激活的点包含的信息得到

$$\begin{aligned} f(\mathbf{t}; \boldsymbol{w}) = & \prod_{t_v \le \tau} \prod_{t_m > \tau} S(\tau \mid t_v; w_{vm}) \\ & \times \prod_{t_z < t_v} S(t_v \mid t_z; w_{zv}) \sum_{u: t_u < t_v} H(t_v \mid t_u; w_{uv}) \end{aligned} \tag{3.6}$$

详细的计算过程可在附录A.3中找到。

由于对多个级联数据中两两都独立的假设，一组级联数据构成的级联数据集 $\mathcal{C}$ 的似然是所有单个级联数据似然的乘积，通过等式 3.6：

$$\prod_{\mathbf{t}^i \in \mathcal{C}} f(\mathbf{t}^i; \boldsymbol{w}) \tag{3.7}$$

类似于 IC 模型，解决问题的目标是找到所有权重 $\boldsymbol{w}$ 使得所有观测到的级联数据 $\mathbf{t}$ 的似然最大化。

$$\begin{array}{ll} \max_{\boldsymbol{w}} & \sum_{\mathbf{t}^i \in \mathcal{C}} \log f(\mathbf{t}^i; \boldsymbol{w}) \\ \text{subject to} & w_{uv} \ge 0, \end{array} \tag{3.8}$$

其中所推断出的网络中的边对应着所求得的点对的参数为非零，既是 $w_{uv} > 0$. 实际上，问题. 3.8 也可以分解到一组彼此独立的更小的子问题上，每一个子问题对应网络中的一个点，其中子问题仅推断每个点的父节点和其对应的这些入边的参数。不失一般性，对于一个特定的点 v，子问题的是如下问题：

$$\begin{array}{ll} \max_{\boldsymbol{w}_v} & \sum_{\mathbf{t}^i \in \mathcal{C}} \mathcal{L}_v(\mathbf{t}^i; \boldsymbol{w}_v) \\ \text{subject to} & w_{uv} \ge 0, \end{array} \tag{3.9}$$

其中 $\boldsymbol{w}_v := \{w_{uv} \mid u = 1, \cdots, n, u \ne v\}$ 是与点 v 有关的变量，

$$\mathcal{L}_v(\mathbf{t}; \boldsymbol{w}_v) = \log \left(\sum_{u: t_u < t_v} H(t_v \mid t_u; w_{uv}) \right) + \sum_{z: t_z < t_v} \log S(t_v \mid t_z; w_{zv})$$

是对于一个观测到激活的点的表达式而 $\mathcal{L}_v(\mathbf{t}; \boldsymbol{w}_v) = \sum_{u:t_u<T} \log S(T \mid t_u; w_{uv})$ 是对于一个为激活的点而言的,这个式子对应与问题. 3.8中涉及到 $\boldsymbol{w}_v$ 变量的项,可以拆分开来。

一部分工作 [NS12, GRSDS16] 提供了对于网络推断问题的极大似然估计求解方法的严格的理论分析并证明网络可被还原的条件和相关的样本复杂度界。除了 **MLE** 方法,其它一些方法包括贪心算法 [NS12] 和 *First-Edge* 算法 [ACKP13] 被提出来解决网络推断问题,上述两个工作也对网络推断问题样复杂度下界 *Lower bound* 进行了分析。

3.2 学习影响力函数

与网络推断问题有稍许不同,学习影响力函数的目标是估计且预测网络中特定种子用户集合最终产生的影响。尽管网络推断的任务可以用来估计整个网络的参数以作为预测影响力的第一个阶段,但影响力函数对模型参数的误差高度敏感(见下一小节的例子),而此前在网络推断问题上的工作并没有给出关于单独一个参数需要达到多少精度才能保证影响力函数的预测能够准确。

3.2.1 影响力函数对模型参数的敏感性

一个常见的通过完全观测数据用以预测影响力的两部曲是:第一步先利用在每个点上的局部影响力的信息来估计模型参数,第二步在利用估计的参数根据传播模型计算特定种子节点集合的最终影响。然而,影响力函数对估计参数的误差会高度敏感:例如,考虑在一条有 n 个节点的链状网络上的 IC 模型,模型中边的参数都为 1;如果所有参数都以常数误差 ε 被低估了,那么链状网络中最后一个点受到影响的概率就会被估计为 $(1-\varepsilon)^n$,这个数字是会比实际值 1 小几个数量级,准确说是指数级别的小,对于大的 n 来说。面对这个问题,有一些相关工作试图去解决。当网络中的参数存在一些误差的时候,健壮优化技术 [CLT+16, HK16] 被发展用以解决这种情形下的影响力最大化问题。此外,在给定模型和网络参数的情况下,计算特定种子集合的影响是一个 #*P-hard* 的问题 [CWW10]。Du 等人 [DLBS14] 就提出了一种直接将影响力函数视作覆盖函数学习的方法,避免了两部曲中的误差可能被放大以及计算困难的问题;Narasimhan 等人 [NPS15] 建立了在多个广泛使用的模型下影响力函数严格的 PAC 可学习性框架,这也给影响力的预测提供了理论上的准确性的保障。

技术上,为了衡量估计误差,本文主要使用平方误差函数,正如 [NPS15, DLBS14] 中使用的。对于两个 n-维向量 $\boldsymbol{x}, \boldsymbol{y}$,平方误差定义为 $\ell_{\text{sq}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n} \cdot ||\boldsymbol{x} - \boldsymbol{y}||_2^2$。当一个或多个研究对象是集合,本文也使用这样的记号:当一个研究对象是集合 S 时,形式化地,我们使用 指示函数 *(indicator function)* $\boldsymbol{\chi}_S$ 将集合转成向量表示,即如果 $v \in S$,则 $\chi_S(v) = 1$,否则 $\chi_S(v) = 0$。特别的,对于一个激活集合 A ,记 $\ell_{\text{sq}}(A, \boldsymbol{F}(S)) = \frac{1}{n}||\boldsymbol{\chi}_A - \boldsymbol{F}(S)||_2^2$ 。

文献 [NPS15, HXKL16] 都形式化的定义从部分观测中学习影响力函数的问题。令 $\mathcal{P}$ 为种子集上的一个分布(例如,在 2^V 上的一个分布),并且固定一种传播模型 $\mathcal{M}$ 和参数,于是我们有对于任意给定种子集的一个激活集合的分布 Δ_S 。

算法输入含有 M 个级联数据的集合 $\mathcal{C} = \{(S_1, A_1), \cdots (S_M, A_M)\}$, 这里每个 S_i 是从 $\mathcal{P}$ 独立采样, 并且 A_i 是 (随机) 激活集合 $A_i \sim \Delta_{S_i}$ 通过某种随机传播过程获得。学习影响力函数 $\boldsymbol{F}$ 的目的是精确地刻画传播过程。此处, 学习影响力函数的精确程度由与真实模型的平方误差来衡量: $\text{err}_{\text{sq}}[\boldsymbol{F}] = \mathbb{E}_{S\sim\mathcal{P}, A\sim\Delta_S}\left[\ell_{\text{sq}}(A, \boldsymbol{F}(S))\right]$. 这里的期望是在种子集合的随机性以及传播过程的随机性上的期望。

3.2.2 影响力函数的 PAC 可学习性

本文延用概率近似正确 (Probably Approximately Correct (PAC)) 学习框架 [Val84] 来描述不完全观测下影响力函数的可学习性。令 $\mathcal{F}_{\mathcal{M}}$ 为在传播模型 $\mathcal{M}$ 下的影响力函数集合,并令 $\mathcal{F}_{\mathcal{L}}$ 为学习算法可以

从中选择最终函数的影响力函数集合。我们说 $\mathcal{F}_{\mathcal{M}}$ 是 PAC 可学习的，仅当存在某个算法 $\mathcal{A}$ 具有以下性质：对任意的 $\varepsilon, \delta \in (0,1)$，给定传播模型的任意一个参数化实例，以及任意一个种子集 S 的分布 $\mathcal{P}$：当给定 部分观测的训练用级联数据集 $\mathcal{C} = \{(S_1, A_1), \cdots, (S_M, A_M)\}$ 有 $M \geq poly(n, m, 1/\varepsilon, 1/\delta)$，$\mathcal{A}$ 会输出 $\boldsymbol{F} \in \mathcal{F}_{\mathcal{L}}$，满足：

$$\text{Prob}_{\mathcal{C}}[\text{err}_{\text{sq}}[\boldsymbol{F}] - \text{err}_{\text{sq}}[\boldsymbol{F}^*] \geq \varepsilon] \leq \delta.$$

这里，$\boldsymbol{F}^* \in \mathcal{F}_{\mathcal{M}}$ 是真实的影响力函数。这里的概率是考虑在训练用级联数据集上的概率，其中包括种子集生成中的随机性和随机传播过程中的随机性。我们说影响力函数学习算法 $\mathcal{A}$ 是 严格意义上的 *(proper)* 仅当 $\mathcal{F}_{\mathcal{L}} \subseteq \mathcal{F}_{\mathcal{M}}$；意思是，算法学习出来的影响力函数会确保是真实传播模型的一个（参数化）实例。否则，我们说 $\mathcal{A}$ 是一个 非严格意义上的 *(improper)* 的学习算法。进一步，在这设定下，如果上述学习算法的运行时间复杂度是以 M 和 G 大小的多项式级别的函数，那么说 $\mathcal{F}_{\mathcal{M}}$ 是 PAC 高效地 *efficiently* 可学习的。另外，称 $\mathcal{F}_{\mathcal{M}}$ 是在 完全观测下 PAC (高效地) 可学习的，仅当上述定义在给定 完全观察的训练级联数据集时成立。

3.3 本章小结

本章介绍了两种背景下，影响力学习的不同问题，一个是网络推断问题，侧重于学习网络结构；一个是影响力函数学习问题，侧重于做出准确预测。针对网络推断问题，介绍了针对不同模型下的一些算法，主要是介绍多种模型下极大似然估计的构建方法及其划分并行子问题的思路。针对影响力函数学习问题，介绍了影响力函数的敏感性问题，及由此引出的 PAC 可学习性问题。附录D中还介绍了严格意义上 PAC 可学习分析框架和非严格意义上 PAC 学习算法。

4 基于多渠道的影响力传播模型

上一章节介绍了一些经典的影响力传播模型及其扩展。这些传播模型都存在一个重要的局限性，即当用这些模型刻画现实世界的影响力传播时，由于现实世界网络非常大，需要大量的模型参数，即至少为网络中边的数量。学习如此大量的参数对于影响力传播相关问题的高效性和可扩张性的应用需求都是一个严峻的问题，更严重的问题是，学习大量参数可能带来过拟合的问题。一旦模型参数过拟合，将导致模型参数的错误估计，而由于影响力函数对参数的敏感性，进一步将导致影响力函数的预测误差被放大。上一章也介绍了部分工作 [BBM13, DSWZ13] 可以用来降低参数空间的维度。其主要关注的是主题建模在影响力传播中的作用：假设用户间传播行为与特定主体模式相关联，同一个主体下的信息传播行为具有相似性，而不同主题的传播行为则可能有较大差别。这部分工作扩展了经典的离散时间模型和连续时间模型。然而，这些扩展的具有隐变量的影响力传播模型是线性的：具体说，这些模型只考虑了可测变量基于隐变量的条件边缘概率与隐变量线性相关的情形，这种局限性是通过主体假设来建模不可避免的。

此外，在网络内部的影响力传播过程中，可能会有多种不同的原因产生相同的传播结果，这样的现象在传统的模型及其现有的扩展中没有被仔细考虑。例如，在病毒营销中，公司营销人员根据预算首先选择高影响力的用户，给予营销产品的试用机会，此后这些用户可能会进而影响他们的亲朋好友来购买产品，被影响的用户又会影响更多的人，这在商业营销中也被称作 口口相传效应 *(word-of-mouth effect)*。有趣的是，用户影响其 “邻居” 的渠道是多样的 [RVB05]，例如视频，语音，邮件短信，微博朋友圈等渠道。一方面，每个通讯渠道可能具有其自身的某种固有属性，即无论传播的内容如何，不同的人群使用相同渠道来施加影响会产生相似的效果。更重要的是，对于大部分信息通过特定的渠道传播例如视频，会比其它渠道例如短信产生更大的影响力。另一方面，尽管从如短信这样的 弱渠道 *(weaker channels)* 所拥有的影响力可能远小于如视频这样的 强渠道 *(dominant channel)*，弱渠道的影响传播效应当是重要且有意义的，是不能被视作噪音忽略不计的。换句话说，在利用多种途径，包括 强和弱的渠道，来传播影响力时，最终的影响力结果应当是某种 组合效应 *(combined*

effect)。在其它场景下，例如通过多种渠道的疾病传播行为，也有类似的问题：对于特定的疾病，存在多种传播渠道例如空气、水、食物等，而其中一种途径的传播感染率可能更高，但最终感染一定是不同途径的综合效果。Myers 等人 [MZL12] 主要对内部 (社交网络内) 和外部 (社交网络外，如大众媒体) 的两种产生传播影响力的原因进行了建模和研究。尽管他们也对传播过程中存在多种原因的现象进行了研究，但其考虑的是网络内与网络外产生影响结果的不同，这与本文考虑的问题– 网络内用户之间产生影响有不同的原因– 有较大的差别。

本论文首次提出了多渠道下的影响力学习问题并对其进行了多层次的研究。多渠道影响力传播是一个较为复杂的现象，为了对这一现象进行有意义并且严格地描述，我们至少需要做一些假设来刻画多渠道传播信息背景下传播行为的关键要素。作为第一个尝试，本文首先关注一种类似或门 *(noisy-or like)* 的两个渠道组合传播的独立级联模型 (Two-channel Independent Cascade model (2-IC))。

4.1 贰渠道独立级联模型 (2-IC)

考虑两个通讯渠道，例如视频交流 a 和文字短信 b。对任意有向边 (u,v)，有 w^a_{uv} 表示用户 u 通过视频交流能独立地成功激活用户 v 的概率；w^b_{uv} 表示用户 u 通过短信文字能独立地成功激活用户 v 的概率。此处假设 a 是 强渠道，b 是弱渠道，先考虑渠道的成功概率为预先设定的与内容无关的常数，如下形式化地定义渠道的强弱：

Definition 4.1 (确定性主导 (Deterministic dominant))**.** 渠道 a 确定性主导渠道 b 当且仅当 $w^a_{uv} > w^b_{uv}$。

Definition 4.2 (λ-确定性主导 (λ - Deterministic dominant))**.** 渠道 a 对渠道 b 是 λ-确定性主导当且仅当 $w^a_{uv} - w^b_{uv} \geq \lambda$。

在这个玩具模型中，本文继续假设对对任意有向边 (u,v)，$w^a_{uv} = \alpha$ 和 $w^b_{uv} = \beta$，即假设渠道的固有性质和无差别性。我们说两种渠道的效应组合是类似或门的，即如果 u 对 v 会使用两种渠道尝试激活，两次尝试的结果与顺序无关，相互独立，有 u 以概率 $1-(1-w^a_{uv})(1-w^b_{uv}) = 1-(1-\alpha)(1-\beta)$ 独立地激活 v。这便是类似或门 *Noisy-or like* 的组合效应结构。

