协同披露Synergistic Disclosure

协同披露是一种保证完美样本隐私的数据披露技术，因为它揭示了整个数据集的特征，但不涉及其任何组成元素。协同披露的目标和普通信息披露相同，都是要推断数据集的统计信息和潜在特征，不同的是它既能提供隐私又能提供实用性，同时又不会违反与数据隐私相关的条件和要求。

整个过程为了避免数据集的信息在统计信息和潜在特征时泄露，生成了一个与数据集整体相关，但与个体数据信息完全独立的中间量，通过中间量同样可推理出数据分析想要的结果，但不会暴露个体信息的隐私。该过程中，将中间量与结果统计信息的最优互信息定义为协同披露信息量，作为协同披露的度量方法。[1][2]

1. 历史

1.1 数据披露 Data Disclosure

数据披露主要是指数据持有者、收集者以某种形式，把数据反映的相关的信息，向社会公众公开披露的行为。数据披露是让数据背后的信息与使用者连接的桥梁。

例如：投资者和社会公众对上市公司数据信息的获取，主要是通过大众媒体阅读各类临时公告和定期报告。投资者和社会公众在获取这些数据信息后，可以作为投资抉择的主要依据。公告和报告就可以认为是数据批量的一种途径。

真实、全面、及时、充分地进行数据披露至关重要，只有这样，才能使数据真正发挥价值，让背后的潜在信息对个人甚至社会，产生有效地帮助。 [1][2]

1.2 数据隐私 Data Privacy

数据隐私（data privacy）是2018年公布的计算机科学技术名词。可以定义为：

（1）数据中直接或间接蕴含的，涉及个人或组织的，不宜公开的，需要在数据收集、数据存储、数据查询和分析、数据发布等过程中加以保护的信息。

（2）保护数据隐私的能力，通常采用数据匿名化、数据扰动、数据加密、差分隐私等技术。

协同披露就属于要在保障数据隐私的同时，完成数据披露的一个重要手段。

2. 基本条件

原始数据集用表示，其中表示组个体数据，统计信息和潜在特征用集合表示，中间量用表示。[1]

2.1马尔科夫性质

协同披露中，数据集，潜在特性，中间量之间存存在马尔可夫动力学的性质。

所谓的马尔可夫动力学是指系统的下一时刻状态只依赖于上一时刻的状态，并且与再之前的状态无关。马尔可夫动力学可以区分为离散时间、连续时间，离散状态、连续状态，以及它们的组合等多种。

由于可以通过直接数据披露产生，故可以定义，于对有马尔可夫依赖。由于是从数据集中变化产生，故对也存在马尔可夫依赖。因此可以用，和两个条件表示对，对的映射关系。这样就可以构成一条马尔科夫链。[1][3]

2.2 中间量的随机映射

是中间量关于数据集的映射，表示由于是从数据集采中变化产生。由于中间量可推理出数据分析想要的结果，但不会暴露个体信息的隐私，故我们需要定义，表示相互独立，在信息论中可以用互信息为0表示，即。

所有可产生满足的映射，可以写成一个集合。[1]

2.3 中间量与所需指标的非独立性

与直接可以产生协同披露，还需要满足，即中间量要和所需要的指标与潜藏信息存在关联，若，则通过无法判别出任何的信息。若无法产生任何满足条件的中间量，那么协同披露也无法完成。[1]

3. 协同披露的度量方法

术语“协同”来自这样一个事实，即协同披露映射揭示了整个数据集的集体属性，这些属性不会损害其“部分”（即样本本身）。为了度量协同披露的效果，引入协同披露信息量的概念，协同披露信息量定义为：

表示在所有中的映射所产生的中，寻找与的最大互信息，该互信息被称为协同披露信息量。[1]

4. 协同披露的性质

收敛性：协同披露能力，即，随着的增长而收敛。

5. 协同披露的应用

5.1在信息保护中的应用

协同数据披露方法，允许公开数据集的集体属性，同时对每个数据样本的值完全保密。隐私和效用都可以通过利用多元统计的反直觉特性来实现，这允许变量与随机向量相关，同时独立于每个分量。[1][2]

5.2在因果涌现中的应用

协同披露提供一种直接的路径来检测涌现情况，而无需重复计算冗余。

数据集可以视作微观态，中间量可以视为宏观态，映射可以视作粗粒化过程。协同披露信息量可以视为因果涌现的度量。从推测可以视为向下的因果关系。

因果涌现特征的预测能力超出了单个成分。当预测力涉及到个别元素时，就会产生向下因果关系；当它涉及自身或其他高阶特征时，会出现因果解耦。[3][4]

参考文献

[1] Rassouli, B. , Rosas, F. E. , & Gunduz, D. . (2019). Data disclosure under perfect sample privacy. IEEE Transactions on Information Forensics and Security, PP(99), 1-1.

[2] Rassouli, B. , Rosas, F. , & Gunduz, D. . (2018). Latent Feature Disclosure under Perfect Sample Privacy. *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE.

[3] Rosas FE, Mediano PAM, Jensen HJ, Seth AK, Barrett AB, Carhart-Harris RL, et al. (2020) Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. PLoS Comput Biol 16(12): e1008289.

[4] Williams, P. L. , & Beer, R. D. . (2012). Nonnegative decomposition of multivariate information. *Physics,* *1004*.