

製薬会社におけるデータサイエンス事例

アステラス製薬株式会社 リアルワールドインフォマティクス機能
アナリティクス&インフォマティクス室
角山和久

新しいくすりを創成する「創薬」には、長い期間と莫大な費用が必要です。その期間は9~17年、1品目あたりの開発費用は数百億円以上かかると言われ、また、その成功確度も極めて低く、研究を始めたくすりの候補品が新薬になる成功確率は3万分の1と見積もられています。その過程では多様な基礎研究と各種試験がおこなわれますが、この期間を短くして確度をより高めることが求められます。一方において、近年における情報科学技術の革新的発展は、いわゆるビッグデータ時代をもたらし、医療分野にも大きな影響を与えています。製薬会社は創薬にビッグデータを活用して、より短期間で確度の高い創薬をおこなおうとしています。

例えば、ヒトゲノム解読処理技術が進化したことにより、国家レベルのヒトゲノム情報の収集がおこなわれています。このデータを利用して疾患の遺伝的要因を分析することで、確度の高い創薬標的タンパク質の同定や、薬剤効果の差を予測できるバイオマーカーの探索がおこなわれています。特にがんに対する創薬においては、国際コンソーシアムによって大規模にがんゲノム情報の解読がおこなわれ、一般にもデータは公表されているため、数千人から得られたゲノムデータを活用した研究が盛んにおこなわれています。

世界中でおこなわれる日々の研究成果は論文としてまとめられますが、その情報量は膨大です。これを素早く分析して創薬に関する知見や新しい知見を取り出すためには、テキストマイニングと呼ばれるデータ分析技術を用います。例えば、疾患とタンパク質の隠れた関係を素早く見出すことが可能です。論文データのテキストマイニングは、新たな創薬標的タンパク質を見つけ出すだけでなく、ある疾患に対する標的が別の疾患へも応用できることを見出す、ドラッグリパーピングと呼ばれる目的にも用いられます。

くすりの候補品の創出や選出にもデータ分析がおこなわれています。低分子化合物の創薬の場合には、数百万規模の化合物ライブラリーを用意して、候補品の種に成り得るものを実験的に見出すハイスループットスクリーニングが一般的におこなわれますが、規模の拡大には限界があります。そこで、より少量のライブラリーを用いて短期間に候補品の種を得ることができるよう、データマイニングや機械学習の手法を用います。或いは、ハイスループットスクリーニングを使わずに、計算機によるシミュレーションのみで候補品を得ようとすることもあります。いずれの場合にも、研究者による化合物合成と薬理的な評価を繰り返しますが、そこでも薬効予測、体内動態、或いは毒性予測にデータマイニングや機械学習を活用して研究スピードの向上を目指します。

ヒトに対する臨床試験、即ち、治験においては、そのデザインから始まり、治験の実施、データマネジメントと統計分析を経て、開発品のヒトへの有効性と安全性のデータを得ますが、どの場面においても統計的な知識が必要とされます。プラセボ（有効成分を含まない薬剤）や対照治療よりも自社開発品が臨床的・統計的に優れていることを示す優越性試験以外にも、自社開発品の有効性が対照薬よりも劣らないことを示す非劣性試験もおこなわれます。治験は莫大な費用と長い期間を

必要としますので、効果が低い、或いは、安全性に問題がある場合は早めにこれを中止し、良いものとなるべく早く次の試験フェーズに進めさせることが必要です。これには中間解析を伴う群逐次検定法やアダプティブ・デザインなど、新たな技法が積極的に導入されています。

リアルワールドデータと呼ばれる、臨床現場で得られるデータも医療ビッグデータの一つとして活用されています。医療現場や患者の実際の動向から、創薬の対象とする疾患をより深く理解して、層別化されうる患者群のアンメットニーズを抽出します。臨床試験は先述の通り膨大な費用と期間を要しますが、これをより効率的におこなえるように治験に最適な場所や組み入れられる患者数を特定、予測します。また、より効果的な営業戦略の立案や、上市品の市販後の調査、医療経済評価にも用います。ここでも多くの数理統計やデータマイニングなどのデータサイエンス技術を利用します。

最近では、生体情報のセンサリング技術やモバイル技術が発達したことから、患者や健常人の日々の環境データやヘルスデータを蓄積、分析することが可能です。これまでは疾患を漫然と一つの患者集団として捉えた創薬がおこなわれてきましたが、先に述べたゲノム情報をはじめとして、これらの個人レベルの詳細な情報を大量に得ることで、遺伝要因と環境要因の相互関係を分析することが可能になり、患者層別化による疾患の精密な診断方法や、創薬を含む治療法、更には予防法を研究することができます。この場合においてもデータマイニングや機械学習のデータ解析手法が多く活用されていることは言うまでもありません。

創薬研究に直接活用できるデータ以外にも、ビッグデータ時代にいる我々は他の多くのデータを分析することができます。統計的シミュレーション等のデータサイエンス技術は、経営・製品戦略の策定にも活用できます。例えば、各製薬会社の薬剤開発の進捗やその概要は各社から公表されています。また論文でも開発医薬品の各種試験での成績が発表されています。これらをまとめて分析することは、自社優位な戦略を考案することに有効です。製品の価値を高めるためには、或いは、自社の製品が他社との競争において有利な価値を創出できるよう自社リソースを最適に配分するためには、いくつかのシナリオを想定したシミュレーションをおこない、研究や開発プロジェクトの加速や継続、中止が定量的な評価によって決定されるべきです。

コンピューティング能力や情報分析技術が日々改良され、情報科学やデータサイエンスの進展が大きく注目されるようになりました。長い間冷遇されていた人工知能研究も GPGPU を用いたディープラーニングの登場により復活し、特に画像分析の分野においては、医師による画像診断の精度を上回る能力を示し、医療分野に大きな衝撃を与えています。もちろん製薬会社においても、急激に発展する人工知能の技術を活用しようと現在盛んに応用がおこなわれています。

本発表では、以上の様に製薬会社においておこなわれている種々のデータサイエンスの事例の中から複数を紹介したいと思います。また、今後に求められるデータサイエンティスト像についても私見を述べたいと思います。