

# 適合性フィードバックを用いた トリパノソーマ原虫症薬剤標的タンパク質選定の改良

曾我 光瑛<sup>1,2,a)</sup> 石田 貴士<sup>1,2,b)</sup>

**概要:** トリパノソーマ原虫症などの感染症では治療のための薬剤標的タンパク質候補は数百種類程度あると言われており、多数の候補の中からどのタンパク質を選べばよいのかという選択の問題が存在する。これまでのトリパノソーマ原虫症薬剤標的タンパク質選定手法としては、原虫のゲノムとアノテーション情報を集約したデータベースシステムの検索機能を用いることが行われてきたが、それらの検索システムは多くの指標を対象として利用者自らが検索クエリを調整しなければならない非常に煩雑なものであった。本研究ではトリパノソーマ原虫症の薬剤標的タンパク質選定において、ユーザの明示的なフィードバックを利用することで検索結果を改善していく情報検索の技術である適合性フィードバックを用いることで検索の改良を行い、より利用者に適した候補を提示し、選定を補助するシステムを提案する。評価実験により、提案した手法は検索上位のタンパク質の適合率が単純なタンパク質の類似度による検索よりも改善されることを示した。

## Improvement of drug target protein selection for trypanosomiasis based on relevance feedback

SOGA MITSUAKI<sup>1,2,a)</sup> ISHIDA TAKASHI<sup>1,2,b)</sup>

**Abstract:** Hundred kinds of drug target protein candidates exist in infectious diseases such as trypanosomiasis, and thus researchers have to select a protein from them based on their demands. For the problem, database system that integrates pathogen's genome and the annotation information has been provided but the search system of those databases is primitive and it is difficult to obtain desirable search results for the user. In this study, we propose a new method to help drug target protein selection of trypanosomiasis based on relevance feedback. A relevance feedback based system improves retrieval results by using user's explicit feedbacks. As the results of evaluation experiments, user's explicit feedbacks improve retrieval of drug target protein selection.

### 1. 背景

トリパノソーマ原虫症は、ヒト以外の病原体が人体に寄生することで惹起される疾患である感染症の一種である。トリパノソーマ原虫症をはじめとする熱帯地域が感染地の中心である疾患群を世界保健機構 (World Health Organization : WHO) は顧みられない熱帯病 (Neglected Tropical Diseases : NTDs) と定義し解決すべき重要な課

題としている。NTDs が蔓延する地域は主に貧困地域で罹患者の殆どが貧困層のため収益を見込めないことから新薬の開発が殆ど行われてこなかった [1]。

一般に新薬の開発には莫大な時間的、金銭的費用が必要である [2] が、治療対象者の多くが経済的問題を抱えている NTDs 新薬開発ではより効率的に開発を行うことが重要な課題となっており、予め疾病に関連する生体分子を特定して薬剤開発を行う手法 (構造ベース創薬) の活用が効率的なトリパノソーマ原虫症薬剤開発として期待されている。感染症の治療には病原体を人体から除去する必要がある、感染症ではその病原体の生存に重要なタンパク質すべ

<sup>1</sup> 東京工業大学 情報理工学院 情報工学系 知能情報コース

<sup>2</sup> 東京工業大学 情報生命博士教育院

a) [soga@cb.cs.titech.ac.jp](mailto:soga@cb.cs.titech.ac.jp)

b) [ishida@c.titech.ac.jp](mailto:ishida@c.titech.ac.jp)

てが薬剤標的候補となる。トリパノソーマ原虫の一種である *T. brucei* には非常に多くの薬剤標的タンパク質の候補が存在することが大規模な生物学的実験により報告されている [3]。選ばれた標的タンパク質によって薬剤開発の効率が大きく変わるため、トリパノソーマ原虫症薬剤開発において多数存在する薬剤標的タンパク質候補の中からどのタンパク質を実際の標的とするのかという選択が一つの問題となる。

