

分散環境における バイオインフォマティクス処理システムの設計*

菅 敏康[†]平石 広典[†]西山 裕之[†]溝口 文雄[‡]東京理科大学 理工学部 経営工学科[‡]

1 はじめに

ゲノムプロジェクトが大量のゲノム配列を生産するとともに、ゲノム配列をコンピュータを用いて処理するバイオインフォマティクスが注目されている。バイオインフォマティクスの長年の研究において配列の比較、パターン検索など基礎的な方法は考えられておりツールとして数多く存在し公開されている。大量のゲノム配列を処理するには、ツールの統合による自動化が必要である。しかし、各ツールは個別に開発されており、配列解析の自動化のためのツールの統合については考えられていない。また処理には、ゲノム配列データベースなど大量のデータと大量のコンピュータリソースが必要であり、ツールを動かすための環境整備も行わなければならない。大量のゲノム配列の処理を行うためには、ゲノム配列処理のためのバイオインフォマティクスツールの統合による自動化とコンピュータリソースを確保するための分散処理が必要になる。

そこで本稿では、エージェント技術を用いて、分散環境においてバイオインフォマティクスツールの統合と自動化を行うバイオインフォマティクス処理システムの設計について述べる。

2 エージェント技術の適用

エージェントとは、特定の機能と機能に関する知識を持ち、知識に応じて自身の振る舞いの制御や他のエージェントとの協調動作を行うソフトウェアコンポーネントである。

バイオインフォマティクス処理においては、公開されているツールを使うのが一般的である。しかし、ツールが出力するデータには統一性がなく、出力データの処理が必要である。また、バイオインフォマティクスツールの多くは UNIX 環境、または、WWW のサー

ビスとして管理・保守されており、配列の類似検索、モチーフ検索では、数百 MB から数 GB にも及ぶデータベースが必要であり、計算量も膨大である。そのため自動化においては、情報技術に関する様々なスキルが必要になる。

そこで本稿では、それぞれのツールをエージェントとして振舞わせることで、

- エージェントの協調動作による問題解決
- 分散環境におけるバイオインフォマティクス処理
- 各バイオインフォマティクスツールの統合と自動化

を行い、バイオインフォマティクスの処理をサポートする環境を構築することを目的としている。

3 システム設計と実装

3.1 エージェント環境

本システムでは、ネットワーク上に分散する計算機においてエージェントの動作の管理を行う環境として、JADE(Java Agent DEvelopment Framework)[1] を利用している。JADE は FIPA (The Foundation for Intelligent Physical Agents)[2] に基づいたエージェント環境である、

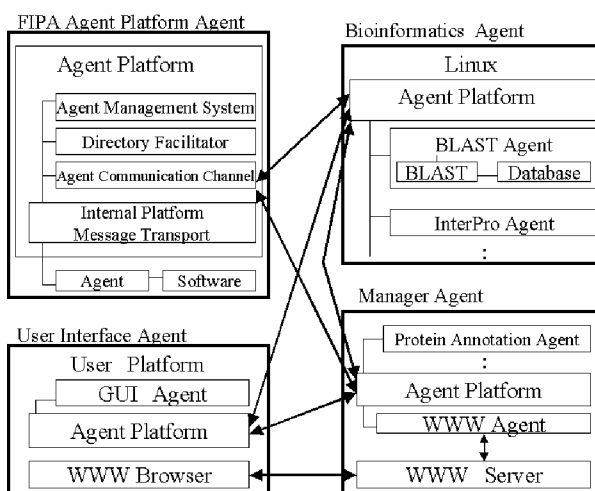


図 1: FIPA エージェント環境と実装したエージェント

*Design of bioinformatics calculation system for distributed Environment

[†]Toshiyasu Suga, Hiraishi Hironori, Hiroyuki Nishiyama, Fumio Mizoguchi

[‡]Dept. Industrial Admin., Faculty of Sci. and Tech., Tokyo University of Science

本システムは、FIPA で定義された通信を用いて協調動作を行っている。FIPA では、エージェント環境にエージェント管理をする Agent Management System, イエローページの役割をする Directry Facilitator, エージェント環境間の通信を行う Agent Communication Channel, メッセージをエージェント環境内のエージェントに伝える Internal Platform Message Transport を配置し、それら FIPA Agent Platform Agent を用いて分散環境でエージェントの協調動作を実現している。エージェントは、FIPA Agent Platform Agent の FIPA エージェント環境に必須なエージェントの他にバイオインフォマティクス用のエージェントとして以下のエージェントがある。