有时候，渠道的成果激活概率可能并不是确定性的，而也具有一些随机性，即 w^a_{uv} 和 w^b_{uv} 是分别服从假设分布 A 和 B 的随机变量，但即便有随机性，渠道也有强弱之分：

Definition 4.3 (随机性主导 (Stochastic dominant))**.** 渠道 a 随机性主导渠道 b 当且仅当

$$\text{Prob}_{w^a_{uv}\sim A}[w^a_{uv} > x] \geq \text{Prob}_{w^b_{uv}\sim B}[w^b_{uv} > x]$$

对任意可能的取值 x 成立，且对部分取值 x，$\text{Prob}[w^a_{uv} > x] > \text{Prob}[w^b_{uv} > x]$ 。

Definition 4.4 (λ-随机性主导 (λ - Stochastic dominant))**.** 渠道 a 对渠道 b 是 λ-确定性主导当且仅当 $\text{Prob}[w^a_{uv} > x+\lambda] \geq \text{Prob}[w^b_{uv} > x]$ 对任意可能的 x 取值成立。

随机性主导意味着尽管不是每次使用强渠道沟通影响成功激活的可能性都更大，但使用强渠道在成功激活的趋势上一定更大，这也一定程度上表现了现实情况，对渠道的随机性假设比起确定性假设可能更符合渠道的固有属性。一个简单的例子是 A 为 $[0.4, 0.9]$ 上的均匀分布，B 为 $[0.1, 0.6]$ 上的均匀分布，w^a_{uv} 显然是随机主导 w^b_{uv}。

对于图结构，本文先假设静态多重图，在上一章的设定基础上，允许多重边，如 $e^a = (u,v)^a$ 表示 u 到 v 通过渠道 a 的一条边。E^a 表示所有为渠道 a 的边集，在贰渠道模型下，$E = E^a \cup E^b$，$E^{a\vee b}$ 表示至少有渠道 a 或 b 中一个渠道的边集。在这样的设定下，有如下几个值得研究的问题：

4.1.1 贰渠道网络推断

通常来说,我们最多只能观察到一些特定的网络中的点被影响激活的时间,但我们可能不知道谁影响了他们;即便我们知道了是谁的影响,也很难知道是通过什么渠道影响了他们。一个信息传播的例子如下,当用户在社交平台发现了新信息后,他们会将其写入该社交平台的微博,但微博很久以前就即可以视频信息,语音信息,也可以是纯文字信息,而我们并不知道该用户在发微博前是看到了哪种或哪几种信息。类似的,在传染疾病传播中,我们能够观测到病人得病了,但不知道病毒是通过什么渠道感染病人的。此外,在病毒营销的设定中,我们能够观测到客户购买了某个产品或接纳了某种特定的行为,但并不能够清楚地了解影响施加者是通过哪种或哪几种渠道沟通的,如何导致了这次接纳或购买行为。因此,问题是重构信息传播的路径,重构这个未知的多渠道社交网络是否可行?形式化地,提供 M 个 完全观测级联数据的信息和 $E^{a\vee b}$ 的信息。$\mathcal{C}=\{C_1=(S_1,A_1^1,\cdots,A_1^{n-1}),\cdots,C_M=(S_M,A_M^1,\cdots,A_M^{n-1})\}$ 或用记号 $\mathcal{C}=\{\mathbf{t}^1,\cdots,\mathbf{t}^i,\cdots,\mathbf{t}^M\}$,我们能否推断出潜在的边集 E,即把两种类型的边 E^a, E^b 区分开?更困难的,不提供 $E^{a\vee b}$ 的信息,只提供级联数据的信息,能否恢复 E^a, E^b?

4.1.2 贰渠道影响力预测

假设贰渠道网络已知或已经被重建,即多重边集 $E=E^a\cup E^b$ 已知,我们能否通过(部分观测)级联数据集 $\mathcal{C}$ 准确地预测给定种子集合的影响力函数?更困难的,能否在不知道多重边集,只知道 $E^{a\vee b}$ 的关于边的信息的情况下来学习影响力函数。

4.2 多渠道独立级联模型 (MIC)

多渠道独立级联模型 (Multi-channel Independent Cascade model (MIC)) 是对贰渠道独立级联模型的自然扩展,考虑 $l\geq 3$ 个传播渠道,对任意有向边 (u,v),有 $w_{uv}^i, i\in[l]$ 表示用户 u 通过第 i 个渠道能独立地成功激活用户 v 的概率;

类似的,如下形式化地定义渠道的强弱:

Definition 4.5 (依序确定性主导 (Sequential Deterministic dominant))**.** 不妨设,渠道 i 确定性主导渠道 $i+1$ 对所有 i 都成立,则有 $w_{uv}^i>w_{uv}^{i+1},\forall i\in[l-1]$。

Definition 4.6 (λ_i-确定性主导 (λ_i - Deterministic dominant))**.** 渠道 i 对渠道 $i+1$ 是 λ_i-确定性主导当且仅当 $w_{uv}^i-w_{uv}^{i+1}\geq\lambda_i$。

这个模型继续假设对对任意有向边 (u,v), $w_{uv}^i=p_i$。我们说两种渠道的效应组合是类似或门的,即如果 u 对 v 会使用 l 个渠道中的部分渠道 L_{uv} 尝试激活,通过不同渠道尝试的结果与顺序无关,相互独立,则有 u 以概率 $1-\prod_{i\in L_{uv}}(1-w_{uv}^i)=1-\prod_{i\in L_{uv}}(1-p_i)$ 独立地激活 v。

关于渠道数 l 的选择. 从理论上讲,渠道数 l 的选取从本质上决定了模型参数学习的精度。如果不存在渠道组合效应的假设,也就是经典的渠道假设,需要模型参数的精度在 ε 范围内,那么需要 $O(1/\varepsilon)$ 个参数就能覆盖区间 $[\mu,1-\mu],\mu\in(0,0.5)$,此处假设参数远离零和一,$\mu$ 是给定的一个常数。如果组合效应是简单的加法,可以使用二进制表示,由此以 ε 的加性误差半径覆盖区间 $[\mu,1-\mu]$ 只需要 $O(\log(1/\varepsilon))$ 个参数。

而使用 *Noisy-or* 的组合形式,需要多少的渠道及其对于的参数,对任意相邻(相邻指两个渠道组合进行 Noisy-or 作用后的影响力大小数值之间不存在一个数值是其它渠道组合得到的)的两种渠道

组合 (A, B), 使得其组合效应的差

$$\Delta = \left| \left(1 - \prod_{i \in A} (1 - p_i)\right) - \left(1 - \prod_{j \in B} (1 - p_j)\right) \right| \tag{4.1}$$

$$= \left| \prod_{i \in A} (1 - p_i) - \prod_{j \in B} (1 - p_j) \right| \tag{4.2}$$

$$= \left| \prod_{i \in A} \bar{p}_i - \prod_{j \in B} \bar{p}_j \right| < 2\varepsilon \tag{4.3}$$

此处, $\bar{p} = 1 - p \in [\mu, 1 - \mu]$。

又由中值定理, 存在 $\xi \in \left(\min\left\{\prod_{i \in A} \bar{p}_i, \prod_{j \in B} \bar{p}_j\right\}, \max\left\{\prod_{i \in A} \bar{p}_i, \prod_{j \in B} \bar{p}_j\right\}\right)$

$$\left| \sum_{i \in A} \log \frac{1}{\bar{p}_i} - \sum_{j \in B} \log \frac{1}{\bar{p}_j} \right| = \left| \log\left(\prod_{i \in A} \bar{p}_i\right) - \log\left(\prod_{j \in B} \bar{p}_j\right) \right| \tag{4.4}$$

$$= \frac{1}{\xi} \cdot \left| \prod_{i \in A} \bar{p}_i - \prod_{j \in B} \bar{p}_j \right| \tag{4.5}$$

$$= \frac{1}{\xi} \cdot \Delta \tag{4.6}$$

因为 $\xi < \max\left\{\prod_{i \in A} \bar{p}_i, \prod_{j \in B} \bar{p}_j\right\} < 1 - \mu$ 恒成立, 现在问题转化为:

需要多少不同的渠道参数 $\log \frac{1}{\bar{p}} \in [\log 1/(1-\mu), \log 1/\mu]$ 能通过简单的加法组合在加性误差 $\frac{\varepsilon}{1-\mu}$ 范围内覆盖所在区间。该问题的结论是至少需要

$$O\left(\log\left(\varepsilon^{-1}(1-\mu)\log\left(\frac{1-\mu}{\mu}\right)\right)\right) = O(\log(1/\varepsilon))$$

个渠道, 也即是原问题结论的充分条件。

进一步, 多渠道的激活概率本身可能也具有一定随机性, 即 w_{uv}^i 是服从分布 D_i 的随机变量, 渠道的随机性强弱之分定义为:

Definition 4.7 (依序随机性主导 (Sequential Stochastic dominant))**.** 不妨设渠道 i 依序随机性主导渠道 $i+1$ 当且仅当 $\text{Prob}_{w_{uv}^i \sim D_i}[w_{uv}^i > x] \geq \text{Prob}_{w_{uv}^{i+1} \sim D_{i+1}}[w_{uv}^{i+1} > x]$ 对任意可能的 x 取值成立, 且对部分取值 x, $\text{Prob}[w_{uv}^i > x] > \text{Prob}[w_{uv}^{i+1} > x]$ 。

Definition 4.8 (λ_i-随机性主导 (λ_i - Stochastic dominant))**.** 渠道 1 对渠道 2 是 λ_i-确定性主导当且仅当 $\text{Prob}[w_{uv}^i > x + \lambda_i] \geq \text{Prob}[w_{uv}^{i+1} > x]$ 对所有可能的 x 取值成立。

对于图结构, 在贰渠道模型的假设基础上, 允许三重及以上重数的边, 如 $e^i = (u, v)^i$ 表示 u 到 v 通过渠道 i 的一条边。E^i 表示所有为渠道 i 的边集, 在多渠道模型下, $E = \bigcup_{i \in [l]} E^i$ 表示多重边集, E^L 表示至少有渠道 L 中一个渠道的边集, 特别的 $E^{[l]}$ 表示至少有一个渠道的边的集合。在多渠道设定下, 同样有几个值得研究的问题:

4.2.1 多渠道网络推断

多渠道网络推断显然是比贰渠道网络推断更为复杂的，因为用户间所选渠道传播的组合指数级地变多了。对于每一对有向点对 (u,v)，所需推断的可能性从 $2^2=4$ 增加到了 2^l 种。形式化地，提供 M 个 完全观测级联数据的信息和 $E^{[l]}$ 的信息。$\mathcal{C}=\{C_1=(S_1,A_1^1,\cdots,A_1^{n-1}),\cdots,C_M=(S_M,A_M^1,\cdots,A_M^{n-1})\}$ 或用记号 $\mathcal{C}=\{\mathbf{t}^1,\cdots,\mathbf{t}^i,\cdots,\mathbf{t}^M\}$，我们能否推断出潜在的边集 E，即把所有类型的边 $E^i, i\in[l]$ 区分开？已知 $E^{[l]}$ 的信息至少能将推断边集的可能性从 n^2 降至 $|E^{[l]}|$，将每个存在边的点对上的多种渠道推断减少一种可能性，最终将所有可能的多重图空间大小从 2^{n^2l} 降至 $2^{|E^{[l]}|(l-1)}$。那么，更困难的，如果不提供 $E^{[l]}$ 的信息，只提供级联数据的信息，能否恢复 $E^1,E^2,\cdots,E^l$？

4.2.2 多渠道影响力预测

同样的，假设多渠道网络已知或已经被重建，即在多重边集 E 已知的条件下，我们能否通过(部分观测)级联数据集 $\mathcal{C}$ 准确地预测给定种子集合的影响力函数？更困难的，在多重边集的未知的情况下，只知道 $E^{[l]}$ 的关于边的信息的情况下，是否能够学习影响力函数。如果能够对多渠道独立级联模型构建严格的 PAC 可学习性架构，学习算法需要多少复杂度级别的样本才能保证影响力函数预测的准确性？

这些问题将后面的章节一一解答。

4.3 广义 MIC 模型

本节主要使用贝叶斯网络的语言概括并推广前几个章节中的模型。

这直接是一个图模型，有隐变量 $c\in\{0,1\}^l$ 和观测变量 $s\in\{0,1\}^{n(n-1)}$，s_{uv} 指示在给定 u 是激活的条件下，u 对 v 的激活尝试是否成功。隐变量 $c_1,c_2,\cdots c_l$ 是独立的并且假设服从伯努利分布。条件分布 $\text{Prob}[s\mid c]$ 的参数由非负权值矩阵 $W\in\mathbb{R}^{n(n-1)\times l}$。本文使用 W^{uv} 表示权值矩阵中所有和 (u,v) 有关的参数，它是一个行向量 $\mathbb{R}^{1\times l}$。在隐变量 c 的条件下，观测变量 s_{uv} 假设独立采样与分布

$$\text{Prob}[s_{uv}=0\mid c]=\prod_{i=1}^{l}\exp\left(-c_iW_{uvi}\right)=\exp\left(-\langle W_{uv},c\rangle\right) \tag{4.7}$$

此处，有 $1-\exp\left(-c_iW_{uvi}\right)$ 可看作是渠道隐变量 c_i 是导致 s_{uv} 成功的原因的概率。在渠道隐变量 c_i 激活的条件下，渠道隐变量 c_i 是导致 s_{uv} 成功的原因的条件概率为 $1-\exp\left(-W_{uvi}\right)=w_{uv}^i$，即为用户 u 通过渠道 i 激活用户 v 的概率，因此也有 $W_{uvi}=\log(1-w_{uv}^i)$。s_{uv} 是成功的仅当至少一个渠道 c_i 的激活尝试成功了 –顾名思义，这也解释了模型类似或门的原因。$s\mid c$ 的条件分布是

$$\text{Prob}[s\mid c]=\prod_{(u,v):u\neq v}\left(1-\exp(-\langle W_{uv},c\rangle)\right)^{s_{uv}}\left(\exp\left(-\langle W_{uv},c\rangle\right)\right)^{1-s_{uv}}$$

模型的名字叫 或门$Noisy-or$ 也是直接来源于一个事实：l 个二值变量 $y_1,y_2,\cdots,y_l$ 的或运算结果是 1 的概率恰好是 $1-\prod_i(\text{Prob}[y_i=0])$。Noisy-or 模型隐式地使用了这种表达；特别的，对于 u 尝试激活 v 这个事件，我们考虑 l 个事件的或运算，其中第 i 个事件是 “渠道 i 没有成功地使 u 激活 v”，这个事件单独的概率是 $\exp(-c_iW_{uvi})$。把这些时间都视作独立事件就能导出等式. 4.7。

只需将 W_{uvi} 设为 $-\log(1-w_{uv}^i)$ 如果 $i\in L_{uv}$；否则 $W_{uvi}=0$，并且把隐变量的伯努利分布的参数设为 1，就可以完全表达上一个章节的模型。除此之外，利用这个模型，可以思考参数的来源本质。

4.4 本章小结

本章基于多渠道的先验假设，对离散时间独立级联模型进行了有机的扩展，定义了 2-IC、MIC 和广义 MIC 模型以及 主导 *dominant* 的概念，也定义了这些新模型的下网络推断问题和影响力函数学习问题，并从直观和理论上多次阐述了用 *Noisy-or* 作为组合效应表述的原因。更多的多渠道模型扩展可参见附录. C。

5 多渠道模型的 PAC 可学习性与样本复杂度

本章节考虑多渠道模型及其相关问题的 PAC 可学习性及样本复杂度的信息论下界等根本性基础问题。

PAC 可学习性框架 [Val84] 是机器学习的理论基础，它通过严格的定义和分析告诉我们什么是可学习的，什么是不可学习的。一般来说，PAC 可学习性框架得到的样本复杂度是算法有关的上界：学习算法要以特定的置信度达到特定的精度至多需要多少样本。

而信息论下界是讨论问题内在的性质，是算法无关的，即考虑问题本身内在的复杂性，解决该问题理论上至少需要多少资源。这个资源可以是时间，比如分析基于比较的排序可以得到时间复杂度的下界。这个资源也可以是样本，在有关学习问题中，分析信息论下界可以得到解决一个学习问题至少需要多少样本，这样一个界可能并没有实际的算法能达到，但是它给算法设计指明了方向。

5.1 多渠道网络推断问题的信息论下界

本节关注对于(近似)学习图结构所需要观察的级联数据个数的下界，并且这个下界是对与任意算法都成立的。显然，我们不能只考虑学习一个图结构的问题，因为在这种情况下，我们可以提出一个专门针对这个图的 “算法”。因此本节参考研究信息论下界中的标准做法，需要考虑图的集合，并研究至少需要多少级联数据才能从这个集合中(近似)找到指定任何一个图。

本节首先在最一般情况下，即对于任何预定义的集合以及近似重构的任务，给出一个下界。然后，我们为多渠道独立级联模型，编辑距离近似在预定义的图集情况下提供在关联性衰减假设下专门的推论。最后在一个简化的图上给出简单情形下的信息论下界。

5.1.1 一般情形的设定.

