

一般口演 | 医療支援

一般口演2 医療支援

2021年11月19日(金) 09:10 ～ 11:10 F会場 (2号館2階224)

[2-F-1-07] コンパクトな疾患知識ベース Illness scriptを利用した問診票推論の実現

*野田 和敬¹、柳田 育孝¹、横川 大樹¹、上原 孝紀¹、平野 陽介²、生坂 政臣¹（1. 千葉大学医学部附属病院 総合診療科, 2. ハロラボ合同会社）

*Kazutaka Noda¹, Yasutaka Yanagita¹, Daiki Yokokawa¹, Takanori Uehara¹, Yosuke Hirano², Masatomi Ikusaka¹（1. 千葉大学医学部附属病院 総合診療科, 2. ハロラボ合同会社）

キーワード：Diagnostic Decision Support System, Symptom Checker, Knowledge Base, Illness Script

背景：病歴に基づく推論を行う症状チェッカーや診断支援システムでは、知識ベースが未だ重要な役割を担っている。しかし、既存ツールは推論時に求められる入力項目数が多く、診断補助ツール普及の妨げになっている可能性がある。一方、診断のエキスパートは問診票の情報だけでも効果的な推論を行うことができる（Int J Gen Med 2013;7:13-9.）。そこで、これらのエキスパートが利用している疾患の特徴を端的に表した illness script を知識ベースとして整備し、少ない病歴情報量での推論の正確性を検証した。

目的：各疾患の illness script を知識ベースとして整備し、問診票相当の入力による推論の正確性を他の既存ツールと比較する。

方法：570疾患について illness script〔有病率、年齢、性別、key feature（症状、発症様式、受診までの期間、症状推移）〕に関する情報を key-value形式のデータ構造で知識ベースに整備し、問診票相当の入力によって疾患リストを生成する単純なプログラム（以下、ポケット問診）を作成した。Semigranら（BMJ 2015;351:h3480）が作成した45症例のヴィネットについて医師3名がポケット問診に入力し、正答率を Semigranらの調査で対象とされたツールと比較した。

結果：ポケット問診の上位1位の正答率は44%（36-52%, 95%CI）、上位3位以内では67%（59-75%, 95%CI）であり、Semigranらが調査した23ツールと比較してポケット問診は上位1位の正答率では3番目、上位3位以内の正答率では4番目に高かった。

結論：Illness scriptを構成要素とした知識ベースにより、問診票相当の病歴情報量でも、既存ツールと同等の診断精度が得られうる。

コンパクトな疾患知識ベース“*Illness script*”を利用した問診票推論の実現

野田和敬^{*1}、柳田育孝^{*1}、横川大樹^{*1}、上原孝紀^{*1}、平野陽介^{*2}、生坂政臣^{*1}

^{*1} 千葉大学医学部附属病院、^{*2} ハロラボ合同会社

Achieving efficient diagnostic reasoning by a few essential medical information “*illness scripts*” derived from interview sheets

Kazutaka Noda^{*1}, Yasutaka Yanagita^{*1}, Daiki Yokokawa^{*1}, Takanori Uehara^{*1}, Yosuke Hirano^{*2}, Masatomi Ikusaka^{*1}

^{*1} Department of General Medicine, Chiba University Hospital ^{*2} Harolabo G.K.

Abstract

INTRODUCTION: The knowledge base still plays an important role in symptom checkers, and diagnostic decision support systems (DDSS) that make inferences from medical history. However, existing tools require many inputs at the time of inference, which can hinder the use of DDSS. Contrastingly, diagnostic experts can make effective inferences, just by using the information present in a questionnaire. Therefore, we prepared an *illness script* used by experts as a knowledge base, that simply expresses the characteristics of diseases and verifies the accuracy of inference with a small amount of medical history information.

OBJECTIVE: We developed a knowledge base using *illness scripts* and compared the accuracy of inferences with other existing tools by inputting data that is equivalent to that of a questionnaire.

METHOD: A knowledge base was designed for 570 diseases, using the information on *illness scripts* [prevalence, age, sex, key features (symptoms, onset pattern, time to consultation, and symptom transition)] in key-value format data structure. We created a simple program, “Pocket Monshin,” that generates a disease list by feeding in data that is equivalent to that of a questionnaire. Three doctors inputted 45 case vignettes compiled by Semigran et al. and compared its accuracy with other tools assessed by Semigran et al.

