

Name : 鳥山 昌幸
Title : 機械学習によるグラフ上でのデータ解析手法に関する研究
Institute : 京都大学化学研究所バイオインフォマティクスセンター
Partner institute : Centre for Computational Biology, Mines ParisTech
Duration : 2013年10月1日～2013年12月26日
Report: はじめに 今回の滞在ではフランスのMines ParisTech に所属する Jean-Philippe Vert 教授のグループの研究室に2013年10月～12月末までの約3か月間滞在させて頂いた。本稿では3か月の滞在期間中のレポートを「研究室での活動」、「研究の進捗」、「日々の生活」の各々についてまとめ、最後に全体の総括を述べる。 研究室での活動について Vert 教授の研究室は、機械学習と呼ばれるアプローチを用いた生物学データの解析を主なテーマとしている。機械学習とはデータからその背後に潜む規則性を自動的に獲得するための手法を研究する情報学分野の一つであり、各分野で爆発的に増加し続ける膨大なデータの解析手法として広く注目されている。私自身は機械学習の基礎方法論を長く主として行ってきたが、その生物学データへの応用に近年トライし始めていた。そのため、機械学習の方法論に重点を置きつつも、実践的な生物学データへの応用も守備範囲とする Vert 教授の研究室に興味を持ち今回の滞在に至った。 研究室は十数名程度のグループであり、主に博士課程の学生とポスドクから構成され、情報学や統計学、数学、生物学など様々なバックグラウンドを持っており、各々の得意分野を中心としつつもお互いに協力しつつ研究を進めていた。研究室としての主な活動は週一回のミーティングと機械学習のテキストを用いた勉強会であり、どちらにも参加させて頂くことができた。ミーティングでは各メンバーの進捗報告に加えて、メンバー(あるいはゲスト)の研究発表、ジャーナル論文誌の輪読会のいずれかが行われた。研究発表では進行中の私の研究についても発表の機会を頂くことができ、教授やメンバーから意見を聞くことができた。ジャーナルの輪読会は毎回メンバーのうち一人がある一つの論文誌について最近の論文をいくつか紹介するという形式で行われた。論文誌はNature などの超有名雑誌に加えて、例えば統計学分野での有名誌(Annals of Statistics)などが取り上げられていた。先に述べた通り様々なバックグラウンドの研究者が集まっているので、メンバーが自身の分野での最新トピックを紹介してくれることで効率的に各分野の話題をチェックでき、メンバー内での情報の共有にもなっているようであった。勉強会は機械学習の有名テキスト(The elements of statistical learning, T. Hastie, et al., 2009) を用いて行われた。このテキストは私にとっても親しみのあるテキストであり、こちらも楽しく議論に参加させて頂くことができた。 その他の研究室関連の活動として Vert 教授が主催者の一人として参加している Statistical Machine Learning in Paris (SMILE in Paris) というセミナーを聴講することができた。このセミナーには Vert 教授のグループのみならず、近隣の大学の機械学習関係の研究者たちが集まりセミナートークが行われた。行われたセミナーの中には機械学習分野の有名研究者である Francis Bach 教授による最適化数学の講義や、機械学習の最高会議である NIPS (Neural Information Processing Systems) の論文紹介などがあり、どれも私にとって興味深いものであった。いずれのセミナーでも多くの参加者が教員、学生に関わらず積極的に議論に参加していたことは印象的であった。

また学生でも英語力は十分に高く、疑問点などについては妥協せず理解に辿りつくまで追及する姿勢がよく見てとられた。

研究の進捗について

今回私は「機械学習によるグラフ上でのデータ解析手法に関する研究」と題したテーマに取り組んだ。この研究では「グラフ」として表現される情報を機械学習の手法の中でどのように扱うかを考える。タンパク質相互作用ネットワークや代謝ネットワークをはじめとして生物学のデータにはグラフとして表現されるデータが多く存在し(図1),そのようなデータからどのように潜在的な相互作用や協調関係を見出すかが非常に重要な課題となっている。特に重要性が高いのはグラフ情報に加えて数値情報が付随するような場合である。これには例えばタンパク質相互作用ネットワークと対応する遺伝子発現量データを合わせて解析するような場合がある。この時、特定の状況下においてネットワークのどの部分が協調的に変化しているのを知ることは生体内のメカニズムの理解の鍵となる(図2)。今回はこのような問題に対してグラフィカルガウシアンモデルと呼ばれるガウス分布にグラフが表現する依存関係を導入したモデルと、その因子分析を用いて、グラフに存在する局所的な変動構造を効率的に発見するアルゴリズムを考えた。