考虑任意一种广义上的级联数据生成过程，生成了完全观察的级联数据 $\mathbf{t} = \{t_i\}$。令 $\mathcal{G}$ 为一个固定的图的集合以及对应的边的概率且令 G 从该图集中均匀独立选择的一个图。生成一个集合 $\mathcal{C}$，其中有 $|\mathcal{C}| = M$ 个独立的级联数据并且记 $\mathbf{t}^{\mathcal{C}}$ 为所有级联数据中点的激活时间信息。令 $\widehat{G}(\mathbf{t}^{\mathcal{C}})$ 为原图的一个估计，这个估计输入级联数据观察信息并输出一个图。本文称一个图 G' 近似地还原了 G 仅当 $G \in \mathcal{B}(G')$，这里 $\mathcal{B}(G') \subseteq \mathcal{G}$ 是任意一个预定义的图的集合，对于每个 G' 都有一个这样事先定义的集合。

举例说明，如果我们要做准确图还原，可以令 $\mathcal{B}(G') = \{G'\}$，即只有它自身的一个图集。如果我们对于与原图编辑距离在 s 内的还原图感兴趣，我们可以将 $\mathcal{B}(G')$ 设为所有与图 G' 在编辑距离 s 内的图的集合。

此处定义图估计量 $\widehat{G}(\cdot)$ 出现错误的概率为

$$P_e(\widehat{G}) \;:=\; \text{Prob}[G \notin \mathcal{B}(\widehat{G}(\mathbf{t}^{\mathcal{C}}))]$$

这里的概率是在 G 自身选择的随机性和在该图上生成级联数据的随机性上计算的。注意这里对错误的定义是无法做到近似还原。

Theorem 5.1. 在上述广义的设定上, 对任意图的估计量, 确保近似估计失败的概率为 P_e, 至少需要

$$M \geq \frac{(1-P_e)\log\frac{|\mathcal{G}|}{\sup_{G'}|\mathcal{B}(G')|}-1}{\sum_{i\in V}H(t_i)}$$

此处 $H(\cdot)$ 是信息熵函数。t_i 为点 i 的激活时间, 在此处是一个随机变量。

Proof. 为了缩减记号, 简单记 $\widehat{G}(\mathbf{t}^{\mathcal{C}})$ 为 $\widehat{G}$。证明次定理用到了几个基本的信息论不等式, 可以在例如文献 [CT06] 等中找到。下面 $H(\cdot)$ 表示信息熵而 $I(\cdot;\cdot)$ 表示互信息。

可以简单得出结论下面的图过程马尔科夫链。

$$G \longleftrightarrow \mathbf{t}^{\mathcal{C}} \longleftrightarrow \widehat{G}$$

有下列不等式成立:

$$\begin{aligned}
H(G) &= I(G;\widehat{G}) + H(G \mid \widehat{G}) \\
&\overset{(\varsigma_1)}{\leq} I(G;\mathbf{t}^{\mathcal{C}}) + H(G \mid \widehat{G}) \\
&\overset{(\varsigma_2)}{\leq} H(\mathbf{t}^{\mathcal{C}}) + H(G \mid \widehat{G}) \\
&\overset{(\varsigma_3)}{\leq} mH(\mathbf{t}) + H(G \mid \widehat{G}) \\
&\overset{(\varsigma_4)}{\leq} m\sum_{i\in V}H(t_i) + H(G \mid \widehat{G})
\end{aligned}$$

此处 (ς_1) 是遵循数据处理不等式 (data processing inequality), (ς_2) 尊寻互信息与信息熵的基本关系, 即两个随机变量的互信息小于两者自身的信息熵, (ς_3) 和 (ς_4) 遵循信息熵的子加性 (subadditivity) 。因为 G 是从 $\mathcal{G}$ 独立均匀选择的, 则有 $H(G) = \log|\mathcal{G}|$ 。现在可以使用 Fano 不等式来界定 $H(G \mid \widehat{G})$。

$$\begin{aligned}
H(G \mid \widehat{G}) &\overset{(\varsigma_1)}{\leq} H(G, Err \mid \widehat{G}) \\
&\overset{(\varsigma_2)}{=} H(Err \mid \widehat{G}) + H(G \mid Err, \widehat{G}) \\
&\overset{(\varsigma_3)}{\leq} H(Err) + H(G \mid E, \widehat{G}) \\
&\overset{(\varsigma_4)}{\leq} 1 + P_e\log|\mathcal{G}| + (1-P_e)\log\sup_{\widehat{G}}|\mathcal{B}_s(\widehat{G})|
\end{aligned}$$

此处 Err 是错误指标随机变量(即等于 1 仅当 $G \notin \mathcal{B}(\widehat{G})$; 否则为 0), 因此, $P_e = \mathbb{E}[Err]$。(ς_1) 遵循信息熵的单调性, (ς_2) 遵循信息熵的链式法则, (ς_3) 遵循条件熵的单调性, (ς_4) 遵循 Fano 不等式。结合上述两个结构, 得到

$$\begin{aligned}
M\sum_{i\in V}H(t_i) &\geq (1-P_e)\log\frac{|\mathcal{G}|}{\sup_{\widehat{G}}|\mathcal{B}(\widehat{G})|}-1 \\
\Rightarrow M &\geq \frac{(1-P_e)\log\frac{|\mathcal{G}|}{\sup_{\widehat{G}}|\mathcal{B}(\widehat{G})|}-1}{\sum_{i\in V}H(t_i)}
\end{aligned} \tag{5.1}$$

□

接下来,为了将这个广泛的结果应用到特定的图的集合 $\mathcal{G}$ 以及特定的近似还原要求 $\mathcal{B}$ 中,我们需要得到 $|\mathcal{G}|$ 的下界和对任意 G' 而言 $|\mathcal{B}(G')|$ 的上界,以及对任意点 i 而言的 $H(t_i)$ 的上界。下面的引理给出在 IC 模型下关联衰减假设下衰减系数为 α 时 $H(t_i)$ 的上界,MIC 模型也可以做相同的假设得到相同的结论。

关联性衰减 (Correlation decay) [NS12]. 大致上说,在图上的随机过程被称为存在 "关联性衰减" 现象,仅当随机过程后期到达的点具有负面效应在影响力传播问题里,这种所谓负面效应即影响力级联不会传播太远,这种现象在实际数据中是常见的。形式化的,假设存在一个数 $\alpha > 0$ 对任意一点 i 满足所有入边概率之和,即这些事件的联合界 $\sum_k p_{ki} < 1 - \alpha$。下面的引理阐述了这个假设对于点的激活时间的意义,其中 p_{init} 为每个点成为种子点的初始概率。

Lemma 5.2. *对任意点 i 和时间 t, 有*

$$\text{Prob}[T_i = t] \leq (1-\alpha)^{t-1} p_{init}$$

因此,一个点被激活过的概率 $\text{Prob}[T_i < \infty]$ 满足 $p_{init} < \text{Prob}[T_i < \infty] < \frac{p_{init}}{\alpha}$。并且,所有激活点到种子点的平均距离至多为 $\frac{1}{\alpha}$。

Lemma 5.3. *对任意具有关联衰减系数 α 的图, 若对任意点 i 有 $p_{init} < \frac{1}{e}$, 则有*

$$\begin{aligned} H(t_i) &\leq \frac{p_{init}}{1-\alpha}\left(\log\frac{1}{p_{init}} + \left(\frac{1-\alpha}{\alpha}\right)^2 \log\frac{1}{1-\alpha}\right) \\ &\quad - \left(1 - \frac{p_{init}}{\alpha}\right)\log\left(1 - \frac{p_{init}}{\alpha}\right) \\ &=: p_{init}\overline{H}(\alpha, p_{init}) \end{aligned}$$

注意到两个图的 *编辑距离* 那些只在一个图中出现的边的个数(即两个图的对称差)。以下第一个结论没有给定图的额外信息,并且近似还原是在全局的近似。

Corollary 5.4. *令 $\mathcal{G}_d$ 表示所有点的入度边小于等于 d 的图的集合, 并且这些图存在多重边, 多重边最多为 l 个, 称为 l-多重图。且 $\mathcal{B}_\gamma(G')$ 表示所有与 G' 编辑距离小于 γ 的 l-多重图的集合。令 $p_{init} < \frac{1}{e}$。那么对任意算法, 要确保估计错误的概率为 P_e, 至少需要*

$$M > \frac{(1-P_e)}{p_{init}} \frac{1-\alpha}{\overline{H}(\alpha, p_{init})}\left(d\log\frac{nl}{d} - \frac{\gamma}{n}\log\frac{n^2 l}{\gamma}\right) - 1$$

Proof. 有

$$\begin{aligned} \log|\mathcal{G}_d| &= \log\binom{nl}{d}^n = (1+o(1))\, nd\log\frac{nl}{d} \\ \log|\mathcal{B}_\gamma(G')| &\leq \log\binom{l\binom{n}{2}}{\gamma} \leq \gamma\log\frac{n^2 l}{\gamma} \end{aligned}$$

使用上述两个式子,配合定理. 5.1 和引理. 5.3 可以得出结论 □

注意到至少需要 $\Omega(d\log nl - \frac{\gamma}{n}\log n^2 l)$ 个级联数据来进行还原。因为 $l << n$,还原多渠道模型和原先的模型所需要的数据下界相差很小。对于准确还原任务,即 $\gamma = 0$,至少需要 $\Omega(d\log nl)$ 数据。

第二个结论是关于有部分先验知识的情况。特别的,假设岁任意点 i 有集合 $\mathcal{S}_i$,大小为 $|\mathcal{S}_i| = D$。考虑 l-多重图的集合 $\mathcal{G}_{D,d}$,所有点的入度为 d。因此对任意点, 需要学习关于它的 d 个入边, 从所有可能的父节点及不同的渠道中选择, 有 Dl 中选择。对任意点 i , 允许至多有 s_i 个局部误差; 令 $\mathcal{B}_s(G')$ 相应的子图图集。

Corollary 5.5. 对上述设定中，任意错误概率为 P_e 的还原图的估计量，样本数 M 至少为

$$\frac{(1-P_e)}{p_{init}}\frac{1-\alpha}{\overline{H}(\alpha,p_{init})}\left(d\log\frac{D}{d}-\frac{1}{n}\sum_i s_i\log\frac{eD}{s_i}+\log\max(s_i,1)\right)-1$$

Remark: 对于特别的完全还原的结果 (i.e. $s_i=0$)，复杂度去除了关于 n 的依赖。

Proof. 对图的集合的大小有一下界定：

$$\log|\mathcal{G}_d|=\log D\binom{l}{d}^n=(1+o(1))\,nd\log\frac{Dl}{d}$$

相似的，

$$\begin{aligned}\log|\mathcal{B}_s(\widehat{G})| &\le \log\prod_{i\in V}\left(\sum_{j=0}^{s_i}\binom{Dl}{j}\right)\\ &\le \log\prod_{i\in V}\left(\max(1,s_i)\binom{Dl}{s_i}\right)\\ &\le \sum_{i\in V}\log\left(\max(1,s_i)\left(\frac{Dle}{s_i}\right)^{s_i}\right)\\ &= \sum_{i\in V}\log\max(1,s_i)+\sum_{i\in V}s_i\log\frac{Dle}{s_i}\end{aligned}\tag{5.2}$$

此处

$$\mathcal{B}_s(\widehat{G})=\{\widetilde{G}\in\mathcal{G}_d:\widetilde{E}_{\to i}\triangle\widehat{E}_{\to i}\le s_i\;\forall\, i\in V\}$$

注意到在第二个不等式中，本文假设 $s_i\le\frac{D}{2}$，否则如果 $d<\frac{D}{2}$，算法可以选择 $\widehat{\mathcal{V}}_i=\Phi$ 作为正确的结果且如果 $d\ge\frac{D}{2}$，算法可以选择 $\widehat{\mathcal{V}}_i=\mathcal{V}_i$ 作为输出。使用定理. 5.1，(5.2) 和引理. 5.3可以得到结果的第一部分。 □

5.1.2 特殊情形：三层传播网络

类似 Park 等人 [PH16]，本节考虑一个三层网络的 IC 模型(如图 1所示)，用以表示一个两层的多渠道 IC 模型。尽管这是一个非常简单的网络，它依然需要至少 $\Omega(d\log nl)$ 个样本来避免网络推断的失败。

在图 1中，圆圈代表点，边的权重写在边上。第一层的 s_1 是种子点，在零时刻激活，以边权为概率尝试激活第二层的邻居。第三层表示每个点可能存在的 l 个渠道，共 nl 个点。第三层的设置将多重图转化为了简单图，因为第二层和第三层是全连接，并且之间的激活概率全为 1；点 $n+1$ 是最后的子节点，恰好有 $d+1$ 个父节点包括 s_2'，另外 d 个父节点是第三层中的 d 个节点，第二层的点总共能且恰好选择 d 个 渠道 来尝试激活 $n+1$。这样一来，多渠道传播模型转化到了传统的传播模型上，如图中第三层灰色点和第二层到第三层的网络结构，表示第二层中指向该灰色点的父节点能够通过该灰色点所代表的渠道尝试激活子节点 $n+1$。唯一不同的是，点 $n+1$ 的激活时间延迟了一个时间片刻。