RESULTS: The top-1 accuracy of Pocket Monshin was 44% (95% confidence interval [CI]: 36–52%), and the top-3 accuracy of it was 67% (95% CI: 59–75%). Compared to the 23 tools investigated by Semigran et al., Pocket Monshin was the third highest in the top-1 accuracy and the fourth highest in the top-3 accuracy.

DISCUSSION: With a knowledge base composed of *illness scripts*, it is possible to obtain diagnostic accuracy equivalent to that of existing tools, even with the amount of medical history information equivalent to that of a questionnaire.

Keywords: Diagnostic Decision Support System, Symptom Checker, Knowledge Base

【緒論】

病歴(問診)に基づく推論をコンピューターが代行する症状チェッカーや診断支援システム(Diagnostic Decision Support System: DDSS)では、知識ベースが未だ重要な役割を担っている。しかし、推論時に求められる入力項目は既存ツールごとに異なっており、効率的な推論のために利用すべき病歴情報が確立されているとは言えない。また、既存ツールは推論時に求められる入力項目数が多く、診断補助ツール普及の妨げになっている可能性がある。

一方、診断のエキスパートは、年齢、性別、主訴、罹病期間、治療歴、既往歴、アレルギー歴、喫煙・飲酒歴などの問診票に記載された情報のみから効果的な推論を行うことができる⁽¹⁾。このことから、比較的限定された情報から成る知識ベースでもこれらの要点を押さえれば、効果的な推論を実現できる可能性があると考えられる。

疾患の特徴を端的にまとめた知識は *illness script* と呼ばれ、医師はこの要約された知識を脳内に形成し、蓄積することでその後の診断推論の効率化に役立っている⁽²⁾⁽³⁾。*illness script* は一般に定型化されていないが、多くの場合、病態生理や疫学的情報、時間経過、主要症状・所見、診断手順、治療法などが構成要素となっている⁽⁴⁾⁽⁵⁾。本研究では、*illness*

script を利用したコンパクトな知識ベースと、少ない病歴情報量から疾患リストを生成する単純なプログラムを用意し、その推論結果の正確性を検証し、病歴推論の際に活用すべき情報を明らかにする。

【目的】

主要疾患の *illness script* を知識ベースとして整備し、問診票相当の入力による推論の正確性を他の既存ツールと比較する。

【方法】

A. 知識ベース

高頻度疾患を中心に 570 疾患について *illness script* に関する情報を知識ベースに整備した。具体的には、年齢、性別のほか、臨床的に鍵となる特徴(以下、**key feature**)として、主訴となり得る症状、発症様式、受診までの期間、症状推移、その他の主たる付随情報(主要随伴症状やリスク因子)について、千葉大学医学部附属病院総合診療科(診断推論学)に所属する医師4名(臨床経験6年～18年)が記述作業を行った。知識ベースに収載した領域別疾患数を図1に示す。年齢は層別化して頻度情報を付与した。主訴症状や付随情報としては、各疾患を想起させるキーワードを疾患ごとに選定し

て登録した。このキーワードにはのべ約 900 用語が登録された。発症様式には「突発、急性、亜急性、緩徐」、受診までの期間には「24 時間未満(時間単位)、1 日～7 日(日単位)、1～4 週(週単位)、1～12 か月(月単位)、1 年以上(年単位)」、症状推移には「悪化、横ばい、改善、間欠期がある」をそれぞれ用意した。登録された key feature のそれぞれに対して、その情報によって各疾患を想起する強度を表すスコアを付与した。知識ベースのデータ構造の例を表 1 に示す。

