## Individualized Treatment Effect Inference

1、SCHWAB P, LINHARDT L, BAUER S, et al. Learning counterfactual representations for estimating individual dose-response curves[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(04): 5612-5619.

2、BICA I, JORDON J, VAN DER SCHAAR M. Estimating the effects of continuous-valued interventions using generative adversarial networks[J]. Advances in Neural Information Processing Systems, 2020, 33: 16434-16445.

## Learning counterfactual representations for estimating individual dose-response curves

## 苏黎世理工

# Importance of the application

* 医疗领域的一个主要挑战是确定某项治疗是否影响或决定了某项结果--例如，确定开具某种药物是否对生存有好处？以及不同的剂量如何影响作用于患者的效果。
* 目前的治疗指南是以随机对照试验为基础，考虑到 病人的平均情况下制定的，但不同的治疗方法会导致不同的效果和结果：对于任何特定的治疗方法，很可能只有一小部分人的实际反应与 "平均 "病人相似。从而制定的诊疗方案缺少对病人个性化的考量。

# Problems cannot be solved with existing methods

* 对于个性化治疗方案推断，数据集中的每个受试者都拥有一些潜在的结果：受试者在应用各种治疗时的结果，以及受试者在不应用治疗时的结果。治疗效果是两个潜在结果之间的差异，但由于我们只能观察到特定治疗分配的 "事实 "结果，而从未观察到相应的 "反事实 "结果，所以在观察性数据集中从未观察到真正的治疗效果的任何例子。这就是个体化治疗效果推断问题与标准监督学习（回归）的根本区别。

# Formal definition of the problem

给定个观察样本，有个治疗前协变量。对于每个样本，潜在结果是第个样本对个可用治疗方案中的一个治疗方案的反应，其剂量，其中和分别为治疗的最小和最大剂量。作为训练数据，作者收到事实样本和它们的观察结果，在应用特定的观察治疗和剂量之后。利用带有事实结果的训练数据，作者希望训练一个预测模型，以产生对所有可用的治疗方案在整个范围内的潜在结果的准确估计值。

# What are the technical challenges with this approach?

* 如何生成连续的剂量反应曲线
* 数据集如何处理

# Proposed Methods for these challenges

* 通过一个层次化的网络，分别对病人特征、治疗方案、治疗剂量进行处理。最终每个疗法的层下面对应多个剂量预测层，每个剂量预测层负责一部分区间内的剂量预测，最终得到总的剂量反应曲线。
* News由纽约时报语料库中的5000篇随机抽样的新闻文章组成，最初是作为反事实推断的benchmark引入的，其设置是两个没有相关dosage参数的treatment。作者扩展了原始数据集的规格，以便能够模拟任何数量的具有相关dosage参数的treatment。样本X是由词语计数xi、代表读者对新闻项目意见的结果y s,t和代表观众阅读时间的归一化dosage参数st∈ (0, 1]组成的新闻文章。treatment t对应于可用于查看新闻项目的各种设备，例如智能手机、平板电脑等。其中t的数量可以有不同的取值。作者在整个纽约时报的corpus上训练一个模型，对文本进行建模，模拟用户在特定新闻类型下，使用何设备进行观看的偏好，对新闻X，剂量st到治疗反应y的关系进行离散的建模，然后通过高斯混合模型得到连续的概率分布，作为最终的剂量反应曲线。

MVICU数据集对ICU中使用呼吸机的病人对不同呼吸机设置的反应曲线进行建模。对于本文，X为观测的体征，y为血氧饱和度，治疗方案t为呼吸机的不同参数，参数的大小代表了剂量参数st。作者同样对数据集进行混合高斯分布，由于MVICU数据集中X是结构化的，不同于News数据集对X进行建模，MVICU直接对X进行相似度计算。

TCGA数据集包含在9659个个体上收集到的不同癌症的基因表达数据，作者使用一个函数对不同治疗方案下的剂量反应曲线进行模拟。

DRNet结构包括base layer、个treatment layer和个head用于处理剂量的参数。base layer对所有样本进行训练，treatment layer只对属于其疗法的样本进行训练。每个treatment layer被进一步细分为个head layer（下图显示了第一组处理的个头层）。每个头层被分配一个剂量区间，将潜在的剂量范围细分为E个宽度相等的分区。每个head layer预测剂量参数值范围内的结果，并且只对属于其各自剂量范围的样本进行训练。