给定模型，但未知第三层与最后第四层子节点 $n+1$ 的边关系以及 M 样本，来推断第三层点到子节点的 d 条边的连接关系。

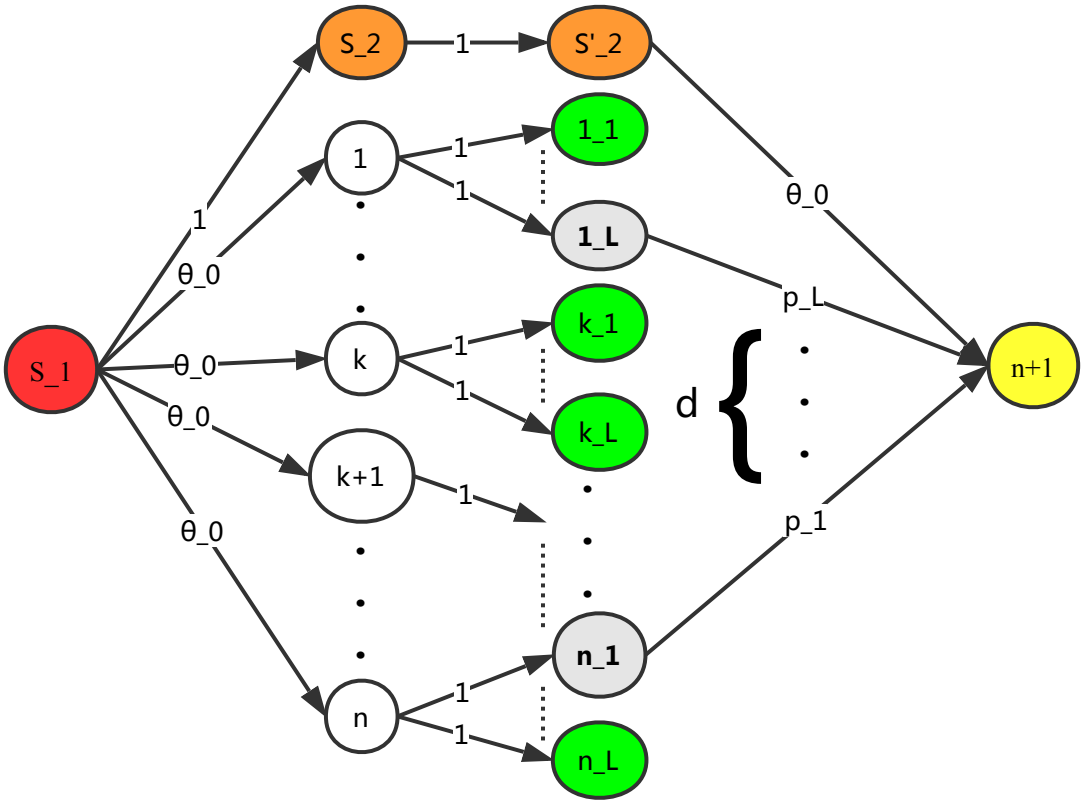

Figure 1: 三层网络的传播模型

假设集 $\mathcal{F}$ 大小为 $|\mathcal{F}| := \binom{nl}{d}$。我们称 π 是一个假设，即作为点 $n+1$ 在第三层的 d 个父节点的集合，满足：对任意的 $i \in \pi$，存在一条边从点 i 到点 $n+1$，边权为对应渠道的固有概率 $p_i \leq \theta$。并令 $\pi^c := \{1_1, 1_2, \ldots, 1_L, \ldots, n_L\} \backslash \pi$ 来作为 π 的补集。给定一个假设 π 和一个样本 $\boldsymbol{t}$，可以写出数据样本的似然：

$$\text{Prob}[\boldsymbol{t}; \pi] = \text{Prob}[\boldsymbol{t}_\pi]\text{Prob}[\boldsymbol{t}_{\pi^c}]\text{Prob}[t_{n+1}|\boldsymbol{t}_\pi t_{s_2}] \tag{5.3}$$

条件概率为

$$\text{Prob}[t_{n+1} = 3|\boldsymbol{t}_\pi t_{s_2}] = 1 - \prod_{i\in\pi}(1-p_i)^{\mathbb{1}[t_i=2]}(1-\theta_0)$$
$$\text{Prob}[t_{n+1} = \infty|\boldsymbol{t}_\pi t_{s_2}] = \prod_{i\in\pi}(1-p_i)^{\mathbb{1}[t_i=2]}(1-\theta_0)$$

不妨设，

$$\theta := 1 - \theta_0^{\frac{1}{d}} \tag{5.4}$$

这个值随子节点 $n+1$ 的父节点增加而减小，这符合直观认识。

下面使用 Fano 不等式来分析对于任意算法而言避免推断失败所至少需要的样本数量。

特殊情形下的信息论下界. 首先,使用基于 Kullback-Leibler (KL) 散度的界 [Yu97] 来界定互信息:

Lemma 5.6. *在离散时间独立级联模型下,对任意两个假设 $\pi, \pi' \in \mathcal{F}$,*

$$\mathbb{KL}(\mathcal{P}_{\boldsymbol{t}|\pi}||\mathcal{P}_{\boldsymbol{t}|\pi'}) \leq \log \frac{1}{\theta_0}$$

Proof. 首先,注意到两个分布 $\mathcal{P}_{\boldsymbol{t}|\pi}$ 和 $\mathcal{P}_{\boldsymbol{t}|\pi'}$ 的最大 KL 散度可以在两个集合 π 和 π' 没有共同的点时,即 $\pi \cap \pi' = \emptyset$。那么可以如下计算两个无交集集合的分布的 KL 散度,

$$\mathbb{KL}(\mathcal{P}_{\boldsymbol{t}|\pi}||\mathcal{P}_{\boldsymbol{t}|\pi'}) = \sum_{\boldsymbol{t}\in\{1,\infty\}^n\times\{3,\infty\}} \text{Prob}[\boldsymbol{t}|\pi] \log \frac{\text{Prob}[\boldsymbol{t}|\pi]}{\text{Prob}[\boldsymbol{t}|\pi']}$$

利用琴生不等式和等式 5.3,有

$$\begin{aligned}
\mathbb{KL}(\mathcal{P}_{\boldsymbol{t}|\pi}||\mathcal{P}_{\boldsymbol{t}|\pi'}) &\leq \log\left(\sum_{\boldsymbol{t}\in\{1,\infty\}^n\times\{3,\infty\}} \text{Prob}[\boldsymbol{t}|\pi]\frac{\text{Prob}[\boldsymbol{t}|\pi]}{\text{Prob}[\boldsymbol{t}|\pi']}\right) \\
&\leq \log\left(\max_{\boldsymbol{t}\in\{1,\infty\}^n\times\{3,\infty\}} \frac{\text{Prob}[\boldsymbol{t}|\pi]}{\text{Prob}[\boldsymbol{t}|\pi']}\right) \\
&= \log\left(\max_{\boldsymbol{t}\in\{1,\infty\}^n\times\{3,\infty\}} \frac{\text{Prob}[\boldsymbol{t}_\pi]\text{Prob}[\boldsymbol{t}_{\pi^c}]\text{Prob}[t_{n+1}|\boldsymbol{t}_\pi t_{s_2}]}{\text{Prob}[\boldsymbol{t}_{\pi'}]\text{Prob}[\boldsymbol{t}_{\pi'^c}]\text{Prob}[t_{n+1}|\boldsymbol{t}_{\pi'} t_{s_2}]}\right) \\
&= \log\left(\max_{\boldsymbol{t}\in\{1,\infty\}^n\times\{3,\infty\}} \frac{\text{Prob}[t_{n+1}|\boldsymbol{t}_\pi t_{s_2}]}{\text{Prob}[t_{n+1}|\boldsymbol{t}_{\pi'} t_{s_2}]}\right)
\end{aligned} \tag{5.5}$$

此前已经讨论过,KL 散度最大值在 $\pi \cap \pi' = \emptyset$ 时取得。因此有

$$\frac{\text{Prob}[t_{n+1}=3|\boldsymbol{t}_\pi t_{s_2}]}{\text{Prob}[t_{n+1}=3|\boldsymbol{t}_{\pi'} t_{s_2}]} \leq \frac{1-\prod_{i\in\pi}(1-p_i)^{\mathbb{1}[t_i=2]}(1-\theta_0)}{1-\prod_{i\in\pi'}(1-p_i)^{\mathbb{1}[t_i=2]}(1-\theta_0)}$$

相似的,有

$$\frac{\text{Prob}[t_{n+1}=\infty|\boldsymbol{t}_\pi t_{s_2}]}{\text{Prob}[t_{n+1}=\infty|\boldsymbol{t}_{\pi'} t_{s_2}]} \leq \frac{\prod_{i\in\pi}(1-p_i)^{\mathbb{1}[t_i=2]}(1-\theta_0)}{\prod_{i\in\pi'}(1-p_i)^{\mathbb{1}[t_i=2]}(1-\theta_0)}$$

可以使用以上表达式来得到等式 5.5的上界。使用等式 5.4,我们有

$$\begin{aligned}
&\mathbb{KL}(\mathcal{P}_{\boldsymbol{t}|\pi}||\mathcal{P}_{\boldsymbol{t}|\pi'}) \\
&\leq \log\left(\max\left\{\frac{1-\prod_{i\in\pi}(1-p_i)(1-\theta_0)}{\theta_0}, \frac{1-\theta_0}{\prod_{i\in\pi'}(1-p_i)(1-\theta_0)}\right\}\right) \\
&\leq \log\left(\max\left\{\frac{1-(1-\theta)^d(1-\theta_0)}{\theta_0}, \frac{1-\theta_0}{(1-\theta)^d(1-\theta_0)}\right\}\right) \\
&\leq \log\left(\frac{1}{\theta_0}\right)
\end{aligned}$$

□

使用上述结论，可以推导出网络推断问题的样本复杂度下界是 $\Omega(d\log nl)$。

Theorem 5.7. *类似的套路，假设“自然”从以 $\mathcal{F}$ 为支撑的假设分布中均匀独立的选取一个“真实”的假设。接下来，级联数据集合在此假设下生成。学习算法要从级联数据集合 $\mathcal{C}$ 中推断 $\hat{\pi}$。在这个简单的三层网络独立级联模型下，存在一个推断网络中 d 个有向边的问题，如果 $M \leq \frac{d\log nl - d\log d - 2\log 2}{2\log\frac{1}{\theta_0}}$，那么推断将以至少 $1/2$ 的概率失败，即，*

$$\text{Prob}[\hat{\pi} \neq \bar{\pi}] \geq \frac{1}{2}$$

对于任意一个学习算法而言都成立。

Proof. 首先用两点 KL 散度来界定互信息 [Yu97]：

$$\begin{aligned}\mathbb{I}(\bar{\pi}, S) &< \frac{1}{|\mathcal{F}|^2}\sum_{\pi\in\mathcal{F}}\sum_{\pi'\in\mathcal{F}}\mathbb{KL}(\mathcal{P}_{S|\pi}||\mathcal{P}_{S|\pi'}) \\ &= \frac{n}{|\mathcal{F}|^2}\sum_{\pi\in\mathcal{F}}\sum_{\pi'\in\mathcal{F}}\mathbb{KL}(\mathcal{P}_{\boldsymbol{t}|\pi}||\mathcal{P}_{\boldsymbol{t}|\pi'})\end{aligned}$$

从引理. 5.6, 互信息的界为：

$$\mathbb{I}(\bar{\pi}, S) < M\log\frac{1}{\theta_0} \tag{5.6}$$

最后，使用 Fano 不等式 [CT06]，等式 5.6 和如下组合系数的界 $\log\binom{nl}{d} \geq d(\log nl - \log d)$，有

$$\begin{aligned}\text{Prob}[\hat{f} \neq \bar{f}] &\geq 1 - \frac{M\log\frac{1}{\theta_0} + \log 2}{\log\binom{nl}{d}} \\ &\geq 1 - \frac{M\log\frac{1}{\theta_0} + \log 2}{d(\log nl - \log d)} \\ &= \frac{1}{2}\end{aligned}$$

解最后一个不等式，有结论，如果 $M \leq \frac{d\log nl - d\log d - 2\log 2}{2\log(1/\theta_0)}$，那么对任何潜在的算法都会以大概率推断失败，$\text{Prob}[\hat{\pi} \neq \bar{\pi}] \geq 1/2$. □

注意到，两种证明思路所得多渠道网络推断样本复杂度信息论下界结果相似，均为 $\Omega(d\log nl)$，与经典模型网络推断问题信息论下界 $\Omega(d\log n)$ 接近，表明考虑多渠道的影响力传播模型并没有大幅增加对应网络推断问题的样本复杂度信息论下界。

5.2 多渠道影响力函数的 PAC 可学习性

本文延用概率近似正确 (Probably Approximately Correct (PAC)) 学习框架 [Val84] 来描述不完全观测下 *多渠道影响力函数*的可学习性。令 $\mathcal{F}_{\mathcal{M}}$ 为在传播模型 $\mathcal{M}$ 下的影响力函数集合，并令 $\mathcal{F}_{\mathcal{L}}$ 为学习算法可以从中选择最终函数的影响力函数集合。我们说 $\mathcal{F}_{\mathcal{M}}$ 是 PAC 可学习的，仅当存在某个算法 $\mathcal{A}$ 具有以下性质：对任意的 $\varepsilon, \delta \in (0,1)$，给定传播模型的任意一个参数化实例，以及任意

一个种子集 S 的分布 $\mathcal{P}$：当给定 部分观测的训练用级联数据集 $\mathcal{C} = \{(S_1, A_1), \cdots, (S_M, A_M)\}$ 有 $M \geq poly(n, m, 1/\varepsilon, 1/\delta)$, $\mathcal{A}$ 会输出 $\boldsymbol{F} \in \mathcal{F}_{\mathcal{L}}$, 满足:

$$\text{Prob}_{\mathcal{C}}[\text{err}_{\text{sq}}[\boldsymbol{F}] - \text{err}_{\text{sq}}[\boldsymbol{F}^*] \geq \varepsilon] \leq \delta.$$

这里, $\boldsymbol{F}^* \in \mathcal{F}_{\mathcal{M}}$ 是真实的影响力函数。这里的概率是考虑在训练用级联数据集上的概率, 其中包括种子集生成中的随机性和随机传播过程中的随机性。我们说影响力函数学习算法 $\mathcal{A}$ 是 严格意义上的 *(proper)* 仅当 $\mathcal{F}_{\mathcal{L}} \subseteq \mathcal{F}_{\mathcal{M}}$; 意思是, 算法学习出来的影响力函数会确保是真实传播模型的一个(参数化)实例。否则, 我们说 $\mathcal{A}$ 是一个 非严格意义上的 *(improper)* 的学习算法。进一步, 在这设定下, 如果上述学习算法的运行时间复杂度是以 M 和 G 大小的多项式级别的函数, 那么说 $\mathcal{F}_{\mathcal{M}}$ 是 PAC 高效地 *efficiently* 可学习的。另外, 称 $\mathcal{F}_{\mathcal{M}}$ 是在 完全观测下 PAC (高效地) 可学习的, 仅当上述定义在给定 完全观察的训练级联数据集时成立。

本章节假设, 对任意有向点对 (u, v), u 对 v 使用的渠道 L_{uv} 已知。

对于多渠道独立级联模型, 其影响力函数有闭式表达, 即表达为在随机无权多重图上, 一个点是否从种子集合可达的期望(类似于传统 IC 模型的解释, 见文献 [KKT03]): 具体的, 因为我们讲所有渠道看作图上的多重边, MIC 的传播过程也可以视作多重边子集的生成过程; 因为每个渠道只会尝试一次激活, 激活的边可以看作是从一次独立的伯努利实验采样而来的。考虑一个激活边的随机子图, 是以概率 w_{uv}^i 选择 $(u, v)^i \in E$ 而来。对于一个给定的生成边的子集和 $A \subseteq E$ 以及给定的种子集合 $S \subseteq V$, 令 $\sigma_v(A, S)$ 作为指示变量表示点 v 从种子集合经过生成的多重图 A 是否可达。由此, MIC 的影响力函数就可以写作在随机生成的多重边子图上 σ 函数的期望:

$$F_v^{\boldsymbol{w}}(S) = \sum_{A \subseteq E} \prod_{(a,b)^i \in E} w_{ab}^i \prod_{(a,b)^i \notin E} (1 - w_{ab}^i) \sigma_v(A, S).$$

尽管这个表达式有指数级别的项数, 可以验证它的梯度是有界的, 即说明 MIC 影响力函数是李普希兹连续的。

Lemma 5.8 (给定 L_1 范数下 MIC 函数的李普希兹连续性)**.** 给定 $S \subseteq V$ 和 $v \in V$。对任意的 $\boldsymbol{w}, \boldsymbol{w}' \in \mathbb{R}^m$ 有 $||\boldsymbol{w} - \boldsymbol{w}'||_1 \leq \varepsilon$, 满足 $|F_v^{\boldsymbol{w}}(S) - F_v^{\boldsymbol{w}'}(S)| \leq \varepsilon$.

