

添付文書情報を用いた医薬品の薬効および副作用の分類 Classification of pharmacological and side effects of drugs based on the package insert

京都大学化学研究所 小寺正明 (Masaaki Kotera)

背景

医薬品は主に人の疾病の診断・治療・予防に用いられるが、薬理作用と同時に副作用を持つため、適正な使用が求められる。本研究では、薬理作用と副作用との因果関係を理解することを目的に、添付文書情報の文書解析を行った。

方法

医薬品データは、ゲノムネット医薬品データベース [1] に登録されている JAPIC 医薬品添付文書情報 [2] および KEGG DRUG データベース [3] から得た。

添付文書は日本語で記述されているため、MeCab プログラム [4] を用いて形態素解析を行い、ライフサイエンス辞書 LSD [5] を用いて英単語・英熟語（以下、「語」と呼ぶことにする）に変換することにより、JAPIC のエントリ毎にそれぞれの語の存在 (1)・非存在 (0) のベクトルとして表した。また、JAPIC 添付文書は XML 形式で記述されているが、文書によって記述方法が異なることが多く統一性に欠いていたため、複数の XML タグを「薬理作用」と「注意事項」に大分類し、それに基づいて語の属性を決めた (表1)。

JAPIC では、添付文書ごとにひとつのエントリとしているため、同じ化合物であっても製造会社や承認番号が異なれば異なるエントリとなる。そこで、KEGG DRUG データベースのエントリ毎に JAPIC エントリをまとめることで、医薬品毎に語の存在・非存在ベクトルを作った。

続いて、語の重要度を評価するためにフィッシャーの正確確率検定を行い、重要であると見なされた語のみからなるベクトルを作成し、医薬品をクラスタリングした。クラスタリングには東京大学新領域創成科学研究科の中谷明弘准教授が作成されたアルゴリズムを使用した。

結果

薬理作用および注意事項の属性が割り当てられた語数は、それぞれ17,113 語、16,853 語であり、重複が16,851語あった。これらの語がそれぞれいくつの薬物で使われているかを示したところ、典型的なべき乗則に従い、半分以上の薬で用いられている語はかなりまれであることが分かった (図1)。頻繁に使われている語は、薬理作用では usage (使用)、unit (単位)、specification (規格、特定)、indication (効能、適用)、effect (効果) などが見られ、注意事項では attention (注意)、use (使用)、and (および)、side effect (副作用)、symptom (症状) などが見られた。これら頻繁に使われる語は、医薬品を特徴づける目的には適さないと考えられる。

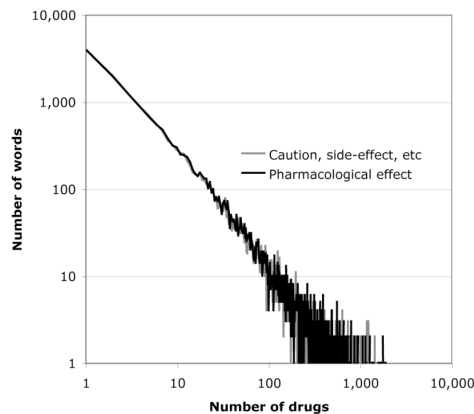


図1. 単語・熟語の出現頻度分布

次に、同じ医薬品に使われている語は意味的につながりがあるであろうことが予想されたので、任意に選んだ2つの語が同じ添付文書で現れる (共起する) かどうかを語の類似性とし、その分布を調べた (図2)。ここで、類似性の定義は、それぞれの語に対し、全ての医薬品の添付文書に使われているか(1)否か(0)をベクトルで表現し、類似性指標としてタニモト係数を用いた (2つの語がともに使われている医薬品

の数を、少なくともどちらか一方が使われている医薬品の数で割った)。その結果、薬理作用と注意事項で分布はほとんど変わらず、また、それぞれの語はほぼ独立に出現していることが分かった。このため、この指標を用いて医薬品を特徴づける語を探すのは困難であることが示唆された。

