

○地引芳乃¹⁾、 土肥栄祐²⁾、 仁宮洸太^{3,4)}
建石 由佳⁵⁾、 山本泰智⁶⁾、 佐々木貴規⁷⁾、 藤原豊史⁶⁾

1) 明治大学大学院 先端数理科学研究科

2) 新潟大学脳研究所生命科学リソース研究センター脳病態解析分野

3) 国立保健医療科学院

4) 東京大学薬学系研究科

5) 科学技術振興機構バイオサイエンスデータベースセンター

6) 情報・システム研究機構ライフサイエンス統合データベースセンター

7) 明治大学総合数理学部

ゲノムデータの活用のためには診療録との紐付けが重要だが、
診療録が非構造化データであるため、データ間の紐付けが容易ではない。



診療録の中で症状及び所見を半自動的に抽出する**POET (Phenotype Ontology Extractor from Text)**を開発し
ゲノム解析により得られた情報と診療録とを組み合わせた活用を可能にするシステムの提供を目指す。

構造化データへの変換

→ 診療録から構造化データである**HPOの用語**を
完全一致検索で抽出する。

医療辞書による語彙の追加

新たなPOET専用のHPO日本語拡張版辞書の作成
10,670語 → 14,270語

助詞(は, の, が, を)の除去

11疾患、32件の症例報告に対して

臨床医のアノテーション結果と比較して性能評価

HPO(Human Phenotype Ontology)



- ・ 2008年発表の、ヒトにおいて生じる症状を詳細かつ標準化したオントロジー
 - ・ 現在16,000以上の用語が英語にて登録済
 - ・ 2018年に日本語版が公開
- **診療録はHPOに準拠した書き方になっていない**

完全一致検索

HPO日本語版には類義語が含まれていないため、
差異のある表現が完全一致検索だと抽出できない
(ex. 認知症/認知症状/痴呆)

ライフサイエンス辞書 (15万語)
万病辞書 (38万語)



[About](#)
[Datasets](#)
[History](#)
[Terms](#)
[RDF](#)
[API](#)
[Contact](#)
[EN](#) / [JA](#)

日本語の診療録から症状を自動で抽出(アノテーション)できます。
尚、ブラウザの機能を利用して処理するため、入力したテキストがインターネット上で通信されることはありません。

以前のアノテーションを再開

患者の診療録を入力

自 診療録の例

現病歴：
38歳ころより四肢の舞踏運動、歩行障害が出現し、徐々に増悪してきた。43歳時某院神経内科に入院したが、44歳ころより医師や看護婦の制止を聞かず絶え間なく食べ物を口にしたり、自分の要求、たとえば酒が飲みたいなどの要求がかなえられないと易怒的となり家人に暴力をふるうなどの人格変化が出現してきた。また、痴呆症状に基づいて、タバコ火の不始末や終日廊下を徘徊する行為もみられた。これらの精神症状出現のため一般病棟での入院治療は困難となり、昭和61年10月27日（45歳時）、当院に転入院した。

症状を抽出

POET動画
<https://youtu.be/ymoiFemBnrA>

抽出結果 (アノテーション結果)

現病歴:

38歳ころより四肢の舞踏運動、歩行障害が出現し、徐々に増悪してきた。43歳時某院神経内科に入院したが、44歳ころより医師や看護婦の制止を聞かず絶え間なく食べ物を口にしたり、自分の要求、たとえば酒が飲みたいなどの要求がかなえられないと易怒的となり家人に暴力をふるうなどの人格変化が出現してきた。また、痴呆症状に基づいて、タバコ火の不始末や終日廊下を徘徊する行為もみられた。これらの精神症状出現のため一般病棟での入院治療は困難となり、昭和61年10月27日(45歳時)、当院に転入院した。

神経学的には顔面、四肢、体幹に激しい舞踏運動が観察され、緊張により増強し、睡眠時には消失した。また下肢の腱反射尤進、仮性足間代、筋緊張低下もみられた。精神医学的には頑固で、自己中心的、易怒的で、自分の要求、たとえば間食にパンを何度もねだるなどの要求がかなえられないと激しい不機嫌状態になるなどの人格変化がみられた。また、間食のこと以外にはほとんど興味を示さず、臥床がちで意欲の低下も認められた。質問に対してはすぐに"わからん"と答えるなどなげやりの態度がみられたこともあったが、問い返すとほぼ正確な返答があった。発語は不明瞭で返答に時間を要することが多かった。以前に入院していた病院のことや数日前に来院した妻のことなど憶えており、記憶力は比較的保たれていた。ただし、同じことを何度も説明するも理解できず、何回となく聞き返したりするなど理解力の低下がみられた。また、3桁の計算もできなくなってきた。舞踏運動に対してはハロペリドール、チアプリドの投与を行ったが、