由于对渠道与具体传播两方无关的假设, MIC 函数 实际不同参数的个数只有渠道数 l。因此, 可以计算 $[0, 1]^l$ (informal) 上的 ε-覆盖并且利用李普希兹性质将参数空间的 ε-覆盖转移到影响力函数空间的 ε-覆盖, 以此获得该空间上的 覆盖数 *covering number* 的界

Lemma 5.9 (MIC 函数的覆盖数)**.** 对多渠道独立级联模型下的影响力函数的严格函数集合空间的 L_∞ 范数的以半径为 ε 的覆盖数是 $O((l/\varepsilon)^l)$。

定义部分观测级联数据 (S, A) 的对数似然函数为

$$\mathcal{L}(S, A|\boldsymbol{w}) = \sum_{v \in V} \chi_{A_i}(v) \log(F_v^{\boldsymbol{w}}(S)) + (1 - \chi_{A_i}(v)) \log(1 - F_v^{\boldsymbol{w}}(S)).$$

算法基于下列优化问题的解输出影响力函数 $\boldsymbol{F}$:

$$\bar{\boldsymbol{w}} \in \text{argmax}_{\boldsymbol{w} \in [\mu, 1-\mu]^m} \sum_{i=1}^{M} \mathcal{L}(S_i, A_i|\boldsymbol{w}).$$

这里采用基于覆盖数的 标准一致收敛 *standard uniform convergence* 论据来对关于估计参数 $\bar{\boldsymbol{w}}$ 和真实参数的期望对数似然的差异进行界定。

Lemma 5.10 (关于对数似然目标函数的样本复杂度保证)**.** 给定 $\varepsilon, \delta \in (0,1)$, 当样本数 $M = \tilde{O}(\varepsilon^{-2}n^3l)$ 时, 以至少 $1-\delta$ (在训练集的采样上) 的概率,

$$\sup_{\boldsymbol{w}\in[\mu,1-\mu]^m} \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\boldsymbol{w})\right] - \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\bar{\boldsymbol{w}})\right] \leq \varepsilon.$$

最终, 上述理论保证将通过代数变换转移到对估计参数 $\bar{\boldsymbol{w}}$ 和真实参数 $\boldsymbol{w}^*$ 的期望平方误差的差异的界定上。

Theorem 5.11 (MIC 模型下影响力函数的 PAC 可学习性)**.** 在 *MIC* 模型下, 影响力函数集合在给定平方误差下是可学习的, 样本复杂度是 $M = \tilde{O}(\varepsilon^{-2}n^3l)$。

Proof.

$$\begin{aligned}
\mathrm{err}_{\mathrm{sq}}[\boldsymbol{F}^{\bar{\boldsymbol{w}}}] - \mathrm{err}_{\mathrm{sq}}[\boldsymbol{F}^{\boldsymbol{w}^*}] &= \mathbb{E}_{S\sim\mathcal{P},A\sim\Delta_S}\left[\ell_{\mathrm{sq}}(A,\boldsymbol{F}^{\bar{\boldsymbol{w}}}(S))\right] - \mathbb{E}_{S\sim\mathcal{P},A\sim\Delta_S}\left[\ell_{\mathrm{sq}}(A,\boldsymbol{F}^{\boldsymbol{w}^*}(S))\right] && (5.7)\\
&= \mathbb{E}_{S\sim\mathcal{P},A\sim\Delta_S}\left[\ell_{\mathrm{sq}}(A,\boldsymbol{F}^{\bar{\boldsymbol{w}}}(S)) - \ell_{\mathrm{sq}}(A,\boldsymbol{F}^{\boldsymbol{w}^*}(S))\right] && (5.8)\\
&= \mathbb{E}_{S\sim\mathcal{P}}\left[(\boldsymbol{F}^{\bar{\boldsymbol{w}}}(S) - \boldsymbol{F}^{\boldsymbol{w}^*}(S))^2\right] && (5.9)
\end{aligned}$$

此处最后一步, 用到了生成数据的事实 $\mathbb{E}_{A\sim\Delta_S}[\chi_A(v) \mid S] = F_v^{\boldsymbol{w}^*}(S)$ 和部分代数运算。

展开下列式子并使用泰特公式可以得到

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in[\mu,1-\mu]^m} \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\boldsymbol{w})\right] - \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\bar{\boldsymbol{w}})\right] && (5.10)\\
&\qquad = \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\boldsymbol{w}^*)\right] - \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\bar{\boldsymbol{w}})\right] && (5.11)\\
&\qquad \geq \frac{1}{n}\mathbb{E}_S\left[2\left(\mathcal{L}(S,A|\boldsymbol{w}^*) - \mathcal{L}(S,A|\bar{\boldsymbol{w}})\right)^2\right] && (5.12)
\end{aligned}$$

由引理. 5.10可知,

$$\mathrm{err}_{\mathrm{sq}}[\boldsymbol{F}^{\bar{\boldsymbol{w}}}] - \mathrm{err}_{\mathrm{sq}}[\boldsymbol{F}^{\boldsymbol{w}^*}] \leq 0.5\varepsilon$$

即满足了要求。 □

PAC 可学习的结果表明在部分观测下的影响力函数学习没有信息论意义上的障碍, 然而, 这并不能直接给出一个 高效的算法。

5.3 本章小结

本章研究了关于多渠道独立级联模型的两个理论问题, 其一是网络推断的样本复杂度信息论下界, 其二是影响力函数的 PAC 可学习性问题。第一部分通过在一般情形下做关联性衰减假设和将简单的两层网络的多渠道 IC 模型转化为三层网络的传统 IC 模型并不做其它假设的两种证明思路得到多渠道网络推断样本复杂度信息论下界结论, 均为 $\Omega(d\log nl)$, 与经典模型网络推断问题信息论下界 $\Omega(d\log n)$ 接近, 表明考虑多渠道的影响力传播模型并没有大幅增加对应网络推断问题的样本复杂度信息论下界。第二部分得到结论在 MIC 模型下, 影响力函数集合在给定平方误差下是可学习的, 样本复杂度是 $M = \tilde{O}(\varepsilon^{-2}n^3l)$, 由于 $l << m$, 因此表明引入多渠道先验知识的模型, 能够大幅降低样本复杂度上界。

6 总结与展望

6.1 总结

新问题与新模型. 本毕业设计首次提出并研究了多渠道场景下的影响力网络与函数的学习问题,这是一个非常具有现实意义以及深刻理论背景的问题。在总结前人工作的基础上,给出了基于多渠道场景的多种建模方式,包括离散时间模型和连续时间模型的扩展。对离散时间独立级联模型进行了有机的扩展,定义了 2-IC、MIC 和广义 MIC 模型以及 主导 *dominant* 的概念,这种扩展是建立在同时考虑 强渠道和多个 弱渠道的影响,并以非线性的 *Noisy-or* 形式刻画它们的组合效应的基础上。从直观和理论上多次阐述了选择 *Noisy-or* 作为组合效应表述的原因。建立完成模型后,考虑了新模型下两个重要的相关问题–网络推断问题和影响力函数学习问题–并对进行了严格的定义。

理论证明的新思路和新结论. 接下来研究了关于多渠道独立级联模型的两个重要理论问题,其一是网络推断的样本复杂度信息论下界,其二是影响力函数的 PAC 可学习性问题。

第一部分通过两种证明思路得到多渠道网络推断样本复杂度信息论下界结论,第一种思路是在一般情形下通过关联性衰减假设进行证明,第二种是考虑简单的特殊情形下的样本复杂度下界,将简单的两层网络的多渠道 IC 模型转化为三层网络的传统 IC 模型,并不做其它假设,这样做的逻辑是在简单情况下样本复杂度的要求是低于一般情况的。两种证明思路所得到的结论相似,均为 $\Omega(d\log nl)$,表明两种证明思路可以相互佐证,对结论中界的紧度和可信度提供了保障。这个结果也与经典模型网络推断问题信息论下界 $\Omega(d\log n)$ 接近,表明考虑多渠道的影响力传播模型并没有大幅增加对应网络推断问题的样本复杂度信息论下界,即引入多渠道和组合效应的先验知识并没有使网络推断问题在本质上变难。

第二部分得到结论是在 MIC 模型下,影响力函数集合在给定平方误差下是可学习的,样本复杂度是 $M=\tilde{O}(\varepsilon^{-2}n^3l)$,而经典的 IC 模型下影响力函数 PAC 可学习的样本复杂度是 $M=\tilde{O}(\varepsilon^{-2}n^3m)$ 由于 $l<<m$,这个结论表明引入多渠道先验知识的模型,样本复杂度从与边数相关变成与渠道的个数有关而与边数无关,能够大幅降低样本复杂度上界。

以上原创研究也回答了绪论中提出了几个重要问题:

1. 对文中提到的组合效应通过 *Noisy-or* 的非线性生成模型可以进行有效的表达。
2. 样本复杂度信息论下界的研究给出了学习这种模型的内在复杂性,结论表明引入多渠道的约束并没有使网络推断问题在本质上变难。
3. PAC 可学习性的研究表明能够通过上述组合效应的先验知识,来降低学习网络中影响力函数的样本复杂度。

在附录中还提出了在多种条件下的 2-IC、MIC 模型下网络推断问题的求解算法,包括极大似然估计、稀疏性正则化和一种频率统计方法。还提出了多渠道连续时间模型和设想了多渠道动态网络的在线学习问题和对抗学习问题。

6.2 展望

模型扩展与理论分析. 本毕业设计提出的广义多渠道模型可以很简单地进行扩展,用以考虑更复杂的级联数据生成过程。附录中还给出了结合在线学习与博弈论的可能建模方式,这种建模方式可以捕捉用户与算法之间的竞争或用户间通讯方式的动态变化等情景,为未来对此课题持续的研究提供了方向。未来的工作包括继续进行有效的建模,补充对更多多渠道模型尤其是连续时间多渠道模型

的 PAC 可学习性分析以及给出考虑部分观察数据的算法和分析极大似然估计或其它求解方法的样本复杂度和时间复杂度。

算法分析与设计. 附录中也提出了部分针对 MIC 模型下网络推断问题的算法, 正在进行的一项重要工作就是通过理论和实验分析评估现有的算法, 给出现有算法的样本复杂度上界, 并尝试提出更高效或更通用的算法。进一步考虑多渠道先验假设在问题求解时起到的作用, 例如*强渠道*与*弱渠道*的差异性对算法设计能够提供的指导意义。

实际数据搜集, 生成数据仿真与实验. 也可以考虑更贴近实际的问题, 例如主要是搜集真实社交网络的多渠道相关数据, 并对实际网络的多渠道属性进行建模和求解, 研究多渠道的实际建模可行性和意义。还有对生成数据与实际数据的相似性与差异进行分析, 进而提出对现有多渠道模型的补充或从全新的角度对影响力进行建模。

Acknowledgments

First of all, I would like to express my gratitude towards my mentor, or my thesis supervisor, *Prof. Kun He* at Huazhong University of Science and Technology. It was she that introduced me the interesting field of social network. She gives me all the freedom to choose my favorite topic, and is always insightful in what we discussed. Being a wonderful mentor for about three years since my second year in university, she was helping me gain several opportunities to visit most influential research institutes in the world including Cornell University and Microsoft Research Asia (MSRA).

Meanwhile, I shall also thank *Doctor and Senior Researcher Wei Chen* at the Theory Group of MSRA for advising me during the whole summer in 2016. It was indeed an unforgettable holiday since I learned so much there, especially finding one of my real research interests in theoretical aspects. After several fruitful meetings with Wei, we finally set up a research project that inspired from both influence propagation and hidden community detection in social network, which is exactly the topic of my undergraduate thesis. Besides, more than acquiring knowledge, I gained several friends there. I will never forget the days and evenings with my dear intern colleagues at MSRA: Xingyou Song, Zheng Yu and Weiran Huang. We played together, but more importantly we lost ourselves in contemplation and enjoyed the feeling of being so immersed in research. I wish all of you a good future!

I am also feeling very fortunate to meet *Prof. John Hopcroft* at Cornell, with brilliant mind and kinder heart. Without his help, we might not have chances to gain research collaborations with Cornell University and MSRA. We had a few meetings on research project including hidden community detection and also this undergraduate thesis. I shall be grateful to him because it was his words and suggestions on research that lead me to do something that really interest myself and not to be utilitarian. More importantly, John dedicates his hard works on improving Chinese higher education. He spends almost four months per year in China, teaching undergraduates and helping young scholars with research. As a Chinese student, I shall thank John again.

Being the student from School of Computer Science in HUST, I must be grateful and proud since the school provided us opportunities to be involved in real frontier of research and gave me travel supports when visiting Cornell. Some great teachers, say *Duoqiang Wang*, *Zhihu Tan*, *Xuanhua Shi* and *Feng Lu*, have given me interesting courses and also different parts of Computer Science.

I would also like to acknowledge my past advisors, *Hao Tu* and *Xuanhua Shi* for their patience and support for SDN Competition and Parallel Application Contest, and some extraordinary friends of mine including *Chao Tu*, *Wenxuan Wang*, *Yunrui Hu* and *Jiajun Lin*. Some of them added so much laughter and joy to my research and built a common dream on research. Some of them together with me build up solid friendship during coding at Unique Studio and participating tech contests. I will never forget all my teachers including those in my junior and high school. It was they raise my initial interest on mathematics, science and music.

Last but not least, I never forget to thank my dearest *parents* for their unceasing support and altruistic dedication to my family.

For all of the above, and the unnamed, please allow me to express my deepest gratitude, now and always.