表1. 語の属性分けに用いた XML タグの例

薬理作用 (Pharmacological effect)	
Effect	効能効果
effect. and. usage	効能効果及び用法用量
efficacy. pharmacology	薬効薬理
Usage	用法用量
注意事項 (Caution, side-effect, etc.)	
Adverseevent	副作用
caution. for. application	適用上の注意
caution. for. handling	取扱上の注意
caution. to. effect	効能効果に関連する使用上の注意
caution. to. usage	用法用量に関連する使用上の注意
caution. touch	接触注意
Important. caution	重要な基本的注意
over. dose	過量投与
Precaution. for. use	使用上の注意
serious. ae	重大な副作用
Warning	警告

続いて、任意に選んだ2つの医薬品の添付文書がどのくらい似ているかを調べた(図3)。ここでも類似性指標としてタニモト係数を用いた(2つの添付文書でともに現れる語の数を、少なくともどちらか一方で現れる語の数で割った)。その結果、薬理作用と注意事項でやはり分布に差は見られず、添付文書同士の類似性は0.2をピ

ークとしていた。以上の結果から、頻繁に使われる語や出現が類似している語の割合は非常に少なく、医薬品の特徴付けに重要でない語が多く含まれていることが示唆された。

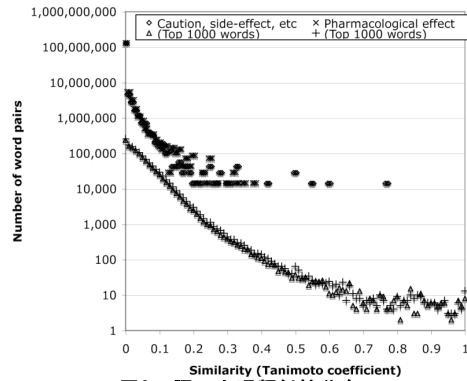


図2. 語の出現類似性分布

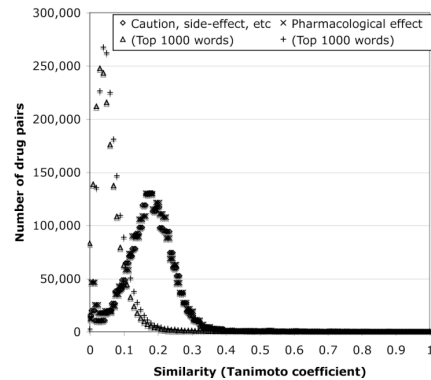


図3. 医薬品の添付文書類似性分布

表2. フィッシャー正確確率検定で上位にランキングされた共起語

20%以上の医薬品で共起する語			1%~20%の医薬品で共起する語		
biology (生物学)	biological (生物学的な)	5.23e-66	Eosinophil (好酸球)	eosinophilic (好酸性)	4.78e-56
half-life (半減期)	halve (半減)	3.86e-62	Teratogenic (催奇形)	teratogenicity (催奇形性)	3.27e-54
Biology (生物学)	organism (生物)	9.27e-62	Eosinophilia (好酸球増加症)	eosinophil (好酸球)	5.78e-51
Stability (安定性)	market (市場)	5.18e-60	Eosinophilia (好酸球増加症)	eosinophilic (好酸性)	2.11e-50
Acceleration (加速、促進)	speculation (推測)	2.71e-59	Matching (適合、整合)	elution (溶出)	1.18e-46
Blinded (盲検、盲目)	blindness (盲目症)	7.79e-59	Electrolysis (電解)	electrolyte (電解質)	7.41e-42
Car (自動車)	drive (運転)	1.06e-58	Agranular (無顆粒)	agranulocytosis (無顆粒球症)	1.85e-41
Receptor (受容体)	accept (受容)	1.07e-57	Conjugate (抱合体)	conjugation (抱合)	2.64e-40
Relative (親戚、相对)	acceleration (加速、促進)	4.28e-57	Ray (光線)	photosensitivity (感光性、光過敏症)	2.71e-40
Doublet (二重)	Blinded (盲検、盲目)	7.95e-57	Cytopenia (血球減少症)	pancytopenia (汎血球減少症)	5.38e-40

そこで、医薬品の薬理作用や副作用を特徴づけるために重要な語とそうでない語を区別するため、フィッシャーの正確確率検定を行うことにした。この検定では、正例・負例を用意し、それぞれについてある事象が成立する数・しない数を数え上げたとき（これを「分割表」と呼ぶ）、「正例か負例か」「その事象が成立するか否か」という2つの因子が独立であるという仮定の下に、ある分割表が得られる確率を調べ、そこからP値を算出する。