合計 6件(症状あり6件、症状なし0件)

• 歩行障害・歩行障害(HP:0001288)



• 筋緊張低下・筋緊張低下(HP:0001252)



• 不随意運動・不随意運動(HP:0004305)



• 構音障害・構音障害(HP:0001260)



• 嚥下障害・嚥下障害(HP:0002015)



• 低体温・低体温(HP:0002045)



アノテーションを保存

POET 辞書拡張前



合計 13件(症状あり13件、症状なし0件)

- 歩行障害・歩行障害(HP:0001288) 🗑️ ⋮
- 人格変化・人格変化(HP:0000751) POET 🗑️ ⋮
- 痴呆症状・痴呆症状(HP:0000726) POET 🗑️ ⋮
- 間代・間代(HP:0002169) POET 🗑️ ⋮
- 筋緊張低下・筋緊張低下(HP:0001252) 🗑️ ⋮
- 人格変化・人格変化(HP:0000751) POET 🗑️ ⋮
- 不随意運動・不随意運動(HP:0004305) 🗑️ ⋮
- 構音障害・構音障害(HP:0001260) 🗑️ ⋮
- 嚥下障害・嚥下障害(HP:0002015) 🗑️ ⋮
- 不能・不能(HP:0000802) POET 🗑️ ⋮

POET 辞書拡張後



合計 35件(症状あり34件、症状なし1件)

- 舞踏運動・舞踏病(HP:0002072) 臨床医 🗑️ ⋮
- 歩行障害・歩行障害(HP:0001288) 🗑️ ⋮
- 制止を聞かず・脱抑制(HP:0000734) 臨床医 🗑️ ⋮
- 絶え間なく食べ物を口にしたり・異常な摂食行動(HP:010000000) 臨床医 🗑️ ⋮
- 自分の要求.たとえば酒が飲みたいなどの要求がかなえられないと易怒的・フラストレーション耐性低下(HP:0000744) 臨床医 🗑️ ⋮
- 家人に暴力をふるう・暴力的行動(HP:0008760) 臨床医 🗑️ ⋮
- 人格変化・性格変化(HP:0000751) POET 🗑️ ⋮
- 痴呆症状・記憶障害(HP:0002354) POET 🗑️ ⋮
- タバコ火の不始末・進行性健忘(HP:0007017) 臨床医 🗑️ ⋮

臨床医のアノテーション結果データ

自動化部分

結果

- ・アノテーションの総数が増えるのでprecisionが下がるのは予測できていたが、recallがあまり上がらず、結果としてF値は下がった。

POET 拡張前

precision	75.5%
recall	32.1%
F	45.1%

問題点

- ・万病辞書 : 上下概念が混ざる
- ・ライフサイエンス辞書 : ただの英単語が混ざる
- ・略称として登録されている英単語が誤って抜かれていることがある

POET 拡張後

precision	56.7%	↓ -18.8%
recall	33.9%	↑ +1.8%
F	42.5%	↓ -2.6%

今後

- ・語彙は約1.5倍程度に増えているので、語彙を増やす手法以外の改善策を考える
(臨床医からのフィードバックがもらいやすい形に改良する)
- ・助詞の抜き取り処理の精度を上げる(ex. 前後が漢字なら抜き出す)

POET 拡張後(助詞処理)

precision	60.0%	↓ -15.5%
recall	34.6%	↑ +2.5%
F	42.7%	↓ -2.4%

手動化部分

今回、臨床医がアノテーションをした際に
明らかとなった重要な課題は
HPOの全体像が掴めない点。

1. アノテーションが出来る単語が
簡単に見つからず、**時間・労力がかかる。**
2. アノテーションが可能な単語があるかが分からない
→自信を持って、この語句/フレーズに
対応するHPOが無い！と言い切れない。



該当するHPOの一つ上/下の概念のみしか確認できない

臨床医にとって分かりやすいHPOの可視化・GUIの実装が半自動システムとしてのアノテーションし易さの向上、
また、臨床医の協力の上での辞書の充実なども含め、必須であると考えられた。