References

[ACKP13] Bruno Abrahao, Flavio Chierichetti, Robert Kleinberg, and Alessandro Panconesi. Trace complexity of network inference. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD)*, pages 491–499. ACM, 2013.

[AHK14] Kareem Amin, Hoda Heidari, and Michael Kearns. Learning from contagion (without timestamps). In *ICML '14: Proceedings of the 31th International Conference on Machine Learning*, pages 1845–1853, 2014.

[BBM13] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. Topic-aware social influence propagation models. *Knowledge and information systems*, 37(3):555–584, 2013.

[BFO10] Allan Borodin, Yuval Filmus, and Joel Oren. Threshold models for competitive influence in social networks. In *International Workshop on Internet and Network Economics*, pages 539–550. Springer, 2010.

[BKS07] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In *International Workshop on Web and Internet Economics*, pages 306–311. Springer, 2007.

[CCC+11] Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *the 2011 SIAM International Conference on Data Mining (SDM)*, pages 379–390. SIAM, 2011.

[CLT+16] Wei Chen, Tian Lin, Zihan Tan, Mingfei Zhao, and Xuren Zhou. Robust influence maximization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, KDD '16, pages 795–804, New York, NY, USA, 2016. ACM.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[CWW10] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD)*, pages 1029–1038. ACM, 2010.

[CWY09] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD)*, pages 199–208. ACM, 2009.

[DLBS14] Nan Du, Yingyu Liang, Maria-Florina Balcan, and Le Song. Influence function learning in information diffusion networks. In *ICML '14: Proceedings of the 31th International Conference on Machine Learning*, volume 14, pages 2016–2024, 2014.

[DSRZ13] Nan Du, Le Song, Manuel Gomez Rodriguez, and Hongyuan Zha. Scalable influence estimation in continuous-time diffusion networks. In *Advances in neural information processing systems (NIPS)*, pages 3147–3155, 2013.

[DSWZ13] Nan Du, Le Song, Hyenkyun Woo, and Hongyuan Zha. Uncover topic-sensitive information diffusion networks. In *Artificial Intelligence and Statistics (AISTATS)*, pages 229–237, 2013.

[DSYS12] Nan Du, Le Song, Ming Yuan, and Alex J Smola. Learning networks of heterogeneous influence. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2780–2788, 2012.

[GBL10] Amit Goyal, Francesco Bonchi, and Laks VS Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the third ACM International Conference on Web Search and Data Mining (WSDM)*, pages 241–250. ACM, 2010.

[GRBS11] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *ICML '11: Proceedings of the 28th International Conference on Machine Learning*, pages 561–568, 2011.

[GRLK10] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD)*, pages 1019–1028. ACM, 2010.

[GRLS13a] M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf. Modeling information propagation with survival theory. In *ICML '13: Proceedings of the 31th International Conference on Machine Learning*, 2013.

[GRLS13b] Manuel Gomez Rodriguez, Jure Leskovec, and Bernhard Schölkopf. Structure and dynamics of information pathways in online media. In *Proceedings of the sixth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 23–32. ACM, 2013.

[GRSDS16] Manuel Gomez-Rodriguez, Le Song, Hadi Daneshmand, and B. Schoelkopf. Estimating diffusion networks: Recovery conditions, sample complexity & soft-thresholding algorithm. *Journal of Machine Learning Research (JMLR)*, 2016.

[HK16] Xinran He and David Kempe. Robust influence maximization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, KDD '16, pages 885–894, New York, NY, USA, 2016. ACM.

[HSCJ12] Xinran He, Guojie Song, Wei Chen, and Qingye Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *the 2012 SIAM International Conference on Data Mining (SDM)*, pages 463–474. SIAM, 2012.

[HXKL16] Xinran He, Ke Xu, David Kempe, and Yan Liu. Learning influence functions from incomplete observations. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2065–2073, 2016.

[KKT03] D Kempe, J Kleinberg, and E Tardos. Maximizing the spread of influence in a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 137–146. ACM, 2003.

[LBGL13] Wei Lu, Francesco Bonchi, Amit Goyal, and Laks VS Lakshmanan. The bang for the buck: fair competitive viral marketing from the host perspective. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD)*, pages 928–936. ACM, 2013.

[LCL15] Wei Lu, Wei Chen, and Laks VS Lakshmanan. From competition to complementarity: comparative influence diffusion and maximization. *Proceedings of the VLDB Endowment*, 9(2):60–71, 2015.

[ML10] Seth Myers and Jure Leskovec. On the convexity of latent social network inference. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1741–1749, 2010.

[MZL12] Seth A Myers, Chenguang Zhu, and Jure Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data mining (KDD)*, pages 33–41. ACM, 2012.

[NDRT13] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In *Advances in neural information processing systems (NIPS)*, pages 1196–1204, 2013.

[NPS15] Harikrishna Narasimhan, David C Parkes, and Yaron Singer. Learnability of influence in networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 3186–3194, 2015.

[NS12] Praneeth Netrapalli and Sujay Sanghavi. Learning the graph of epidemic cascades. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 211–222. ACM, 2012.

[PH15] Jean Pouget-Abadie and Thibaut Horel. Inferring graphs from cascades: A sparse recovery framework. In Francis R. Bach and David M. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 977–986. JMLR.org, 2015.

[PH16] Keehwan Park and Jean Honorio. Information-theoretic lower bounds for recovery of diffusion network structures. *CoRR*, abs/1601.07932, 2016.

[RNG16] Nir Rosenfeld, Mor Nitzan, and Amir Globerson. Discriminative learning of infection models. In *the Ninth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 563–572. ACM, 2016.

[RVB05] Arvind Rangaswamy and Gerrit H Van Bruggen. Opportunities and challenges in multichannel marketing: An introduction to the special issue. *Journal of Interactive Marketing*, 19(2):5–11, 2005.

[Val84] Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[Yu97] Bin Yu. *Assouad, Fano, and Le Cam*, pages 423–435. Springer New York, New York, NY, 1997.

[YZ13] Shuang-Hong Yang and Hongyuan Zha. Mixture of mutually exciting processes for viral diffusion. In *ICML'13: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages II–1–II–9. JMLR.org, 2013.

[ZZS13] Ke Zhou, Hongyuan Zha, and Le Song. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 31, pages 641–649, 2013.

A Missing Proofs

A.1 引理. 5.3的证明

对任意具有关联衰减系数 α 的图,若对任意点 i 有 $p_{init} < \frac{1}{e}$,则有

$$\begin{aligned} H(t_i) &\leq \frac{p_{\text{init}}}{1-\alpha}\left(\log\frac{1}{p_{\text{init}}} + \left(\frac{1-\alpha}{\alpha}\right)^2 \log\frac{1}{1-\alpha}\right) \\ &\quad - \left(1-\frac{p_{\text{init}}}{\alpha}\right)\log\left(1-\frac{p_{\text{init}}}{\alpha}\right) \\ &=: p_{init}\overline{H}(\alpha, p_{init}) \end{aligned}$$

Proof. 注意到引理.5.2有结论 $\text{Prob}[T_i = t] \leq (1-\alpha)^{t-1} p_{\text{init}}$。此处的证明及用以上结论来计算 $H(T_i)$. Since $p_{\text{init}} < \frac{1}{e}$, we have the following

$$\begin{aligned} H(T_i) &= -\sum_{t=1}^{n} \text{Prob}[T_i = t]\log\left(\text{Prob}[T_i = t]\right) \\ &\quad - \text{Prob}[T_i = \infty]\log\left(\text{Prob}[T_i = \infty]\right) \\ &\leq -\sum_{t=1}^{n} (1-\alpha)^{t-1} p_{\text{init}} \log (1-\alpha)^{t-1} p_{\text{init}} \\ &\quad - \left(1-\frac{p_{\text{init}}}{\alpha}\right)\log\left(1-\frac{p_{\text{init}}}{\alpha}\right) \\ &\leq -\sum_{t=1}^{\infty} (1-\alpha)^{t-1} p_{\text{init}} \log p_{\text{init}} \\ &\quad - \sum_{t=1}^{\infty} (t-1)(1-\alpha)^{t-1} p_{\text{init}} \log(1-\alpha) \\ &\quad - \left(1-\frac{p_{\text{init}}}{\alpha}\right)\log\left(1-\frac{p_{\text{init}}}{\alpha}\right) \\ &= -p_{\text{init}} \log p_{\text{init}} \frac{1}{1-\alpha} - p_{\text{init}} \log(1-\alpha)\frac{1-\alpha}{\alpha^2} \\ &\quad - \left(1-\frac{p_{\text{init}}}{\alpha}\right)\log\left(1-\frac{p_{\text{init}}}{\alpha}\right) \\ &\leq \frac{p_{\text{init}}}{1-\alpha}\left(\log\frac{1}{p_{\text{init}}} + \left(\frac{1-\alpha}{\alpha}\right)^2 \log\frac{1}{1-\alpha}\right) \\ &\quad - \left(1-\frac{p_{\text{init}}}{\alpha}\right)\log\left(1-\frac{p_{\text{init}}}{\alpha}\right) \end{aligned}$$

□

A.2 引理. 5.8的证明

Proof.

$$F_v^{\boldsymbol{w}}(S) = \sum_{A \subseteq E} \prod_{(a,b)^i \in E} w_{ab}^i \prod_{(a,b)^i \notin E} (1 - w_{ab}^i)\sigma_v(A, S).$$

固定边 $(c,d)^j$ 并取 $F_u^{\boldsymbol{w}}(S)$ 关于 w_{cd}^j 的偏导数：

$$\begin{aligned}
\left|\frac{\partial F_v^{\boldsymbol{w}}(S)}{\partial w_{cd}^j}\right| &= \left|\frac{\partial}{\partial w_{cd}^j}\left(w_{cd}^j \sum_{A\subseteq E\setminus\{(c,d)^j\}} \prod_{(a,b)^i\in E} w_{ab}^i \prod_{(a,b)^i\notin A,(a,b)^i\neq(c,d)^j} (1-w_{ab}^i)\sigma_v(A\cup(c,d)^j,S)\right.\right. \\
&\qquad \left.\left. + (1-w_{cd}^j) \sum_{A\subseteq E\setminus\{(c,d)^j\}} \prod_{(a,b)^i\in E} w_{ab}^i \prod_{(a,b)^i\notin A,(a,b)^i\neq(c,d)^j} (1-w_{ab}^i)\sigma_v(A,S)\right)\right| \\
&= \left|\left(\sum_{A\subseteq E\setminus\{(c,d)^j\}} \prod_{(a,b)^i\in E} w_{ab}^i \prod_{(a,b)^i\notin A,(a,b)^i\neq(c,d)^j} (1-w_{ab}^i)\sigma_v(A\cup(c,d)^j,S)\right.\right. \\
&\qquad \left.\left. - \sum_{A\subseteq E\setminus\{(c,d)^j\}} \prod_{(a,b)^i\in E} w_{ab}^i \prod_{(a,b)^i\notin A,(a,b)^i\neq(c,d)^j} (1-w_{ab}^i)\sigma_v(A,S)\right)\right| \\
&\leq \left|\sum_{A\subseteq E\setminus\{(c,d)^j\}} \prod_{(a,b)^i\in E} w_{ab}^i \prod_{(a,b)^i\notin A,(a,b)^i\neq(c,d)^j} (1-w_{ab}^i)\right| \\
&= 1.
\end{aligned}$$

因此，$||\nabla_{\boldsymbol{w}} F_u^{\boldsymbol{w}}(S)||_\infty \leq 1$，便证明了李普希兹连续性。 □

A.3 级联数据似然3.1.2的计算详细过程

给定一个级联数据 $\mathbf{t}$，我们首先计算观察到的激活 $\mathbf{t}^{\leq\tau} = (t_1,\cdots,t_n \mid t_v \leq \tau)$ 的似然。因为我们假设激活对于给定激活节点的父节点而言是条件独立的，这个似然可以分解到每个点上：

$$f(\mathbf{t}^{\leq\tau};\boldsymbol{w}) = \prod_{t_v\leq\tau} f(t_v \mid \mathbf{t}_{-v};\boldsymbol{w}) \tag{A.1}$$

进一步转化成了计算每个点的激活时间的条件似然，给定除该点信息以外的级联数据。作为递进模型的特点，一个点一旦被第一个父节点激活就一直保持激活状态。给定一个激活的点 v，我们计算所有可能的父节点 u 作为第一个激活 v 的父节点的似然，通过应用等式. 3.5

$$f(t_v \mid t_u;w_{uv}) \times \prod_{z\neq u,t_z<t_v} S(t_v \mid t_z;w_{zv}) \tag{A.2}$$

等式. A.1中的条件似然可以通过对所有每个潜在的父节点作为第一个激活的父节点这些互斥事件的似然之和计算，

$$f(t_v \mid t_{-v};\boldsymbol{w}) = \sum_{u:t_u<t_v} f(t_v \mid t_u;w_{uv}) \prod_{z\neq u,t_z<t_v} S(t_v \mid t_z;w_{zv}) \tag{A.3}$$

并且一个级联数据的似然是

$$f(\mathbf{t}^{\leq\tau};\boldsymbol{w}) = \prod_{t_v\leq\tau} \sum_{u:t_u<t_v} f(t_v \mid t_u;w_{uv}) \prod_{z\neq u,t_z<t_v} S(t_v \mid t_z;w_{zv}) \tag{A.4}$$

通过一些代数变换，我们可以移除式子中的条件 $z\neq u$，使得这个乘积独立于特定的点 u，

$$f(\mathbf{t}^{\leq\tau};\boldsymbol{w}) = \prod_{t_v\leq\tau}\left(\prod_{t_z<t_v} S(t_v \mid t_z;w_{zv})\right) \sum_{u:t_u<t_v} \frac{f(t_v \mid t_u;w_{uv})}{S(t_v \mid t_u;w_{uv})} \tag{A.5}$$

除此之外,那些为被视为的点也对解决问题提供了重要的信息。因此,最终似然的构成是

$$f(\mathbf{t};\boldsymbol{w}) = \prod_{t_v\le\tau}\prod_{t_m>\tau} S(\tau \mid t_v; w_{vm}) \times \prod_{t_z<t_v} S(t_v \mid t_z; w_{zv}) \sum_{u:t_u<t_v} H(t_v \mid t_u; w_{uv}) \tag{A.6}$$

B 多渠道模型下网络推断算法

B.1 贰渠道模型的简单解法

本节结果是在给定完全观察级联数据 $\mathcal{C} = \{(S_1, A_1^{1:n-1}), \ldots, (S_M, A_M^{1:n-1})\}$ 下根据上述假设与性质将两层转化为单层下的学习问题, 并最终转化为统计量估计问题而解决的。这里, $A^{1:n-1} = \{A^1 \ldots, A^{n-1}\}$,而 A^t 恰好在时间 t 被激活的点的集合,与此前章节描述略有不同。假设每对节点都使用两种渠道进行通讯。在已知强渠道参数 α 的情况下,如何简单迅速求解弱渠道参数 β?