本研究ではまず、薬理作用ベクトルと注意事項ベクトルとの組み合わせで、正しい組み合わせ（正例）に対し、間違っただけ作成した。次に、薬理作用のある語と注意事項のある語が、正例と負例とでそれぞれどれだけ共起するかを数え上げ、フィッシャーの正確確率検定を行った（表2）。表2の左側と右側はそれぞれ、20%以上の医薬品、1%~20%の医薬品で有意に共起する

語をP値の小さい順に並べたものである。医薬品を特徴づけるためには後者の方が適せるであろうことが分かる。

次に、1%~20%の医薬品で共起する語のトップ1000を用いて医薬品の薬理作用ベクトルと注意事項ベクトルを作成し、それぞれ別個にクラスタリングした結果が図4（それぞれ左上と右下）である。これらの図では、縦軸と横軸はそれぞれのクラスタリング結果に基づいて医薬品エンタリが並べられており、交差する点の灰色の濃さが対応する医薬品の類似度を反映するようになっている。薬理作用ベクトル、注意事項ベクトルでそれぞれ特徴的なクラスターが得られたが、これは最初のベクトル（全ての語を用いたもの）では得られなかったものである。

図4の左下は、横軸を薬理作用、縦軸を注意事項で並べたときの類似度分布である（それぞれ左上と右下の図に対応している）。薬理作用、注意事項どちらでも同一のクラスターに分けられた部分を黒色で示しており、含まれる医薬品の多い順に #1 ~ #4 まで番号を付けた。

それらのクラスターに含まれる医薬品や、頻出する語の例をまとめたのが図5である。#1を除き、おおそ既知の薬効分類に従った分類結果になっており、化学構造も比較的類似している。#1は比較的ゆるい結合をしたクラスターであり、様々な薬理作用と構造の医薬品が属していた。

今後は重要語を選び出すプロセスやクラスタリングのプロセスについて詳細に検討したいと考えている。この研究は薬理作用と副作用の理解や、医薬品検索システムの利便性向上に役立つことが期待される。

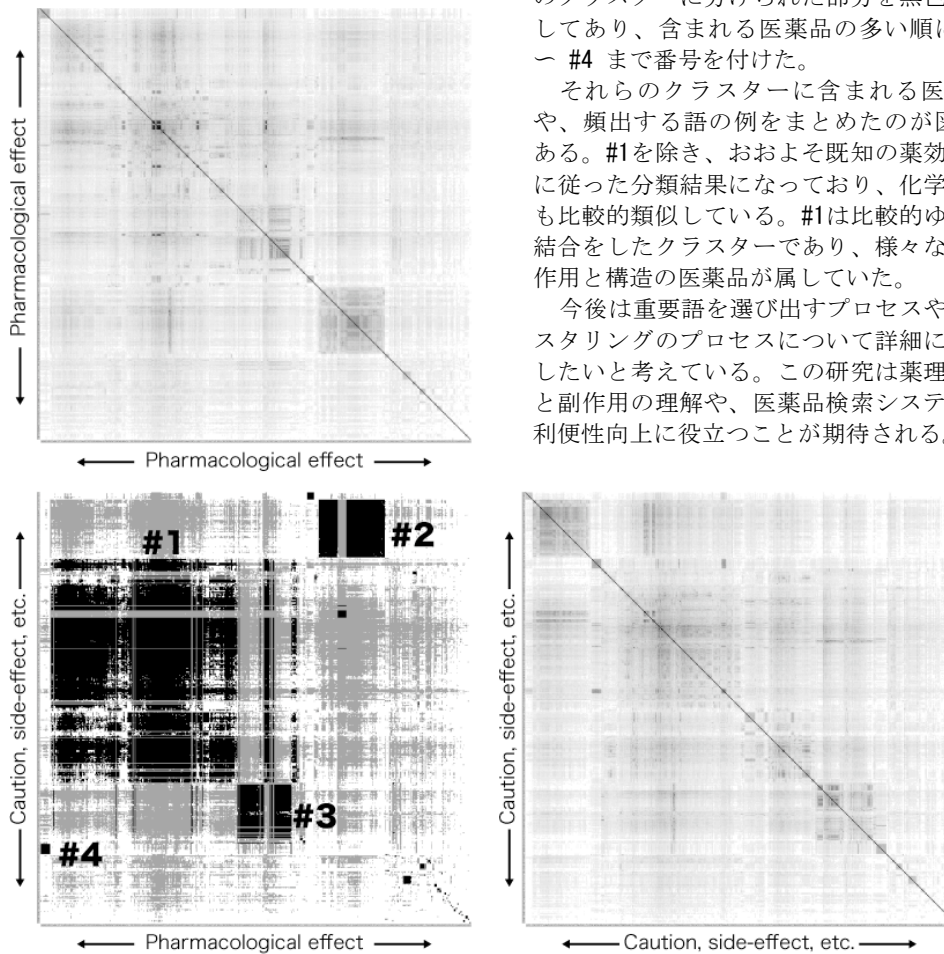
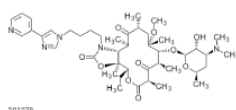


