

抄読会

機械学習を用いた様々なデータベースの CDISC 標準への変換に関する研究

博士課程 3 年 小田拓馬

概要

医薬品の承認申請時に各国の規制当局への CDISC 形式でのデータ提出が今後義務化されていく。このため、製薬企業やアカデミアは承認申請のために CRF data や Legacy data を CDISC 形式に変換しなければならない。また、この他にも市販後調査データ、MID-NET をはじめとした医療情報データ、電子カルテなど様々な医療データが様々な形式で保持されている。これらを CDISC 標準という一つの形式にまとめることができれば、統合可能であり、今後の医薬品開発のためにも非常に有用である。そこで今回、機械学習を利用して上記のような医療データを効率的かつ高品質に CDISC 形式に変換することを目的とし、研究を行うこととした。

まず、上記の医療データの中でデータの入手の容易さとデータバリエーションの観点から、CRF data と Legacy data の SDTM 形式への変換を対象データとすることにした。

今回、上記研究の進捗と今後の予定について報告を行う。