トリパノソーマ原虫症の薬剤標的タンパク質選定を支援するものとして、原虫のゲノム情報とそれぞれのタンパク質に関する生物学的実験などのアノテーション情報を集約している TriTrypDB[4], TDRTargets[5], iNTRODB[6] などの Web システムが存在し、これまでの薬剤標的タンパク質選定はそれらのシステムが提供する検索機能を利用し行われてきた。しかし、そこで利用可能な検索機能はタンパク質に関する多くの指標を利用者自らがそれぞれの主観的な嗜好に合うように指標間の関係や優先度などを考慮しながら調整しなければならない煩雑なものであった。利用者の嗜好の違いから単一の指標や利用者同士の類似性によるタンパク質の評価が困難であるため、利用者に応じた検索機能の開発が求められる。

ここで想定されている利用者は、一般的な Web 検索のように一度きりの検索を行いその結果の最初のみを重要視するような利用者ではなく、検索の改善に協力的であることが想定されるため検索結果を改善するために何度かシステムとやり取りをすることについて大きな抵抗を感じないという仮定をおくことができる。本研究ではトリパノソーマ原虫症における薬剤標的タンパク質選定の改良するための手法を提案する。適合性フィードバックは検索結果に対して利用者の明示的なフィードバックを繰り返し適用し検索クエリを改善することでより良い検索結果を得る手法である。利用者の嗜好に応じた検索の改善という点に注目し適合性フィードバックを薬剤標的タンパク質選定に適用する。

## 2. 適合性フィードバック

### 2.1 適合性フィードバック

情報検索システムにおいて一度の検索のみで利用者にとって理想的な結果を得られることは多くなく、そのような場合に利用者はクエリを自ら改良することによりこの問題を解決しようとする。適合性フィードバックは検索結果に応じてクエリを改良するローカルな手法で、最終的な結果の集合を改善するように検索プロセス内に利用者を含めることでクエリの改良する。基本的な手順は、あるクエリで検索したときの結果に対して利用者が適合か不適合かをマークしてシステムがそのフィードバックに応じてクエリを変更しその改善された結果を表示する、という一連の作業を繰り返すことである。利用者が検索対象のコレクショ

ンをよく知らない場合、最適な検索クエリを自力で構築するのが困難であるが特定のアイテムを判断するのは容易であるという考えに基づいているため検索結果に対して利用者の明示的なフィードバックを得るといふプロセスは理にかなっているとされる [7]。

### 2.1.1 教師あり機械学習手法に基づく適合性フィードバック

情報検索では、検索の利用者が自身の情報要求に従ってクエリを検索システムに投げるようなアドホックな検索だけでなく、あるアイテムがどのクラスに属するのかを決定する分類問題も扱うことができる。分類問題には教師あり機械学習の手法が用いられており、それらの手法を適合性フィードバックに組み合わせた手法が存在する。それらの手法は検索結果の上位数件に対して適合か不適合かのフィードバックを受け、そのデータを用いて更新された学習機を用いて新しい検索を行うというものである。使用する学習アルゴリズムとして様々な手法 [8], [9], [10], [11] が提案されているが、主に Support Vector Machine (SVM) を用いた手法が主流である。

### 2.2 適合性フィードバックの応用分野

適合性フィードバックは主に情報検索の分野で用いられてきたが、応用分野として画像検索がある。Content Based Image Retrieval (CBIR) において画像の低レベルな特徴量と主観的な解釈のギャップを埋めるものとして適合性フィードバックを用いるシステム MARS[12] が提案された。その後、SVM に基づいた適合性フィードバックを適用することでより高い精度で検索を行う手法が開発された [8]。