由于任意有向对 (u, v) 的最终影响概率可写为 $p = \alpha + (1-\alpha)\beta$,令 $q = 1 - p$,求解 β 的问题可以转化为求解 p 或 q 的问题。

- **方法一:** 对每个点 u, 统计 M 个样本中 $u \in S_i \cup A_i \setminus A_i^n$ 的频数 M_u, 记录下 u 所在的集合 A_i^t, for some $t < n-1$,若 $u \in S$,则记 $t = 0$; 再统计 M 个样本中 $N_{out}(u) \cap A_i^{t+1} \neq \emptyset$ 的频数 $M_{N_{out}(u)}$; 则有当 u 作为传播者时,其邻居一个都没被激活的频率:

  $$1 - \frac{M_{N_{out}(u)}}{M_u} = (1-p_u)^{|N_{out}(u)|} = q_u^{|N_{out}(u)|}$$

  而 $q_u = (1 - M_{N_{out}(u)}/M_u)^{-|N_{out}(u)|}$ 可作为 q 的估计量, 我们最多有 n 个这样的估计量,那么我们的问题就变成了求解 q 使之最接近所有的估计量的问题了。

- **方法二:** 考虑任意点 v,统计 M 个样本中,恰好有其 k 个入邻居 $u \in N(v)$ 都满足 $u \in S$ *or* $u \in A_i^{t-1}$, for some $t < n$ 的频数 $M_{k_{atpt}}$; 在此基础上再统计 $v \in A_i^t$ 的频数 $M_{k_{act}}$; 则有对所有在一个时间片内被尝试激活了 k 次的点没有在 k 次尝试后仍未被激活的频率:

  $$1 - \frac{M_{k_{act}}}{M_{k_{atpt}}} = (1-p_k)^k = q_k^k$$

  而 $q_k = (1 - M_{k_{act}}/M_{k_{atpt}})^{-k}$ 可作为 q 的估计量, 我们最多有 $MAX\ Degree < n$ 个这样的估计量,那么我们的问题就变成了求解 q 使之最接近所有的估计量的问题了。

求解完 q 后,可通过简单的线性变换解出 β。

下面小节考两次更通用的多渠道模型求解问题, 并给出已知渠道信息和未知渠道信息等多种情形下的不同解法。

B.2 已知渠道信息下网络传播参数推断

B.2.1 级联数据的似然

假设传播模型为 MIC 模型,且已知任意点对 (u, v) 的渠道信息 L_{uv}, 求每个渠道的激活成功概率。

首先有,一个 完全观测级联数据 C 的对数似然是:

$$\mathcal{L}(C \mid \boldsymbol{w}) = \sum_{t=1}^{n-1} \sum_{v \in A^t} \left(\log \left(1 - \prod_{u \in A^{t-1}} \prod_{i \in L_{uv}} (1 - w_{uv}^i) \right) + \sum_{u \in A^{t-2}} \sum_{i \in L_{uv}} \log(1 - w_{uv}^i) \right) + \sum_{v \in V \setminus A} \sum_{u \in A} \sum_{i \in L_{uv}} \log(1 - w_{uv}^i) \quad \text{(B.1)}$$

这主要包含两项似然:成功激活的点的似然及未被激活的点的似然,这和基本的 IC 模型无太大差别,差别主要在 L_{uv},若 u 对 v 未使用任何一种传播影响力,那么 L_{uv} 为空集。

B.2.2 子问题: 点的似然最大化

与预备知识中完全相同,此处问题也可以分解到所有的点上。用上对不同渠道传播概率的假设,式子变为

$$\begin{aligned} \mathcal{L}_v(C \mid \boldsymbol{w}) &= \mathcal{L}_v(\mathbf{t} \mid \boldsymbol{w}) \\ &= \sum_{u:t_u < t_v - 1} \sum_{i \in L_{uv}} \log(1 - p_i) + \log \left(1 - \prod_{u:t_u = t_v - 1} \prod_{i \in L_{uv}} (1 - p_i) \right) \end{aligned} \quad \text{(B.2)}$$

同样的,给定级联数据集包含 M 个级联数据,基于此分解,可以通过结果 n 独立的子问题来并行化给定 M 个级联数据总的 **MLE** 问题:

$$\boldsymbol{w}_v = \arg\max_{\boldsymbol{w}_v} \mathcal{L}_v(\mathcal{C} \mid \boldsymbol{w}_v), \quad \text{(B.3)}$$

其中 $\boldsymbol{w}_v$ 是对于点 v 的入边的 非负边权的集合。这个问题由于已知大量信息,应当是容易解的。

B.3 未知渠道信息下多渠道网络推断

依然假设相同渠道的传播概率相同。

$$\begin{aligned} \mathcal{L}_v(C \mid \boldsymbol{w}) &= \mathcal{L}_v(\mathbf{t} \mid \boldsymbol{w}) \\ &= \sum_{u:t_u < t_v - 1} \sum_{i \in [l]} \log(1 - d_{uv}^i p_i) + \log \left(1 - \prod_{u:t_u = t_v - 1} \prod_{i \in [l]} (1 - d_{uv}^i p_i) \right) \end{aligned} \quad \text{(B.4)}$$

此处 $d_{uv}^i \in \{0, 1\}$ 表示 u 对 v 是否使用了通道 i。求解 $\boldsymbol{w}$ 变成了对 p_i 和 d_{uv}^i 的联合求解:

$$\begin{array}{ll} \max_{\{d_v, p\}} & \mathcal{L}_v(\mathcal{C} \mid d_v, p), \\ \text{subject to} & d_v \in \{0, 1\}^l, p \in [\lambda, 1 - \lambda]^l \end{array} \quad \text{(B.5)}$$

可以加入一个限制帮助问题求解:

$$\begin{array}{ll} \max_{\{d_v, p\}} & \mathcal{L}_v(\mathcal{C} \mid d_v, p), \\ \text{subject to} & d_v \in \{0, 1\}^l, \|d_v\|_1 \leq K, p \in [\lambda, 1 - \lambda]^l \end{array} \quad \text{(B.6)}$$

上述问题为整数规划,可能为 NP-hard 问题,故尝试对二值变量连续化:

$$\begin{array}{ll} \max_{\{d_v, p\}} & \mathcal{L}_v(\mathcal{C} \mid d_v, p), \\ \text{subject to} & d_v \in [0, 1]^l, \|d_v\|_1 \leq K, p \in [\lambda, 1 - \lambda]^l \end{array} \quad \text{(B.7)}$$

将限制写入优化目标：

$$\max_{\{d_v,p\}} \mathcal{L}_v(\mathcal{C} \mid d_v, p) - \mu \left\| d_v \right\|_1 \tag{B.8}$$

改用广义模型的写法重写似然表达式：

$$\mathcal{L}_v(\mathbf{t} \mid W_v) = \sum_{u:t_u<t_v-1} \sum_{i\in[l]} (-W_{uvi}) + \log\left(1 - \prod_{u:t_u=t_v-1} \prod_{i\in[l]} \exp(-W_{uvi})\right) \tag{B.9}$$

其中，

$$W_{uvi} = -\log(1 - d^i_{uv} p_i) = \begin{cases} -\log(1-p_i), & d^i_{uv} = 1, \\ 0, & d^i_{uv} = 0. \end{cases} \tag{B.10}$$

引入记号 $\tilde{d} \in [0,1]^l, \tilde{p} \in \mathbb{R}^l$，其中

$$\tilde{p}_i = -\log(1-p_i)$$

，

$$\tilde{W_{uvi}} = d^{\tilde{i}}_{uv} \tilde{p_i} = \begin{cases} -\log(1-p_i), & d^{\tilde{i}}_{uv} = 1, \\ 0, & d^{\tilde{i}}_{uv} = 0. \end{cases} \tag{B.11}$$

有相近的新表达式：

$$\tilde{\mathcal{L}}_v(\mathbf{t} \mid \tilde{d}_v, \tilde{p}) = \sum_{u:t_u<t_v-1} \sum_{i\in[l]} \left(-d^{\tilde{i}}_{uv} \tilde{p_i}\right) + \log\left(1 - \exp\left(- \sum_{u:t_u=t_v-1} \sum_{i\in[l]} d^{\tilde{i}}_{uv} \tilde{p_i}\right)\right) \tag{B.12}$$

以及对应的新问题，

$$\max_{\{\tilde{d}_v,\tilde{p}\}} \tilde{\mathcal{L}}_v(\mathcal{C} \mid \tilde{d}_v, \tilde{p}) - \mu \left\| \tilde{d}_v \right\|_1 \tag{B.13}$$

新问题是对原始问题的放松，且变得简单求解。添加 L_1 范数正则化项是为了刻画网络连接的稀疏性，此前的工作表明，网络间边的连接是稀疏，并且对一个点来说，根据小世界网络的理论，社交网络中的用户所拥有的联系人数不会太大，应当在常数级别。我们也可以将此扩展到假设用户对其邻居所拥有的通信渠道总数也不会很大，因为毕竟沟通也是消耗时间与精力的，这是较为合理的假设。并且增加稀疏性和渠道假设一样可以降低参数个数，提高学习算法的健壮性，避免过拟合。还有一些情况，例如仅仅已知部分观测数据，如何求解这些情况下的问题也是值得考虑的。

C 多渠道模型扩展

C.1 多渠道连续时间独立级联模型 (MCIC)

正文中扩展了经典的离散时间独立级联模型，自然的，对于连续时间独立级联模型，我们也可以使用一种新机制将其扩展为多渠道连续时间独立级联模型 (Multi-channel Continuous-time Independent Cascade model (MCIC))。MCIC 模型在连续时间中展开。任意有向点对 $e = (u, v)$ 关联一个或多个延迟时间分布，对应用户 u 对 v 传播影响所使用的渠道 L_{uv}，分别以 $w^i_{uv}, i \in L_{uv}$ 作为分布的参数。当一个点 u 在时间 t 变成新激活的点，对其所有未激活的邻居 v，从多个渠道对应的多个**延迟分布** ***(delay distribution)*** 中分别采样得到延迟时间 d^i_{uv}。延迟时间 d^i_{uv} 是指 u 经过渠道 i 激活 v 所需要的时间，它可以是无限的（如果 u 没有成功地通过该渠道激活 v）。点在这个过程结束时被视为激活的，

仅当它们是在一个指定的观察窗口 $[0,\tau]$ 内被激活的。如果一个点 v 是被不同邻居通过不同渠道都影响了，只用第一个激活点 v 的邻居才是**真实父节点** *(true parent)*，并且该父节点使用的渠道中第一个到达 v 的渠道才是**真实渠道** *true channel*。这样产生的结果是，尽管这样的社交通信网络可以是任意的多重有向网络，但每一次接触传播过程会诱导出一个有向无环图 Directed Acyclic Graph (DAG)。

在连续时间模型中，渠道的强弱体现在延迟分布上，具体体现在不同渠道延迟时间的快慢上，可以这样解释，影响作用强的渠道有更快传播成功的趋势，而弱的渠道可能存在较慢传播的趋势。形式化地说明，依然是与随机性主导相关的定义：

Definition C.1 (依序随机性主导 (Sequential Stochastic dominant))**.** 不妨设渠道 i 依序随机性主导渠道 $i+1$ 当且仅当 $\text{Prob}[d_{uv}^{i} > x] \geq \text{Prob}[d_{uv}^{i+1} > x]$ 对任意可能的 x 取值成立，且对部分取值 x，$\text{Prob}[d_{uv}^{i} > x] > \text{Prob}[d_{uv}^{i+1} > x]$ 。

注意到依序随机性主导可以诱导出序号在前的渠道都随机性主导序号在后的渠道这个结论。

C.2 多渠道动态网络、在线学习与博弈论

正文主要假设了影响力传播网络是静态的，然而现实并非如此，不同用户间影响力强弱可能随着时间变化，Gomez 等人 [GRLS13b] 就研究类似的问题。本文给出另一种可能的假设，用户间影响力在多渠道网络中随时间变化是由于用户沟通的方式在每次不同的信息传播中，不是一成不变的，例如在一对用户在成为朋友一段时间后，可能关系更加紧密，从前只用短信联络，而一个月后开始使用短信加电话，两三个月后短信电话以及视频或直接见面等方式都开始使用；另一种可能是关系变得疏远，联络频率和方式都变少。这其中有一个假设是关系变化不直接导致影响力的变化，而是关系变化导致的沟通渠道和频率的变化进而导致影响力强弱的变化。这种变化和算法的预测无关。

还有一种有趣的问题是，用户不希望学习算法学习到他们的行为方式，希望保护自己的隐私，于是用户以**自适应**的方式在传播不同信息的时候选择不同的渠道组合，这种选择的变化是基于学习算法的预测，目的是欺骗学习算法。形式化的，级联数据在这类背景下是以在线的方式输入，该问题可以表述为：

给定影响力函数生成模型及其函数集合 $\mathcal{F}_{\mathcal{M}}$ ，

for $i = 1, 2, \ldots, T, \ldots$

(1) 收到对种子集合 S_i 的影响力预测请求；

(2) 根据给定模型预测点之间的连接方式及权重，**确定性地**(或根据**随机策略**)给出影响力函数 $\boldsymbol{F}_i$；

(3) 收到关于种子集合 S_i 的真实影响力结果 A_i (网络可能(自适应地)做出调整，也可能不做任何调整)；

(4) 计算损失 $\ell(F_i, (S_i, A_i))$。

计算前 T 轮中关于函数集中可能的真实函数 $\boldsymbol{F}^* \in \mathcal{F}_{\mathcal{M}}$ 的**悔过** *Regret* 如下：

$$Regret_T(\boldsymbol{F}^*) = \sum_{i=1}^{T} \ell(\boldsymbol{F}_i, (S_i, A_i)) - \sum_{i=1}^{T} \ell(\boldsymbol{F}^*, (S_i, A_i)) \tag{C.1}$$

而算法关于整个假设函数集合 $\mathcal{F}_{\mathcal{M}}$ 的悔过为：

$$Regret_T(\mathcal{F}_{\mathcal{M}}) = \max_{\boldsymbol{F}^* \in \mathcal{F}_{\mathcal{M}}} Regret_T(\boldsymbol{F}^*) \tag{C.2}$$

以上式子中 $\boldsymbol{F}^*$ 也可能随时间变化。在这类背景下，学习算法的目标是最小化关于函数集合 $\mathcal{F}_{\mathcal{M}}$ 的*悔过*：$\min Regret_T(\mathcal{F}_{\mathcal{M}})$。当式子中的 $\boldsymbol{F}^*$ 是根据此前算法的预测做出改变的，这可能涉及到博弈论中的*重复博弈 Repeated Game*，学习算法需要根据什么策略才能应对用户的欺骗，或者用户至少需要做出多大的渠道改变才能够欺骗学习算法？这些都是值得思考的非常有趣的问题。详细细节可参考在线学习或博弈论的书籍及论文，本文此处仅考虑了一种影响力传播背景下的涉及在线学习或博弈论的场景，值得注意的是，这类场景没有学者研究过。

D 经典模型下影响力函数 PAC 可学习性

D.1 严格意义的 PAC 可学习性

本章节将建立在 IC 和 LT 模型下影响力函数的严格意义上的 PAC 可学习性框架。对于这两种传播模型，$\mathcal{F}_{\mathcal{M}}$ 可以用参数向量 $\boldsymbol{w}$ 进行参数化，这里参数向量的每个分量 w_e 是激活概率（在 IC 模型中）或边的权值（在 LT 模型中）。建立可学习性框架的目标是找到影响力函数 $\boldsymbol{F}^{\boldsymbol{w}} \in \mathcal{F}_{\mathcal{M}}$ 能够输出准确的边缘激活概率。尽管我们的目标是*严格意义上的 (proper)* 的学习 — 意味着这个函数必须来自 $\mathcal{F}_{\mathcal{M}}$ — 但并不要求学习（推断）出来的参数一定要和真实的边的参数 $\boldsymbol{w}$ 完全匹配。主要的理论结果将概括在 Theorem D.1 和 Theorem D.2 中。

Theorem D.1 (参考文献 [NPS15, HXKL16])**.** *Let* $\lambda \in (0, 0.5)$. *在所有边上的激活概率满足* $w_e \in [\lambda, 1-\lambda]$ *的 IC 模型下，影响力函数集合 PAC 可学习的。样本复杂度是* $\tilde{O}(\frac{n^3 m}{\varepsilon^2})$. *当在带有保留比例* r *的不完全观察情况下，样本复杂度为* $\tilde{O}(\frac{n^3 m}{\varepsilon^2 r^4})$.

