

CBRC Newsletter

<http://www.cbrc.jp>

A
U
T
U
M
N
2
0
0
5
13

生物学と機械学習の交流から得られるもの



津田 宏治
Koji TSUDA

数理モデルチーム
研究員

私は産総研に入所して7年目ですが、その内の3年をドイツで在外研究*をして過ごさせていただきました。このような生活の良い面は、様々な価値観を持つ人と出会い共同研究ができるという点です。調べてみると、私はこれまで25人と共著の論文を書いています。ただ単に日本に留まっていたならば、この数はもっと少なくなっていたと思います。

二つの国を往復したことも、視野を広げるのに役立ちましたが、CBRCに参加し、生物学の世界を体験できたことは、純粋に計算機科学の人間だった私にとっては、非常に大きい出来事でした。

生物学データに使われている数々の手法を勉強するうち、バイオインフォマティクスの手法は、他の計算機科学の分野(機械学習、画像処理、テキスト処理など)と

交わることなく、独自の発展を遂げていることがわかってきました。例えば、自然言語処理と、ゲノム配列処理は、記号列を処理するという点で共通点が多いにも関わらず、バイオインフォの論文には、自然言語処理の論文が引用されることなどほとんどありません(逆もない)。そこで、CBRCに入った当初の私は、バイオインフォと機械学習の間にポジションを取って、互いの手法を流通させ、統合することによって、いい研究ができるのではないかと思っていました。

そのような試みの全てが成功したわけではないし、思っていたよりも難しかったのも事実ですが、小さな成功例もいくつかあります。例えば、3年前に提案したMarginalized kernelという手法は、元々は、隠れマルコフ

モデルに基づいて、二本のDNA配列の類似度を計算するという方法でしたが、近年、センサーアレイや、ロボティクスの分野にも応用されています。またカーネル行列補完の研究も、バイオと、それ以外の分野の両方で応用されています。

強調したいのは、つまるところ生物学的なデータにしか使えない手法などないということです。どんなに生物学的な知識を入れて、ある対象に特化させたとしても、結局は数字の処理をしているだけなので、バイオ以外の対象にも必ず適用できるはず。皆さんは、目の前のデータの処理に忙しいことと思いますが、時間があるときに、他の計算機科学分野との関連を探してみると、新しい発想が生まれるかもしれません。

* Max Plank Institute for Biological Cybernetics, Tübingen, Germany

P.1
エッセー

P.2
トピックス(シンポジウム)

P.3
研究紹介(長崎)

P.4
研究員紹介・お知らせ