

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 応募書類

Name: 志賀元紀
Title: 疾患遺伝子同定のためのデータマイニングの開発
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 of your choice : Department of Bioengineering, Boston University
Duration of your choice: 6月15日～9月2日
Plan : 滞在の目的 共同研究する予定であるボストン大学 Simon Kasif 教授の研究室 Computational Genomics Laboratory では、私の所属するバイオインフォマティクスセンターと同様に、コンピュータを使用したゲノムおよび生物データの解析法を研究している。Kasif 教授は計算機科学のバックグラウンドを基礎として生物学の解析法の研究に従事しており、また、研究室のメンバーは、生物系の研究者と計算機科学の研究者の両方が所属している。私と Kasif 教授のグループのバックグラウンドが大きく重なるために、研究に関する討論を行い易いと考えている。 また、私の所属する馬見塚研究室と大きく異なる点は、我々の研究室ではデータマイニングおよび機械学習の研究に重点をおくのに対して、Kasif 教授のグループは生物データの解析に重点をおくことである。そのため、ボストン大学内またはハーバード大学医学部などのウェット研究室と積極的に共同研究を行っている。 上記の2つの利点を元にして、この度の研究滞在では、現在の自分の研究（下記に記述する）の発展および関連する新たな研究テーマの見出すこと、また、新たな共同研究パートナー、特に、ウェット研究に関わるものとの関係を築くことを目的とする。 研究計画 私の主な研究テーマは、多様な生物データ解析のためのデータマイニング法の開発である。最近の成果としては、遺伝子発現量と遺伝子ネットワークを統合するクラスタリング法を開発し、ISMB/ECCB や KDD といったバイオインフォマティクスまたはデータマイニングのトップレベル(採択率が約15%)の国際会議に研究発表している。データマイニングの側面からの現在の課題は、これらの手法をより一般的な方法へと拡張し、よりロバストな解析法を開発する事である。また、生物学的な側面では、より具体的な実データ解析への応用である。現在は、疾患データの開発や創薬に関わる応用研究を考えている。 ヒトの塩基配列が解読された後のポストゲノムと言われる昨今では、遺伝子に関わる情報、例えば、遺伝子発現量、タンパク質相互作用、SNP（一塩基多型）などを、安価かつ高品質に取得可能となっており、疾患メカニズムの解明への研究が益々加速している。特に、個体単位でのゲノム情報が入手できるようになったために、癌や糖尿病などの複雑かつ多因子性の疾患の研究が注目されている。こうした背景から、ボストン大学では、疾患関連遺伝子群の同定法に関する研究を行いたい。

Plan (Continued)

近年、Kasif 教授らの研究グループでは、複雑な病気の遺伝子解析法を開発しており、最も興味がある研究成果として、論文

Liu M, Liberzon A, Kong SW, Lai WR, Park PJ, Kohane IS, Kasif S.

Network-based analysis of affected biological processes in type 2 diabetes models.

PLoS Genet. 2007 Jun 15; 3(6): e96.

が挙げられる。Liu らによる手法では、Gene Set Enrichment Analysis(GSEA)と呼ばれる疾患関連遺伝子の同定法を、よりロバストにするため、タンパク質相互作用 (PPI) を用いる手法 Gene Network Enrichment Analysis (GNEA)を開発しており、開発手法をII型糖尿病の解析に応用している。彼らの手法は2型糖尿病の解析のみに応用しているものの、方法自体は、癌などの他の疾患に応用できる汎用的かつ重要なものである。Liu らの手法は、まず、正常細胞と疾患細胞の遺伝子発現量差をスコアリングして、スコアが平均的に高い PPI の部分ネットワークを抽出し、そして、部分ネットワークと関連する遺伝子集合 (Gene Ontology term など) を同定する手法である。この手法の主要な問題点は、直接的に解けないために、他者の開発したソフトウェアを用いるヒューリスティックな方法の組み合わせで問題を解く方法を提案している。私の研究目的は、統計的なモデルを用いて、一貫性のあるアルゴリズムの開発、および、よりロバストな手法を構築することである。

そのために、まず、我々の開発法、遺伝子発現量とネットワークを統合する統計的モデルおよび最適化手法を改善することから始める。私の手法の主要問題は、1)アルゴリズムの実行速度、2)悪質な局所解に陥り易いこと、3)クラスタ数などの入力パラメータが多いこと の3点である。これらの問題を改善するために、ベイズ統計モデルを用いた学習法を導出する。問題1の解決は難しいものの、問題2と3の解決を確実に期待できる。具体的な手法は、Hofman (Phys. Rev. Lett. 2008)の拡張として導出できるものと考えている。そして、このモデルの学習結果を用いて、正常細胞と疾患細胞の遺伝子発現量の差を統計的にスコアリングする手法を構築する予定である。研究を進める際に、上記の論文の著者や Kasif 教授と積極的に討論して、研究を迅速に進めてゆきたい。

また、新たな研究パートナーを見つけるために、セミナーや討論への積極的に参加したい。私が滞在する期間には、ボストン大学のバイオインフォマティクスプログラムにて、本国際交流プログラムが関わるワークショップ、The 9th Annual International Workshop on Bioinformatics and Systems Biology が開催される。さらに、私は、ワークショップにて、研究発表する予定である。発表予定の内容は、上記で述べたクラスタリング手法の遺伝子機能予測への応用である。この発表への討論や参加者との積極的な討論により、現在の研究の問題解決の手がかりを見つけるのみならず、新しい研究パートナーを見つけるために積極的な交流をしてゆきたい。

以上の計画を通して、研究の進展のみならず、国際的な共同研究を行うためのコミュニケーション能力の向上のために努めていきたい。