方法論については予めある程度設計しておいたため、その内容を研究室のミーティングで発表させて頂き議論を開始した。Vert 教授からモデルの解釈や実験比較に関するアドバイスを頂きつつ、手法の洗練化と、遺伝子発現量とタンパク質相互作用ネットワークによる実験を進めた。最終的に、提案するアプローチによってタンパク質ネットワークから活発に相互作用を起こしていると推定される部分ネットワークを抽出し、抽出された部分ネットワークが一般的な手法と比較して多くの(各遺伝子に対応付けられた)アノテーションを共有していることを示唆する結果を得た。滞在中に完了することはできなかったが、帰国後この内容について国際会議への投稿論文の執筆を行っている。

日々の生活について

今回の3か月の滞在は私にとって過去最も長い海外への滞在となったため、研究以外の日々の生活全体でもよい経験ができた。人や街の文化や慣習の違いに最初は戸惑う(あるいは気づけない)こともあったが、時間が経つほどにそういった事にも多少は慣れていけた。日本での所属研究室にも海外からの短期滞在研究者は頻りに訪れるため、今回の経験を彼らとの接し方にも活かすことができるのではないかと思う。また余暇を観光に充てることもできたため、公私ともに充実した時間を過ごすことができた。特に、言わずと知れた「芸術の都」であるパ

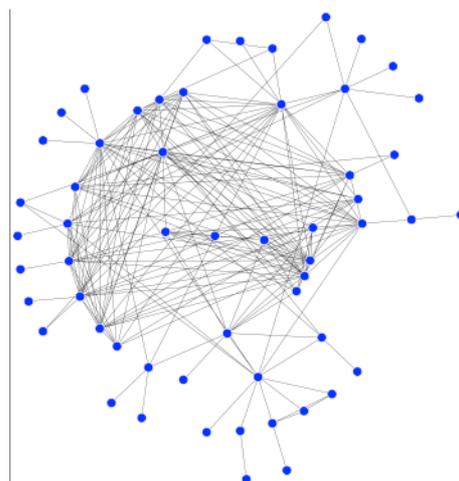


図1 グラフの例. タンパク質相互作用ネットワークの場合、丸い点(ノード)がタンパク質に対応し、点間の直線が相互作用の存在を意味する

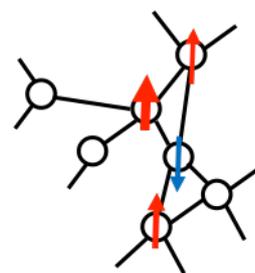


図2 グラフ上の局所的な共分散構造の概念図。(実際にはより巨大な)グラフの一部で各ノード(例えばタンパク質)が協調的に変動(増減)する。

若手研究者インターナショナル・トレーニング・プログラム(ITP)

バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム 報告書

りの美しい世界遺産の数々はどれも心に残るものであった。滞在も終わりに近づいたクリスマスには研究室のメンバーでパーティが行われ、私も参加させて頂くことができた。短い期間ではあったが、これを機にメンバーたちとの親交も続けていければこの滞在の意義もより深まるだろうと思う。

総括

今回の滞在は私にとって初めての海外研究室でのまとまった期間の滞在であり、多くのことを経験することができた。海外研究者達の研究のやり方や姿勢など取り入れるべきものは取り入れ、この経験を今後の研究生活にフィードバックすることでこの滞在の価値を高めていけるよう努めていきたい。

最後に、今回の貴重な機会を与えて頂いた馬見塚拓教授、滞在を快く承諾していただき、期間中もすべてにおいてサポートして下さったJean-Philippe Vert教授、また研究室のメンバー、スタッフの方々に感謝致します。