- Bioinformatics Agent
バイオインフォマティクスツールをエージェント化したエージェント
- Manager Agent
ゲノム配列処理に対し必要なデータ管理を行うエージェント
- User Interface Agent
処理の依頼・閲覧するためのユーザインタフェースを提供するエージェント

Bioinformatics Agent は、バイオインフォマティクスツールの動作する環境が必要になるが、Manager Agent や User Interface Agent は、Java で実装しているので、マルチプラットフォームで動作する。本システムでは、主にバイオインフォマティクスツールの多くが動作する Linux とユーザーの環境をシームレスにつなぐように実装している。また、エージェントへの処理依頼や出力結果を閲覧するために、WWW ブラウザの利用も考慮している。

3.2 エージェントの動作

エージェントは、それぞれ固有の機能に基づきバイオインフォマティクス処理を行っている。Bioinformatics Agent は、ツールを利用し、各エージェントからの処理依頼に対し応答する。Manager Agent は、User Interface Agent と特定の Bioinformatics Agent に関する知識と Bioinformatic Agent からの処理結果を処理する機能を持ち、処理結果の情報管理と情報抽出を行う。User Interface Agent は、各エージェントへの処理依頼と処理結果の表示を行う。

バイオインフォマティクスのエージェント間の通信は、FIPA-request, FIPA-query プロトコルを使って他のエージェントに動作要求し、受信したエージェントが、その動作を実行するか、またはその動作を実行できないことを示す応答を返すことになっている。以下に FIPA-query プロトコルを示す。

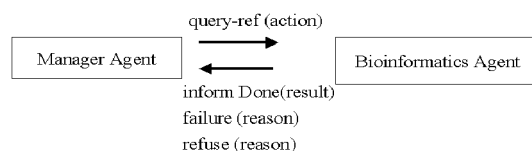


図 2: FIPA-query プロトコル

3.3 エージェントの実装

Bioinformatics Agent の実装として、類似性検索の BLAST, 複数のタンパク質の機能予測ツールとモチーフ検索の InterPro のエージェント化を行い、タンパク質の機能予測の支援を行うシステムを作成した。Fasta 形式の配列を入力データとして、Protein Annotation Agent によって各 Bioinformatic Agent の処理結果から情報抽出され、HTML レポートとして出力される。タンパク質の機能予測として、シグナルペプチド予測、膜貫通領域予測、局在予測を行っている。

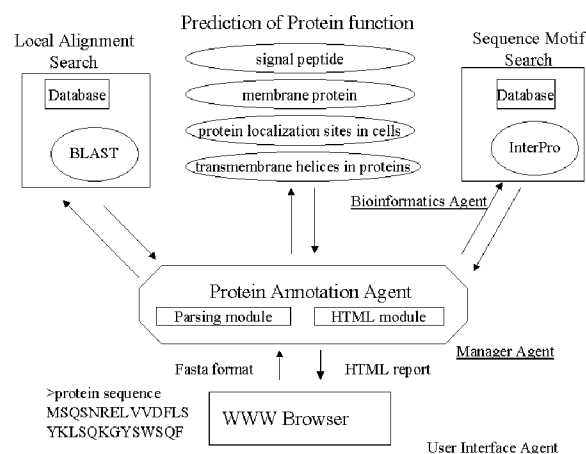


図 3: 機能予測支援システム

4 おわりに

本稿では、エージェント技術の応用としてバイオインフォマティクスの処理にエージェントを用い、分散環境で処理を行うシステムの設計について述べた。現在、分散環境においてバイオインフォマティクスツールの統合と処理の自動化を行っている。今後は、本設計に基づいて実装を続け、エージェントを用いた分散環境でのコンピュータリソースの確保や分散処理を実現させる予定である。

参考文献

- [1] Fabio Bellifemine, Agostino Poggi, Giovanni Rimassa, "JADE A FIPA-compliant agent framework", Proceedings of PAAM'99, 1999.
- [2] Foundation for Intelligent Physical Agents. Specifications ; <http://www.fipa.org>