Theorem D.2 (参考文献 [NPS15, HXKL16])**.** *令* $\lambda \in (0, 0.5)$，*考虑每个边的权值满足* $w_e \in [\lambda, 1-\lambda]$，*以及每个点* v *满足* $1 - \sum_{u \in N(v)} w_{uv} \in [\lambda, 1-\lambda]$ *的 LT 模型下的影响力函数集合。这样的函数集是 PAC 可学习的。样本复杂度是* $\tilde{O}(\frac{n^3 m}{\varepsilon^2})$. *当在带有保留比例* r *的不完全观察情况下，样本复杂度为* $\tilde{O}(\frac{n^3 m}{\varepsilon^2 r^4})$.

本章节还将展示一些证明的直观认识以及这两个定理的证明骨架。证明细节可以在原文献找到。

- 第一步是要证明影响力函数 $F_v^{\boldsymbol{w}}$（是）在给定 L_1 范数下是 *1-李普希兹 (Lipschitz)*。（即为参数中有界的变化只会产生对与函数值有界的变化）。这在下述引理D.3中。

  Lemma D.3 (给定 L_1 范数下影响力函数的李普希兹连续性)**.** *给定* $S \subseteq V$ *和* $v \in V$。*对任意的* $\boldsymbol{w}, \boldsymbol{w}' \in \mathbb{R}^m$ *有* $||\boldsymbol{w} - \boldsymbol{w}'||_1 \leq \varepsilon$，*满足* $|F_v^{\boldsymbol{w}}(S) - F_v^{\boldsymbol{w}'}(S)| \leq \varepsilon$.

- 第二步是建立在参数空间 $[0,1]^m$ 上的 ε-覆盖并且利用李普希兹性质将参数空间的 ε-覆盖转移到影响力函数空间的 ε-覆盖，以此获得该空间上的*覆盖数 covering number* 的界

  Lemma D.4 (影响力函数的覆盖数)**.** *对影响力函数的严格函数集合空间的* L_∞ *范数的以半径为* ε *的覆盖数是* $O((m/\varepsilon)^m)$。

- 定义部分观测级联数据 (S, A) 的对数似然函数为

$$\mathcal{L}(S, A|\boldsymbol{w}) = \sum_{v \in V} \chi_{A_i}(v) \log(F_v^{\boldsymbol{w}}(S)) + (1 - \chi_{A_i}(v)) \log(1 - F_v^{\boldsymbol{w}}(S)).$$

  算法基于下列优化问题的解输出影响力函数 $\boldsymbol{F}$：

$$\boldsymbol{w}^* \in \text{argmax}_{\boldsymbol{w} \in [\lambda, 1-\lambda]^m} \sum_{i=1}^{M} \mathcal{L}(S_i, A_i|\boldsymbol{w}).$$

这里采用基于覆盖数的 标准一致收敛 *standard uniform convergence* 论据来对关于估计参数 $\boldsymbol{w}^*$ 和真实参数的期望对数似然的差异进行界定。

Lemma D.5 (关于对数似然目标函数的样本复杂度保证)**.** 给定 $\varepsilon, \delta \in (0,1)$，当样本数 $M = \tilde{O}(\varepsilon^{-2} n^3 m)$ 时，以至少 $1-\delta$（在训练集的采样上）的概率，

$$\sup_{\boldsymbol{w}\in[\lambda,1-\lambda]^m} \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\boldsymbol{w})\right] - \mathbb{E}_{S,A}\left[\frac{1}{n}\mathcal{L}(S,A|\boldsymbol{w}^*)\right] \le \varepsilon.$$

- 最终，上述理论保证将通过代数变换转移到对估计参数 $\boldsymbol{w}^*$ 和真实参数的期望平方误差的差异的界定上。

PAC 可学习的结果表明在部分观测下的影响力函数学习没有信息论意义上的障碍。然而，这并不能直接给出一个 高效的算法。并且，对于经验风险最小化的目标函数的计算需要在所有的隐藏变量上进行边缘化 (marginalizing) 操作。严格意义上的 PAC 可学习性框架结果也还没有推广到 CIC 模型和其它传播模型上。这是由于缺乏 CIC 模型下影响力函数的简洁描述方法，而对于 IC、LT 模型行啊的影响力函数有这样的简单描述。因此，在下一章节将介绍非严格意义上的学习方法，目标是设计实用的算法并能在一大类传播模型上建立可学习性框架。

D.2 非严格意义的 PAC 学习算法

本章节将介绍可以进行影响力函数高效学习的非严格意义上的学习算法。这里不用以边的参数来参数化影响力函数，而采用一种 “模型无关 (model-free)” 的影响力函数学习框架，InfluLearner，由 Du 等人 [DLBS14] 提出，它将影响力函数表达成一些基函数的加权之和。下面考虑单独固定一个点 v 的影响力函数 $F_v(S)$ 。

影响力函数的参数化表达. 对于三种传播模型（IC，DIC，LT），传播过程可以用生存图 *(live-edge graphs)* 进行等价描述。具体的，[KKT03, DSRZ13] 的结果表明对于 IC，DIC 和 LT 模型的任意一个实例，都存在在生存图 H 上的一个分布 Γ ，每个H 对应一个产生该图的概率 γ_H 满足 $F_v^*(S) = \sum_{H:S \text{ 至少有一个点在} H \text{ 中有路径到达 } v \text{ in}} \gamma_H$.

为了简化表达，注意到如果只是为了激活 v，我么称两个不同的生存图 H, H' 是 “等价的” 仅当在 H and H' 中 v 恰好从相同的一些点出发可达.

因此，对任意点集 T，令 $\beta_T^* := \sum_{H:\text{恰好 } T \text{ 中的点在生存图 } H \text{ 中有路径到达 } v} \gamma_H$. 这里用描述向量作为特征向量 $\boldsymbol{r}_T = \boldsymbol{\chi}_T$，这里可以将 u 的分量解释成 u 在一个生存图中是否由一条路到 v。更精确地，令 $\phi(x) = \min\{x, 1\}$，并且 $\boldsymbol{\chi}_S$ 作为 S 的描述向量。那么，$\phi(\boldsymbol{\chi}_S^\top \cdot \boldsymbol{r}_T) = 1$ 当且仅当 v 从S 可达，那么有

$$F_v^*(S) = \sum_T \beta_T^* \cdot \phi(\boldsymbol{\chi}_S^\top \cdot \boldsymbol{r}_T).$$

这种表示有指数个特征项(每个 T 有一个)。为了让这个问题能够被解决，我们从合适的分布中采样少量集合 $\mathcal{T}$ 表示 K 个特征，相当于隐式地将所有其它特征地权重 β_T 都置为 0. 那么，所需学习地影响力函数被参数化成

$$F_v^{\boldsymbol{\beta}}(S) = \sum_{T\in\mathcal{T}} \beta_T \cdot \phi(\boldsymbol{\chi}_S^\top \cdot \boldsymbol{r}_T).$$

这里的目标就是学习采样出来的特征的参数 β_T 。(它们来自某个分布，例如，$||\boldsymbol{\beta}||_1 = 1$ 且 $\boldsymbol{\beta} \ge 0$.) 问题的关键是要表明足够少的 K 个特征 (例如，采样的集合) 能够进行较好的近似，并且这些权重可以从有限的 (不完全) 级联数据中有效地学习出来。特别地，我们考虑对数似然函数 $\ell(t,y) = y\log t + (1-y)\log(1-t)$，且通过如下极大似然估计问题学习参数向量 $\boldsymbol{\beta}$:

$$\begin{array}{ll}\text{Maximize} & \sum_{i=1}^{M} \ell(F_v^{\boldsymbol{\beta}}(S_i), \chi_{A_i}(v)) \\ \text{subject to} & ||\boldsymbol{\beta}||_1 = 1, \boldsymbol{\beta} \geq 0.\end{array}$$

不完全观察下的处理. 原始的极大似然估计不能直接应用到不完全级联数据上, 因为我们不知道 A_i (只知道不完全观测的 $\tilde{A}_i$). 要解决这个问题, 我们指导 MLE 实际上是一个实用对数误差的二元分类问题并且实用 $y_i = \chi_{A_i}(v)$ 作为标签。而不完全性观测可以考虑成标签上的类别条件噪音 (class-conditional noise)。令 $\tilde{y}_i = \chi_{\tilde{A}_i}(v)$ 作为 v 在不完全级联数据 i 中是否被激活的 *observation* . 那么,

$$\text{Prob}[\tilde{y}_i = 1 | y_i = 1] = r \quad \text{and} \quad \text{Prob}[\tilde{y}_i = 1 | y_i = 0] = 0.$$

换句话说,不完全观测 $\tilde{y}_i$ 相比完全观测y_i 只有有单边误差。已经有相应的技术来解决这个问题。

根据 Natarajan 等人 [NDRT13] 的结果 He 等人 [HXKL16] 用不完全观测 $\tilde{y}$ 可以构造一个 $\ell(t, y)$ 的无偏估计量,如下引理所示:

Lemma D.6 (文献 [NDRT13] 中引理 1 的结论)**.** *令 y 为 v 的真实激活信息, 而 $\tilde{y}$ 为不完全观测。那么可以定义,*

$$\tilde{\ell}(t, y) := \frac{1}{r} y \log t + \frac{2r - 1}{r} (1 - y) \log(1 - t),$$

对任意 t, 有 $\mathbb{E}_{\tilde{y}} \left[\tilde{\ell}(t, \tilde{y}) \right] = \ell(t, y)$.

基于这个引理,He 等人用稍作调整的似然函数 $\tilde{\ell}(t, y)$ 来进行极大似然估计:

$$\begin{array}{ll}\text{Maximize} & \sum_{i=1}^{M} \tilde{\ell}(F_v^{\boldsymbol{\beta}}(S_i), \chi_{\tilde{A}_i}(v)) \\ \text{subject to} & ||\boldsymbol{\beta}||_1 = 1, \boldsymbol{\beta} \geq 0.\end{array} \tag{D.1}$$

He 等人 [HXKL16] 分析了 (D.1)的解法, 提供了非严格意义上的 PAC 可学习性保证的条件, 这些条件可以适用于上述全部三个传播模型。这些条件与 Du 等人 [DLBS14] 工作中引理 1 的条件相似, 并且考虑了可达分布 (reachability distribution) β_T^* 的可近似性。特别的, 令 q 为在点集 T 上的概率分布, 对于所有点集 T 满足 $q(T) \leq C\beta_T^*$。令 $T_1, \ldots, T_K$ 为从分布 q 中采样的 K 个独立同分布 (i.i.d.) 样本。那么现在特征为 $\boldsymbol{r}_k = \boldsymbol{\chi}_{T_k}$。He 等人和 Du 等人都利用参数 λ 对影响力函数稍作修整 $F_v^{\boldsymbol{\beta},\lambda}(S)$ [2]:

$$F_v^{\boldsymbol{\beta},\lambda}(S) = (1 - 2\lambda) F_v^{\boldsymbol{\beta}}(S) + \lambda.$$

令 $\mathcal{M}_\lambda$ 为所有这样的修正版本的影响力函数的集合,并且 $F_v^{\tilde{\boldsymbol{\beta}},\lambda} \in \mathcal{M}_\lambda$ 为从优化问题(D.1)中获得的影响力函数。如下定理 (可在原始文献找到证明) 建立了所学习函数的准确性。

Theorem D.7 (参考文献 [HXKL16])**.** *假设学习算法在其构建的影响力函数中使用了 $K = \tilde{\Omega}(\frac{C^2}{\varepsilon^2})$ 个特征, 并且给定[3] $M = \tilde{\Omega}(\frac{\log C}{\varepsilon^4 r^2})$ 个保留率为 r 的不完全级联数据。那么, 以至少 $1 - \delta$ 的概率, 学到的对于每个点 v 和种子分布 $\mathcal{P}$ 的影响力函数 $F_v^{\tilde{\boldsymbol{\beta}},\lambda}$ 满足*

$$\mathbb{E}_{S \sim \mathcal{P}} \left[(F_v^{\tilde{\boldsymbol{\beta}},\lambda}(S) - F_v^*(S))^2 \right] \leq \varepsilon.$$

这个定理蕴含了:只要有足够的不完全级联数据,一个学习算法可以以任意的精度来逼近真实的影响力函数。因此,所有三个传播模型在不完全观测下都是非严格意义的 PAC 可学习的。最终的样本复杂度没有包含图的大小这一项, 但这一项在 C 中隐式地刻画了, 因为 C 和整个图以及分布 β_T^* 的近似好坏程度有关。注意到当 $r = 1$ 时, He 等人 [HXKL16] 的关于 M 的界是关于 C 的对数级别的而不是 Du 等人 [DLBS14] 中的多项式级别。这一改进的进一步解释可在原始文献中找到。

[2] 原文献中与定理 D.7相同的定理会展示如何选择 λ.

[3] $\tilde{\Omega}$ 记号省略了除 $\log C$ 以外的所有对数项,因为 C 可能是关于点数的指数级别或更大级别的数量

高效实现. 就像上面提到的,特征 T 无法从实际的可达分布 β_T^* 采样,因为这个分布是复杂的,难以完整刻画。为了从定理 D.7 获取有用的理论保证,He 等人 [HXKL16] 和 Du 等人 [DLBS14] 的方法是以从观测到的级联数据中估计边际分布,并以该边际分布的的乘积形式来近似分布 β_T^* 。

优化问题 (D.1) 是凸的,因此由凸优化的结论可知,该优化可以在以特征数 K 的多项式时间函数内解决。然而,需要较多特征来保证定理 D.7中的理论有意义。为了实用的高效算法,我们牺牲了一些理论保证,并实用固定数量的特征。

注意到优化问题 (D.1) 可以对每个点 v 单独的解决; 所学习到的函数$F_i(S)$ 可以进行组合得到 $\boldsymbol{F}(S) = [F_1(S), \ldots, F_n(S)]$。因为优化问题可以分解在每个点上单独解决,因此这个方法显然可以并行化,因此可以适用于大规模的网络。

在实现中还需要注意的是: 理论分析假设观测的保留率 r 是学习算法已知的。但实际上是未知的,而保留率可以使用交叉验证来估计。He 等人 [HXKL16] 也通过实验验证了其算法对保留率的选择并不敏感。