

ベイジアンネットによる反復投与毒性評価システム ToxBay の構築

ToxBay: A System for Repeated Dose Toxicity Evaluation by Bayesian Net

岡田 孝^{*1}
Takashi Okada

大森 紀人^{*1}
Norihito Ohmori

^{*1} 関西学院大学理工学部
School of Science & Technology, Kwansai Gakuin University

ToxBay is a system to evaluate repeated dose toxicity of rats based on Bayesian net. The net consists of 3 layers: chemical categories, diagnoses and endpoints. We used cascade model to extract basic active structures causing each diagnosis. Links between diagnoses and endpoints are set following suggestions by toxicologists. Some categories showed unique tendencies to cause endpoints, and we added direct connections between these nodes. CPT values among diagnoses and endpoints were set by EM learning. Users input chemical structures by SMILES and see the resulting posterior probabilities by node colors. When the evidence contains endpoints, the system shows local conflicts among nodes by coloring edges. If its conflict measure exceeds a threshold value, the system suggests most questionable observations. ToxBay is disclosed on Internet for the use by toxicologists.

1. はじめに

大量に生産される化学物質の安全管理は重要な課題であるが、10万種にも昇る物質すべての毒性を調べることは実質的に不可能であり、*in silico* の評価法が期待されている。特に数週間に渉る投与により評価される反復投与毒性は、化学構造との因果関係の認識が困難なことで知られている。2007年度から5年間の NEDO/METI プロジェクト「構造活性相関手法による有害性評価手法開発」では、国内外からラットの反復投与毒性データを収集して DB 化し、さらに評価手法の開発も行った[1]。

我々はこのデータにもとづいて、反復投与毒性評価システム ToxBay を構築した。化学構造マイニングにより構造カテゴリーを抽出し、毒性学者の知識とデータ分布の特異性にもとづいてネットのトポロジーを構築した。ToxBay は矛盾の検出にもとづき、データ評価のための豊富な情報を与えることができる。ネットの構築法と利用者インターフェースについて報告する。

2. データベースとネットの基本構成

データベースは各種の試験報告書から抽出したおよそ 500 化合物の詳細な定量的データからなるが、本研究で使用したものは毒性兆候表として要約されたもので、20 種の endpoint (赤血球減少や GPT 増加など) に対して各化合物の影響の有無が記されている。それぞれの endpoint は、これまでの毒性学研究により、溶血性貧血や肝細胞毒性など 7 種の診断項目とおおよそ関係づけられている。

この知見に従い、ネットの構造は図1のように最上層に化合物カテゴリー(アニリン、フェノールなど)を置き、その下に診断さらに endpoint を置いた 3 層構成とした。またすべてのノードは T/F の 2 値とした。診断と endpoint 間のリンク設定は毒性学者の示唆に従った。化合物カテゴリーを定めるためには、観測項目でない診断と化学構造の関係をマイニングする必要がある。そこで、溶血性貧血では RBC, HGB の双方が減少した場合、肝細胞毒性では GOT, GPT のいずれかが増加した場合を毒性とし、他の診断項目では、もっとも関連の深いと示唆された endpoint をその診断での毒性指標とした。

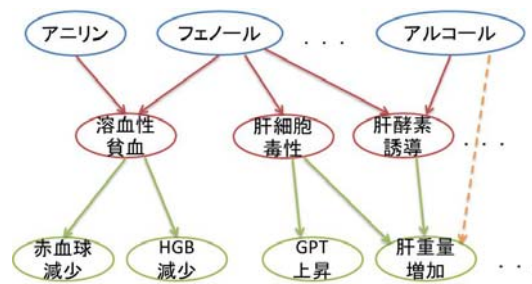


図1. ネットの基本構成

3. 活性構造の抽出と化合物カテゴリーの設定

7 種の診断に対して毒性を引き起こす特徴的な基本活性構造 (BAS) を、カスケードモデルを用いたデータマイニングの技法[2]により抽出した。これらの構造群は、対象となる部分構造と除外すべき mask 構造群を SMARTS 表記で指定し、さらに必要に応じて分子量と疎水性を表す logP 値で制限をかけたものである。発掘した結果は BASiC サイトで、各 endpoint を直接対象とした BAS 群と共に公開している[3]。なお、化合物数が少ないため、2化合物にのみ存するような統計的に有意でない構造も、化学的な反応性を考慮して積極的に採用している。