生物学や医学系分野において適合性フィードバックは主に学術記事の検索 [13] や画像検索 [14] に用いられており、薬剤標的タンパク質選定に用いられていない。

本研究で対象とするトリパノソーマ原虫症薬剤標的タンパク質選定でも、CBIR と同様に検索においてタンパク質に対するユーザの主観的な嗜好をタンパク質の特徴として表現することが困難であることが課題となっておりこれまでの検索はユーザの負担に依存していた。

## 3. 適合性フィードバックを用いたトリパノソーマ原虫症薬剤標的タンパク質選定

本研究では、適合性フィードバックを用いた手法により利用者の明示的なフィードバックを利用することでトリパノソーマ原虫症薬剤標的タンパク質選定を改良する。

### 3.1 利用者の明示的なフィードバックを利用した

#### トリパノソーマ原虫症薬剤標的タンパク質選定支援

図 1 のように、利用者はシステムに対して検索結果が適合か不適合かの明示的なフィードバックを与え、システム

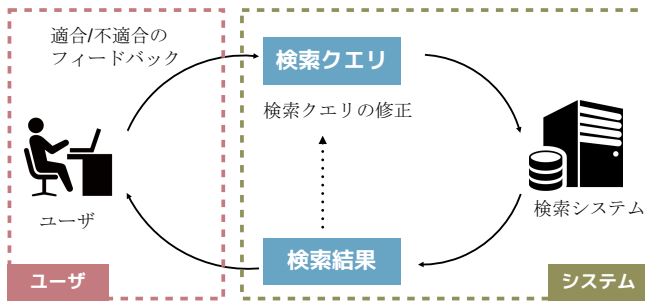


図 1 提案するシステムの概念図

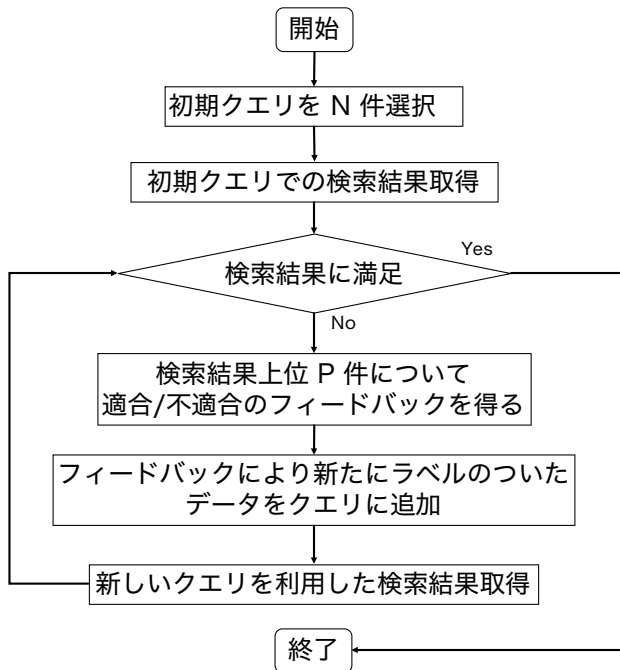


図 2

はそのフィードバックから検索クエリを修正して新たな検索結果を示す。このプロセスを利用者が検索結果に満足するまで繰り返すというのが本研究で提案するシステムである。このシステムは、情報要求に合った検索クエリを最初から利用者自身が構成するのは困難であるが検索結果に対して評価を行うのが比較的容易であるという前提に立っている。

### 3.2 提案手法

図 2 のフローチャートにより本研究で実装したシステムの概要を説明する。最初に検索開始時点でクエリとなるタンパク質を N 個選びそのタンパク質をクエリとして最初の検索を行い、検索のスコアによって並び替えられたタンパク質の一覧 (検索結果) を得る。検索結果に対して満足すればそこで検索を終了し、最終的な結果として現在の検索結果を出力して終了する。そうでなければ、検索結果の上位 P 件に対して適合か不適合かのフィードバックを行い、その P 件を新たにクエリに加えて検索モデルを修正し、そのモデルでの検索結果を出力する。出力結果に満足