DRNets的分层结构使它们能够在所有样本（基础层）、治疗方案（治疗层）和剂量层（头部层）之间共享hidden representation，同时保留和相应的作用域。

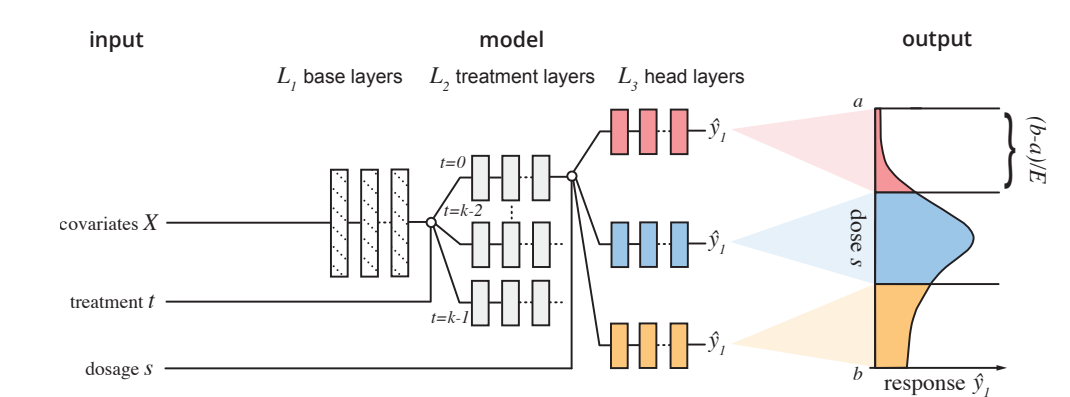


图1. DRNet的概述图

**Experiments and Datasets**

* 数据集

论文采用了三个数据集：News，MVICU，TCGA。其中用于验证的反事实的生成方式上文已表，不再赘述。

表格

描述已自动生成

图2. 数据集统计信息

* 实验
  + 指标

为了衡量一个模型在多大程度上覆盖了个体剂量-反应曲线的整个范围，作者使用真实剂量-反应和预测剂量-反应之间的，在个样本、所有疗法和整个剂量范围中计算得到。

为了进一步衡量模型为每个个案确定最佳剂量点的能力，作者计算了平均剂量决策误差（）。平均剂量决策误差是指在个样本和所有治疗方案T中，预测模型估计的最佳剂量点与真正的最佳剂量点间的平均平方误差。

决策误差()衡量模型确定个体病例最佳治疗决策的能力，即使用估计的最佳治疗方案与真正的最佳治疗方案和剂量相比，结果会差多少。是根据ground truth y得出的最佳治疗方案，是预测模型得出的最佳治疗方案。

将和一起考虑，对于全面评估反事实推断模型非常重要。例如，如果一个模型除了最佳疗效区间以外，都能很好地拟合剂量反应曲线，就能获得一个较低的MISE，因为MISE关注整体的剂量反应曲线拟合效果。但对于确定特定病人最佳结果的治疗和剂量选择来说，就不是一个好模型，导致PE的值较高。通过考虑多个指标，作者可以确保预测模型既能恢复整个剂量反应，又能选择最佳治疗和剂量选择。一般来说，在不了解结果产生过程的情况下，作者无法计算出MISE、DPE或PE，因为真正的剂量-反应函数是未知的。这就需要在数据集中进行数据的合成。合成方式上文已表，不再赘述。