これらの BAS 群で化学的に類似したものを統合し、化合物カテゴリーノードを設定した。なお、NITE では本研究とは独立に、毒性学文献を元に作用機序の明らかなものをまとめて、カテゴリーを設定している。このカテゴリーの物質がたとえ DB 中に存在しない場合でも、これらを ToxBay のカテゴリーの一部として組み込んだ。化合物カテゴリーの種類は 50 種程度である。

4. ネットワークの構築

4.1 トポロジーの精細化

前節の活性構造は診断項目に対応するため、その間のリンクを設定できる。また、各診断と endpoint 間のリンクは毒性専門家の意見に従い設定した。ただし因果関係が明確なメトヘモグロビン生成による溶血性貧血に対応するため、カテゴリーと溶血性貧血の間にメトヘモグロビン生成のノードを作成した。また、脾臓色素沈着も赤血球減少の下流に置いた。さらに、観測され

ていない endpoint でも、評価に有用と考えられる endpoint はネットの構成に含めた。

このようなネットでは、特定の診断にリンクされている複数の endpoint 群は、CPT 値の差によりその程度は異なってもすべて同様の毒性発現パターンを有する。しかし、BAS 毎に各 endpoint への影響を調べると特異的な相関が見られる。例えば次の表1に示す肝重量増は全学習化合物中の過半で発生する endpoint であり、肝細胞毒性の BAS 群に採用されている化合物に限ってもほぼ 2/3 で観測されている。しかし、表の下3行に示した BAS では毒性比率が小さい。この場合は、アルコールもカルボン酸に酸化されることを考慮すると、「酸は肝重量増を起こしにくい」と結論できる。

表1 肝細胞毒性と肝重量増 endpoint 間の特異的相関

肝細胞毒性 BAS	肝重量増化合物数(毒/非毒)
全学習化合物	190 / 125
BAS 採用群	88 / 41
カルボン酸	1 / 4
スルホン酸	0 / 4
エーテル&アルコール	0 / 4

このような特異性は、肝細胞毒性における肝細胞壊死、酵素誘導における肝重量増、および貧血における脾臓色素沈着でも発見された。これらの特異的相関がある場合は、化合物カテゴリーから当該の診断へのリンク以外に、カテゴリーから endpoint への直接のリンクを追加した。

4.2 CPT 値の設定

化合物カテゴリーから診断への条件付確率表は、各カテゴリーにおける毒性発生頻度に従って確率を定め、NosiyOr 関数により表現した。ただし、False negative の予測を少なくするために、頻度分布よりも 0.1 程度高い値を設定している。また、漏れの確率は 0.1 に設定した。各診断から endpoint への確率表は EM learning 法により値を定めた。

なお、NITE によるカテゴリーで、上記設定法が適用できない場合は、カテゴリーによる診断の生起確率を 0.7、診断による endpoint の生起確率を 0.5 とした。

5. ToxBay システム

5.1 実装

作成したネットを毒性専門家に提供するために、IE 上で動作する web application ToxBay を作成し公開した[4]。このシステムは、Javascript で端末側の処理を行い、python でサーバ側の処理を実装している。なお、サーバ側では(1)OpenBABEL[5]を利用して物性値推算とカテゴリー設定、(2)Hugin[6]を用いた確率推論を行っている。なお、Hugin とのインターフェース作成のために一部 C 言語による実装も行った。