がいくまでこの作業を繰り返す。

イテレーションで利用するアルゴリズムとして、以下の分類を行うアルゴリズムを用いた。薬剤標的タンパク質選定での適合性フィードバックを利用した研究例が無いため、どのアルゴリズムが今回のような検索に適するか検証を行うために複数のアルゴリズムでの比較を行う。

#### 3.2.1 初期検索

一番最初にどのアルゴリズムでも共通で検索を行う初期検索にはベクトル間のユークリッド距離を用いた。適合性フィードバックでは不適合ラベルよりも適合ラベルの方が重要とされ、最初のイテレーションで検索結果の上位のタンパク質に対してフィードバックを行う際に有用な情報を得る上でも初期検索は重要である。本研究では利用者が求めるタンパク質ベクトルは互いに類似しているものと仮定したため、初期検索でユークリッド距離を用いた。

#### 3.2.2 適合性フィードバックのアルゴリズム

Rocchio のアルゴリズムは 1970 年代に提案された適合性フィードバックの実装のための古典的なアルゴリズムである。ある情報要求に適合する文書同士はベクトルも類似したものであるという根拠のもと、適合と不適合の分類を各クラスを表すクエリ集合の重心からの距離により行っている。以下の式にしたがって各イテレーションでクエリを更新する。ただし  $q_0$  を初期クエリ、 $D_r$ 、 $D_{nr}$  をそれぞれ適合、不適合文書の集合とする。

$$q_{opt} = \alpha q_0 + \beta \frac{1}{|D_r|} \sum_{d_j \in D_r} d_j - \gamma \frac{1}{|D_{nr}|} \sum_{d_k \in D_{nr}} d_k \quad (1)$$

最終的なスコアは更新されたクエリベクトルと各タンパク質ベクトルとのユークリッド距離で、昇順を検索結果として出力する。このアルゴリズムはベクトル空間モデルにおいて、ベクトルの類似度に基づいて検索を行う最も単純な実装である。

#### 3.2.3 教師あり機械学習アルゴリズム

##### 3.2.3.1 SVM ベースのアルゴリズム

Rocchio のアルゴリズムによる分類は、各クラスの重心からの距離に基づく最も単純なアルゴリズムである。SVM による分類は Rocchio の分類よりもテキスト分類において効果的であるとされる。

本研究では、SVM の実装の一つである Python ライブラリ scikit learn での実装を用いた。ハイパーパラメータとしてカーネル関数、誤認識率とマージンの比率  $C$ 、RBF カーネルの式 (2) に出てくる変数  $\gamma$  について、訓練データセットに対して 5-fold cross validation を行いカーネル関数とパラメータを定めた。カーネル関数は線形カーネルと RBF カーネルから選択し、スコア関数は Average Precision を利用した。

最終的なスコアは決定関数の値として、降順を検索結果として出力する。

$$K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2) \quad (2)$$

### 3.2.3.2 Random Forest ベースのアルゴリズム

Rocchio のアルゴリズムや SVM と異なり、Random Forest による分類は複数の木構造の分類器により分類を行う。

本研究では、Random Forest の実装の一つである Python ライブラリ scikit learn での実装を用いた。ハイパーパラメータとして弱学習器 (木) の数  $n\_estimators$  とランダムにサンプリングされる特徴量の最大数  $max\_features$  を訓練データセットに対して 5-fold cross validation を行い値を定めた。最終的なスコアは森の中で適合と判定した木の割合 (Class Probability) として、降順を検索結果として出力する。

### 3.2.3.3 ロジスティック回帰ベースのアルゴリズム

テキスト分類には確率的に分類する手法があり、ナイーブベイズなどの手法が用いられる。今回は適合か不適合かの 2 値分類であることから、ロジスティック回帰を用いる。ハイパーパラメータとして正則化項の値  $C$  を訓練データセットに対して 5-fold cross validation 行い決定した。検索結果は Class Probability の降順で並べたものを結果として出力する。