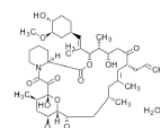
図4. 薬理作用と注意事項に基づいた医薬品のクラスタリング

#1: 270 drugs

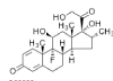
主な頻出語とその頻度:
 0.60 hypertension (高血圧)
 0.59 myalgia (筋肉痛)
 0.59 exacerbation (増悪)
 0.54 creatinine (クレアチニン)
 0.53 arthralgia (関節痛)



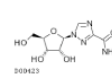
Telithromycin
(Antimicrobial)



Tacrolimus
(Immunosuppressant)



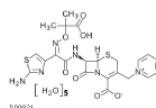
Dexamethasone
(Glucocorticoid)



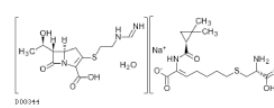
Ribavirin
(Antiviral)

#2: 80 drugs

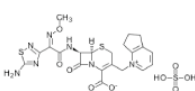
主な頻出語とその頻度:
 1.00 antibacterial (抗菌)
 0.98 pneumonia (肺炎)
 0.94 antibiotic (抗生物質)
 0.93 nephritis (腎炎)
 0.90 glossitis (舌炎)



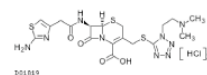
Ceftazidime
(Antibacterial)



Imipenem hydrate - cilastatin sodium
(Antibacterial)



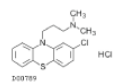
Cefpirome sulfate
(Antibacterial)



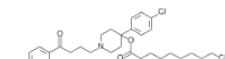
Cefotiam hydrochloride
(Antibacterial)

#3: 70 drugs

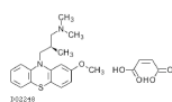
主な頻出語とその頻度:
 0.93 machine (機械)
 0.91 coma (昏睡)
 0.81 malignant (悪性)
 0.79 swallowing (飲み込み)
 0.79 epilepsy (てんかん)



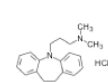
Chlorpromazine hydrochloride
(Anti-emetic, Antipsychotic)



Haloperidol decanoate
(Antipsychotic)



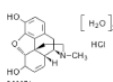
Levomepromazine maleate
(Analgesic [central nervous system depressant])



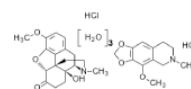
Imipramine hydrochloride
(Antidepressant)

#4: 14 drugs

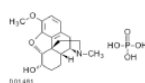
主な頻出語とその頻度:
 1.00 yawning (あくび)
 1.00 vasopressor (昇圧剤、血管収縮)
 1.00 urination (排尿)
 1.00 thyroid (甲状腺)
 1.00 teratogenicity (催奇形性)



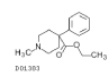
Morphine hydrochloride
(Narcotics)



Compound oxycodone
(Narcotics)



Dihydrocodeine phosphate
(Narcotics)



Pethidine hydrochloride
(Analgesic [narcotic])

図5. 各クラスターに属する医薬品と頻出語

文献・ウェブサイト

- [1] ゲノムネット医薬品データベース <http://www.genome.jp/kusuri/>
- [2] JAPIC <http://www.japic.or.jp/>
- [3] KEGG DRUG <http://www.genome.jp/kegg/drug/>
- [4] MeCab <http://mecab.sourceforge.net/>
- [5] Life Science Dictionary Project (LSD) <http://lsd.pharm.kyoto-u.ac.jp/ja/index.html>