5.2 利用者インターフェースと矛盾検出

ToxBay にアクセスすると、図2のような画面を表示する。右上ペインにはネット構造が無色で表示される。ここで I の化合物を評価してみよう。① 利用者は分子の構造を SMILES 表記により入力し、② logP 値と分子量を推算すると、③ 該当する化合物カテゴリーが左下ペインに evidence として設定される。④ goBN ボタンをクリックして推論を行うと、事後確率の値に従って各ノードが色付けて表示される。この例では、ニトロベンゼンとハロゲ

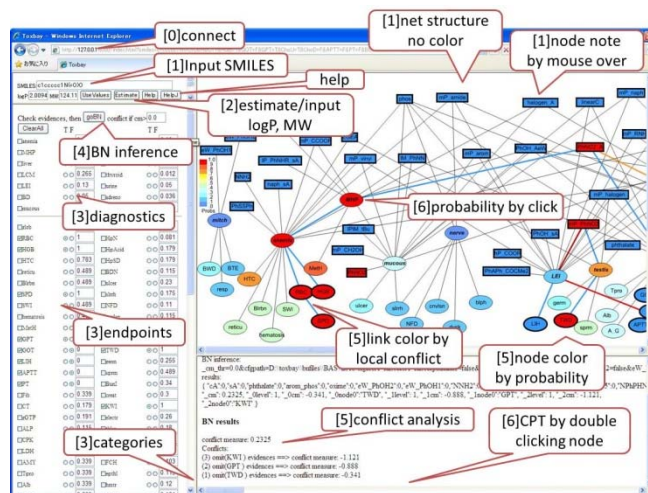
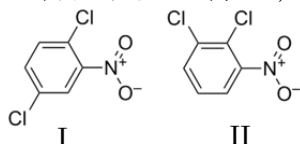


図2. ToxBay の画面と利用者インターフェース

ンに係わるカテゴリーが発火し、貧血がもっとも起こりやすく、ついで肝臓関係および精巣に係わる症状がでやすいことがわかる。

データベース中で I の化合物に類似した構造を有するものを探すと II が見つかる。この化合物では③で endpoint の観察結果も併せて evidence として指定してみる。その結果、貧血と肝臓の酵素誘導が起こっていることが分かる。

結果表示で注目すべきは、貧血(赤)→赤血球減少(青)→脾臓色素沈着(赤)のようにノードが色づけられており、さらにこの間のリンクが赤で着色されていることである。この赤色は、当該リンクが矛盾を起こしていることを示している。対照的に貧血から HGB 減少は、双方で T の値が高くもってもらいリンクとして青色で表示されている。なお、もってもらい因果関係は、リンク両端が F である場合にも成立するが、これは「原因がないから結果もない」ことを意味しており、利用者にとっては煩雑で無用の情報であるため、黒色でリンクを表示している。

この赤色の矛盾表示は、赤血球減少が観測されていないのに、上流の貧血と下流の色素沈着が起こっていることに注意を促している。このように ToxBay は実験結果の評価にも用いることができる。なおこの例では、毒性兆候表で赤血球減少が F とされていたが、定量的データでは減少していると判断してもよい事例であると判明している。

右下側ペインは、ネット全体としての矛盾情報とデバッグ用の計算過程の詳細を表示するために用いている。ネット全体の conflict measure 値が閾値以上の場合に、その値を最も大きく低下させる 3 種までの evidence を表示し、利用者により矛盾解消のための手懸かりを与えている。

6. まとめ

ToxBay はインターネット上で公開されている。なお、API を通して毒性兆候表のデータを evidence として与えて評価したり、化合物構造とマッチする BAS を検索する機能も有する。今後多くの毒性関係者に利用されるものと考えている。

参考文献

- [1] 林他: <http://www.nedo.go.jp/content/100095236.pdf>.
- [2] Okada et al: *Chem. Central J.* vol.4, 1, 2010.
- [3] 岡田他; BASiC, <http://www.dm-lab.ws/BASiC/>.
- [4] 岡田他: ToxBay, <http://211.8.17.77/>.
- [5] O'Boyle et al: Open Babel: An Open chemical toolbox, *Journal of Cheminformatics* 2011, 3:33.
- [6] HUGIN EXPERT: <http://www.hugin.com/>.