## 3.3 評価指標

検索結果の集合について、適合率 (precision) (P) と再現率 (recall) (R) は以下で定義される。

$$\text{適合率} = \frac{\#(\text{取得された正解要素数})}{\#(\text{取得された要素数})} \quad (3)$$

$$\text{再現率} = \frac{\#(\text{取得された正解要素数})}{\#(\text{正解要素数})} \quad (4)$$

適合率と再現率はどちらも集合に対して定義される値である。本研究では、検索結果はそれぞれのアルゴリズムで定義したスコアに従ってすべてのタンパク質を並び替えたランク付きの検索結果であるため、以下の 3 つの指標で検索システムの効果を測定する。

### 3.3.1 上位 k 件における適合率

利用者にとって重要であるのはその中でも上位のものである。上位 k 件における適合率は、検索結果上位 k 件までの集合における適合率であり、利用者にとって重要な部分を直感的に評価することができる。様々な評価指標の中で安定しないとされているが、本研究の目的は多数ある中から効率的に選定を行うことであり実際の利用場面で利用者が最も重視するのは上位にどれだけ適合するものを示せるかであることから、こちらの指標を重視する。

### 3.3.2 11 点補間平均適合率

ある再現率レベル  $r$  における補間適合率 (interpolated precision)  $p(r)$  は  $r' \geq r$  であるすべての再現率の中で一番

高い適合率、つまり以下の式により定義される。

$$P_{interp}(r) = \max_{r' \geq r} p(r') \quad (5)$$

この補間適合率を  $r = \{0.0, 0.1, \dots, 1.0\}$  の 11 点について測定し、適合率-再現率曲線で示す。この評価指標は TREC のアドホック検索評価の最初の 8 回で利用された [15]。

### 3.3.3 MAP

MAP (Mean Average Precision) は現在 TREC のコミュニティで標準的に利用されている指標で、再現率にまたがった検索システムの標準的な性能を示す。ある情報要求  $q_j \in Q$  の関連文書が  $\{d_1, \dots, d_{m_j}\}$  で、文書  $d_k$  を得るまでの最上位からのランク付き検索結果の集合を  $R_{jk}$  として以下の式で表される。

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} Precision(R_{jk}) \quad (6)$$

テストコレクションの違いのため、一般に MAP の絶対値はそこまで重要ではないが、今回の実験では単一の情報要求に対する検索でそれぞれのアルゴリズムに対してすべて同じコレクションを利用しているため MAP の値をアルゴリズムの性能の比較に利用する。

## 4. 実験

### 4.1 実験の設定

一回の実験は、検索の初期クエリとして 1 つのタンパク質を選択し、検索結果に対してフィードバックを 10 回行う。各フィードバックでは検索結果上位 10 件までに順位付けされた、既にフィードバックに利用されていないタンパク質を適合タンパク質と不適合タンパク質集合にそれぞれ追加して各アルゴリズムの検索モデル (分類器) の更新を行った。実験は 30 回行い、実験ごとに初期クエリをランダムに選定した。以下に示す結果の数値は 30 回の実験の平均値である。これらはすべて自動で行った。

### 4.2 使用したデータセット

iNTRODB で公開されている *Trypanosoma brucei* の全 9078 タンパク質を対象とし、公開データを元にそれぞれのタンパク質を 14 次元の特徴ベクトルとして表現した。

#### 4.2.1 *Trypanosoma brucei* タンパク質情報

Yoshino ら (2017) [16] の標的候補選定の手法で用いられた、主に以下の規準で選定された 195 タンパク質を適合、それ以外を不適合とする。これらのタンパク質は iNTRODB の検索機能と専門家の手による選別により選定された。

- RNAi 実験で生存に重要であることが判明している
  - トリパノソーマ科寄生原虫間で遺伝子が保存している
  - 類縁タンパク質の結晶構造解析の報告がある
- また、欠損値には平均値を代わりに用いて、z-score で