* + 性能测试

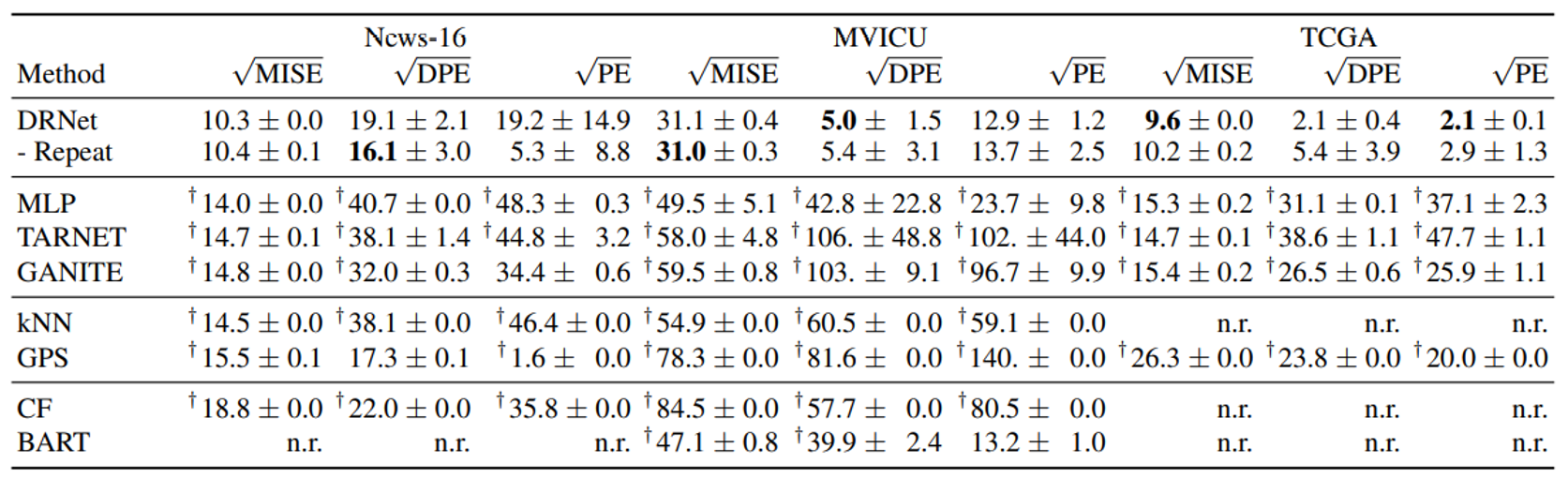


图3. 整体性能比较

相较于baseline，DRNet在MISE，DPE，PE三个指标上都获得了明显的提高，证明了模型生成反事实对决策起到了帮助效果。

# Differences of this Method compared to others

作者提出了一种训练神经网络进行反事实推断的新方法。与现有的方法不同，该方法适用于估计任何数量的治疗方案和连续剂量参数的反事实结果。

**Limitation of this work or your discovery**

因为观察性数据集是基于不可检验的假设，所以从观察性数据集中估计因果效应的方法存在着限制，必须进行数据的合成。在实践中，尤其是当人们对基本的因果过程没有太多的先验知识时，做出合理的合成是很困难的，导致最终反事实的结果可能与现实出入甚大，不利于最终的决策辅助。

## Estimating the Effects of Continuous-valued Interventions using Generative Adversarial Networks

## (牛津大学，UCLA)

# Problems cannot be solved with existing methods

DRNet使用多头layer对区间内的剂量反应曲线进行预测，然而多头layer的设置是固定的，无法动态确定取值范围，因此失去了大部分的灵活性。作者使用GAN生成反事实，不同于DRNets的监督学习方法。通过实验证明SCIGAN优于DRNets。

# Formal definition of the problem

与上文相同，目标是从观察性数据中学习，如电子健康记录，其中包含关于病人特征的事实的治疗方案和分配给他们的相应剂量，代表一个事实治疗。最终训练神经网络得到不同治疗方案下的剂量反应曲线。

# What are the technical challenges with this approach?

* + 数据应该符合怎样的约束
  + 如何生成连续的剂量反应曲线

# Possible methods for this challenges

* + 为了能够从观察数据中确定这些潜在的结果，作者需要做出以下两个假设：重叠和无隐藏的混杂因素。重叠是指每次每个病人接受每种治疗的概率不为零，无混杂因素是指需要观察到所有影响治疗方案分配和病人结果的变量
  + 要注意的是，无隐藏的混杂因素是一个很难实现的假设，在实践中无法检验，与上文同样的，只能通过模拟去实现近似结果。
  + 因为只有事实结果存在于数据中，而反事实结果从未被观察到，如果要得到最佳的治疗方案，需要在多种治疗方法中进行选择。这个问题在干预措施是连续取值的情况下更加棘手，因为反事实的数量是无限的。SCIGAN采用GAN，可以用来估计这种连续的干预措施的效果。生成器用来生成反事实的剂量反应曲线，判别器作用于每条生成的曲线上的有限点集，并从实际结果集和生成的反事实中判断是否为事实结果。