手法	MAP
Rocchio	0.491
SVM	0.415
ロジスティック回帰	0.553
Random Forest	0.489

表 1 各アルゴリズムを実装したシステムの MAP

正規化を行う.

$$Z\text{-score} = \frac{X_i - \mu}{\sigma} \quad (7)$$

ただし  $\mu$  は母集団平均,  $\sigma$  は標準偏差とする. iNTRODB で公開されている *Trypanosoma brucei* の全 9078 タンパク質を用いる. タンパク質はそれぞれ, 実験結果などを示す 14 次元のベクトルとして表現される.

### 4.3 実験結果

各アルゴリズムにおける上位  $k$  件 ( $k = 20, 195$ ) のイテレーションごとの変化を図 3,4 に示す. 最も単純な手法である Rocchio のアルゴリズムではフィードバックを繰り返しても初期のクエリ (タンパク質 1 つ) による類似度検索と結果がほぼ変化しないということから, 本研究の問題設定は単純なベクトルの類似度だけで検索を行うことが困難であるということがいえる.

その他の教師あり機械学習アルゴリズムでは, 初回に検索結果の適合率が悪くなるものの, それ以降は適合率を上げている. 本研究では実際に利用した際にユーザが重視するであろう検索結果上位の適合率により注目している. Random Forest ベースのアルゴリズムが上位 20 件における適合率で最も良い結果を残している. ユーザからのフィードバック回数による上位  $k$  件の適合率の変化から, Random Forest ベースのアルゴリズムと SVM ベースのアルゴリズムではユーザから 4 回程度のフィードバックを得られれば十分だと言え. 一方ロジスティック回帰ベースのアルゴリズムでは他の手法と同様の 4 回程度のフィードバックでは性能が不十分なものの回数を繰り返しても検索結果が改善され, 上位 195 件では最も高い適合率であった.

図 5 と表 1 では, 10 回フィードバックした後の検索結果についてそれぞれ 11 点補間適合率と MAP により再現率にまたがった性能を示している. イテレーション 10 回終了後の検索結果での 11 点補間適合率と MAP の値を示す. 再現率 0.1 (適合タンパク質上位 20 件程度) までの適合率を見ると Random Forest が最もよい結果となっているが, それ以降の再現率レベルではロジスティック回帰ベースのアルゴリズムが優れており, 再現率レベルが 1.0 の適合率は Rocchio のアルゴリズムが最も優れている. すべての再現率レベルにおける適合率を考慮した MAP ではその傾向が見て取れ, ロジスティック回帰ベースのアルゴリズムを実装したシステムが最もよい結果を残している.

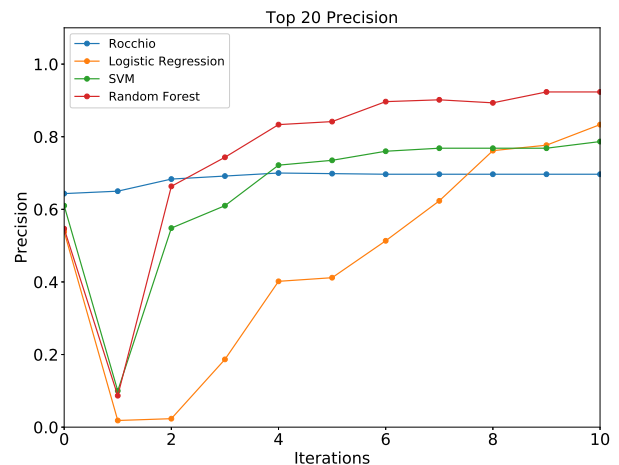


図 3 各種法の上位 20 件における適合率の比較

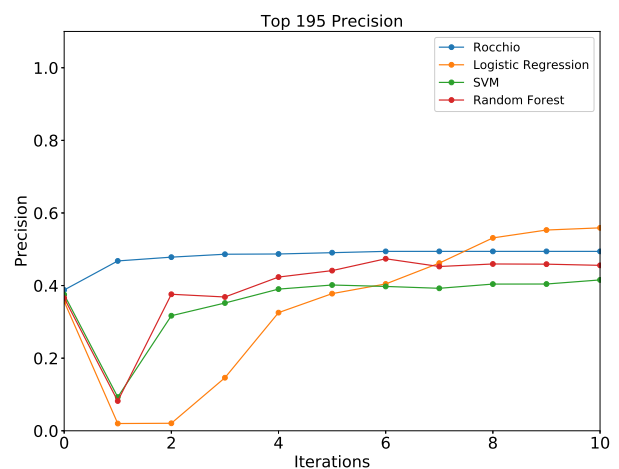


図 4 各種法の上位 195 件における適合率の比較

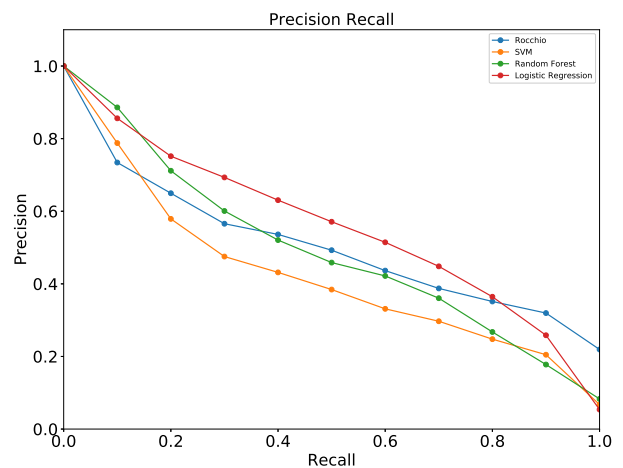


図 5 11 点補間適合率での適合率-再現率グラフ

### 4.4 考察

利用者が最も重視すると想定される上位にランク付けされたタンパク質の適合率では, フィードバックによる改善の速さからも Random Forest が最も適していると言える. Yoshino ら (2017) の研究 [16] では, 最終的には 4 つま

で絞り込んでいる。二段階目以降の絞り込みでは、タンパク質の特徴以外の情報も利用しているため、タンパク質の特徴のみから最終的な 4 件を得ることは困難だと考えられるがその候補の数を効率的に絞りこめることが望ましい。Random Forest ベースのアルゴリズムでは、少ないフィードバックの回数で上位 20 件までで最も効率的に絞り込むことが可能である。

システム全体の評価として 11 点補間平均適合率の適合率-再現率グラフと MAP で比較した場合、ロジスティック回帰が一番結果の良かった手法だと言える。図 5 では上位の適合率を見ると Random Forest が一番適合率の高い結果となっているが、それ以降は適合するタンパク質を多く検索することができていることがわかる。しかし、上位 20 件程度しか気にしないのであれば 11 点補間平均適合率や MAP の結果にもかわらず Random Forest が最も優れた手法であると言える。

## 5. 結論

### 5.1 本研究の結論

本研究では、トリパノソーマ原虫症薬剤標的タンパク質選定についての検索の特徴や既存システムの問題点から、利用者の検索結果に対する明示的なフィードバックを利用する手法である適合性フィードバックの考え方を利用することで検索結果を改善できると期待し、それを実現する手法として適合性フィードバックを応用したトリパノソーマ原虫症薬剤標的タンパク質選定システムを提案した。検索結果に対して明示的なフィードバックを利用することで、利用しない場合と比較して適合率を改善させることを示した。また、検索に用いるアルゴリズムとして Random Forest が最も適切な手法であることを示した。