图表

描述已自动生成

图4. SCIGAN模型架构

反事实生成器将病人的特征x，事实结果yf，接受的治疗类型wf和剂量df以及一些噪音，并输出每种类型治疗的剂量反应曲线。对于生成器，类似于DRNet，采用了一个多任务模型，首先输入共享层，然后对于每个疗法w，引入了一个多任务头，再输入剂量参数d产生结果。不同于DRNet对区间进行等分，SCIGAN的反事实生成器将d看作一个整体。

判别器作用于从每个生成器生成的剂量反应曲线中的几个随机采样点。对于每个治疗w，比较nw个剂量样本。详细的，对于每个治疗w，作者从剂量域中抽取nw个剂量水平。然后，使用生成器来获得每个治疗w的剂量样本的反事实结果，作者形成这里所示的剂量结果对。对于事实治疗wf，作者用事实剂量df替换生成器的输出样本，用事实结果yf替换生成器的结果。 这样作者就得到了一组所有治疗类型w的剂量结果对，并为作者提供了判别器的输入。

推理网络具有与生成器相同的结构，但由于推理时所拥有的数据仅仅为病人的特征X，所以推理网络只接收病人的特征作为输入，然后估计每个病人个性化的剂量反应曲线。

**Experiments and Datasets**

* 数据集

对于数据集，由于无法观察到所有的反面事实，治疗效果估计问题无法在真实世界的数据集上进行有意义的评估。因此作者提出了自己的半合成的数据模拟模型。使用真实的数据集TCGA,NEWS和MIMIC3。

然后考虑三种治疗，每种治疗都有相应的剂量，每种治疗类型w都有一组参数v，每个实验都是随机抽样。 每个治疗fw的剂量反应曲线，以及对应最佳剂量的闭式表达如图5所示。这里治疗效果与病人的特征相关。例如在第一个式子中，第一个项可以代表提高化疗剂量和减少肿瘤大小的疗效，而二次项可以代表随着剂量的增加，化疗的毒性增加。

表格

描述已自动生成

图5. 数据合成

* 性能测试

相较于之前的SOTA（DRNet），SCIGAN在MISE与PE指标上得到了显著的性能提升，体现了模型的有效性。

表格

描述已自动生成

图6. 整体性能比较

* 消融实验

消融实验可以看到，本文中提出的监督损失、多任务层、层次化判别器以及互换不变和互换等效层是有效的。

表格

描述已自动生成

图7. SCIGAN消融实验

# Differences of this Method compared to others

DRNets基于神经网络来估计多种治疗和连续剂量的反事实。DRNets由一个三层结构组成，包括所有疗法的共享层、每个疗法的multi-head layer和一个用于预测该疗法下剂量反应曲线的multi-head layer。具体来说，对于每个治疗w，剂量区间[a,b]被细分为E个同等大小的子区间，E个head分别计算该区间内的剂量反应曲线。然而，作者认为对剂量区间使用多任务头的主要优点是增加了模型的灵活性，可以在剂量区间的不同区域学习潜在的非常不同的功能。DRNets不能动态地确定这些区间，因此失去了大部分的灵活性。作者使用GAN生成反事实，不同于DRNets的监督学习方法。

**Limitation of this work or your discovery**

总的来说，个性化治疗预测能够更好地估计个体化对剂量的反应，将有助于我们选择能使病人结果得到改善的治疗。此外，临床医生和病人往往需要考虑几种不同的结果（如潜在的副作用）；对这种结果进行更好的估计，可以使病人做出更明智的、适合他们的决定。

作者认为，这篇工作的限制如下。SCIGAN至少需要几千个训练实例，因为GAN通常需要较多数据来支持收敛，这通常在医学领域是缺少的。此外，进一步的工作将考虑如何纳入有关这些响应函数的先验知识。作者认为也可以考虑如何将工作扩展到时序数据的建模上。

而且，比较重要的一点是，本文提出的工作只是理论上的，evaluation部分采用的是合成的数据集。在这种方法能够用于实践之前，还需要进行大量的测试，例如临床试验。特别是错误的估计会威胁到生命，面临道德风险。

通过确保此类模型作为支持系统与临床医生一起使用，而不是代替临床医生，可以大大减轻错误分配治疗的风险。但是我认为，此类模型作为支持系统的情况下，对医生可能会引入决策上的偏差，这类支持系统对辅助诊断的有效性可能还需要心理学、行为学上的更多论证。而且作者认为，从另一方面看，由于决策者可以获得各种疗法的剂量反应曲线，希望造成伤害的人可以利用估计的结果来选择最坏的结果。