### 5.2 今後の課題

今後の課題として以下の点が挙げられる

#### (1) 実際の検索システムへの実装

実際のシステムにおいては、ユーザが許容してくれるフィードバック回数や判断の正確性など今回の実験では評価の対象としていない部分も考慮し、用いる手法を検討する必要がある。

#### (2) SVM の手法の改善

画像検索分野では、SVM に基づく適合性フィードバックで弱学習木の組み合わせ [11] や、特徴量の重み付け手法などで検索を改善する研究が存在する。本研究の分野においても、SVM ベースのアルゴリズムにそれらのような手法を適用することで改善を期待できる。

## 参考文献

[1] Nagle, A. S. et al.: Recent Developments in Drug Discovery for Leishmaniasis and Human African Trypanoso-

- miasis, *Chemical Reviews*, Vol. 114, No. 22, pp. 11305–11347 (2014).
- [2] Mullard and Asher: New drugs cost US\$2.6 billion to develop, *Nature Reviews Drug Discovery*, Vol. 13, p. 877 (2014).
- [3] Alsford, S. et al.: High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome., *Genome research*, Vol. 21, No. 6, pp. 915–924 (2011).
- [4] Aslett, M. et al.: TriTrypDB: a functional genomic resource for the Trypanosomatidae, *Nucleic Acids Research*, Vol. 38, No. suppl.1, pp. D457–D462 (online), DOI: 10.1093/nar/gkp851 (2010).
- [5] Agüero, F. et al.: Genomic-scale prioritization of drug targets: the TDR Targets database, *Nature Reviews Drug Discovery*, Vol. 7, pp. 900 EP – (2008).
- [6] : iNTRODB Integrated Neglected TROPical disease DataBase, (online), available from (<http://www.bi.cs.titech.ac.jp/introdb/index.html>) (accessed 2018-01-20).
- [7] Manning, C. D. et al.: *Introduction to Information Retrieval*, Cambridge University Press, New York, NY, USA (2008).
- [8] Tong, S. and Chang, E.: Support Vector Machine Active Learning for Image Retrieval, *Proceedings of the Ninth ACM International Conference on Multimedia*, pp. 107–118 (online), DOI: 10.1145/500141.500159 (2001).
- [9] Cai, F. et al.: Your Relevance Feedback Is Essential: Enhancing the Learning to Rank Using the Virtual Feature Based Logistic Regression, *PLOS ONE*, Vol. 7, No. 12, pp. 1–9 (2012).
- [10] Yu, H. et al.: Enabling multi-level relevance feedback on PubMed by integrating rank learning into DBMS, *BMC Bioinformatics*, Vol. 11, No. Suppl 2, pp. S6–S6 (2010).
- [11] Tao, D., Tang, X., Li, X. and Wu, X.: Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 7, pp. 1088–1099 (2006).
- [12] Rui, Y. et al.: Content-based image retrieval with relevance feedback in MARS, *Proceedings of International Conference on Image Processing*, Vol. 2, pp. 815–818 vol.2 (online), DOI: 10.1109/ICIP.1997.638621 (1997).
- [13] Abdulla, A. A. A. et al.: Improving biomedical information retrieval by linear combinations of different query expansion techniques, *BMC Bioinformatics*, Vol. 17, No. Suppl 7, p. 238 (2016).
- [14] Rahman, M. M. et al.: A Learning-Based Similarity Fusion and Filtering Approach for Biomedical Image Retrieval Using SVM Classification and Relevance Feedback, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 15, No. 4, pp. 640–646 (2011).
- [15] Teufel, S.: *An Overview of Evaluation Methods in TREC Ad Hoc Information Retrieval and TREC Question Answering*, pp. 163–186, Springer Netherlands (2007).
- [16] Yoshino, R. et al.: In silico, in vitro, X-ray crystallography, and integrated strategies for discovering spermidine synthase inhibitors for Chagas disease, *Scientific Reports*, Vol. 7, No. 1, p. 6666 (2017).