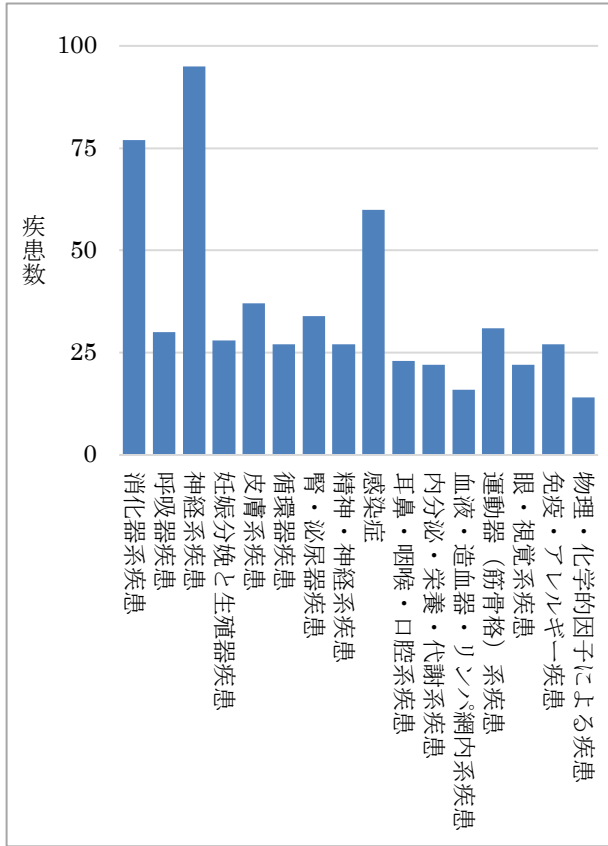


図1 知識ベースに収載した領域別疾患数。

表1 知識ベースのデータ構造の例。(実際のデータとは異なる)

急性心筋梗塞	
急性	8
胸痛	7
冷や汗	7
⋮	⋮

B. 推論プログラム

次に、問診票相当の入力、すなわち、年齢、性別、主訴、いつからの症状か、発症様式(突然、急性、亜急性、緩徐)、症状の程度の推移、その他の症状、の入力によって、可能性の高い疾患のリストを生成する単純なプログラム(以下、ポケット問診)を作成した。入力はプログラムが呈示する項目を選択する形で行われるものとし、それらの項目は検索によって候補が呼び出されるようにした。プログラムの入力値および出力結果の例を図2に示す。

年齢：64歳 性別：男性

主訴

胸痛 × 1日以上7日未満 急性 悪化する

他の症状

冷や汗 × 呼吸困難 × 胸部圧迫感 ×

他の症状を追加

推論結果

急性心筋梗塞

不安定狭心症

労作性狭心症

心タンポナーデ

大動脈解離

大動脈弁狭窄症

パニック障害

冠攣縮性狭心症

特発性食道破裂

縦隔気腫

自然気胸

急性肺塞栓症

図2 推論プログラム(ポケット問診)の入力値および出力結果の例。

C. 推論精度の評価

Semigran ら⁽⁶⁾が作成した 45 症例のヴィネットを検証対象データとした。知識ベースの構築やポケット問診の開発に関与していない医師 3 名にポケット問診への項目入力と推論結果の保存を依頼した。このとき協力医師には研究の目的を開示せず、症例ヴィネットを Google 翻訳にて日本語に自動翻訳したものを配布した。それぞれの症例において入力する項目や内容については、医師ごとに判断させた。3 名の推論結果で、出力された疾患リストの上位 1 位(以下、TOP1)および上位 3 位以内(以下、TOP3)に正解の疾患が含まれる症例の割合をそれぞれ算出した(以下、正答率)。TOP1 および TOP3 の正答率を Semigran らの調査⁽⁶⁾で対象とされたツールの結果と比較した。

【結果】

検証に用いた症例の診断および医師 3 名の入力による推論結果を表2に、正答率のまとめを表3に示す。検証したのべ 135 症例全体でみると、TOP1 正答率は医師1が 33%、医師2が 49%、医師3が 51%で、平均 44%であった。TOP3 正答率は医師1が 60%、医師2が 69%、医師3が 71%で、平均 67%であった。緊急度別の正答率では非緊急症例での TOP3 正答率がいずれの医師でも高い傾向にあったが、その他のものも含めて有意差はなかった。

Semigran らの調査結果⁽⁶⁾では、調査された 23 ツールのうち、TOP1 正答率では上位から順に DocResponse 50%(33-67、95%CI)、Family Doctor 47%(31-62、95%CI)、Isabel 44%(29-60、95%CI)、Ask MD 43%(26-59、95%CI)となっている。TOP3 正答率では同様に Symcat 71%(57-85、95%CI)、Isabel 69%(55-83、95%CI)、Ask MD 68%(52-83、95%CI)、

DocResponse 67% (50-83, 95%CI) となっている。したがって、ポケット問診をそれらと比較すると、TOP1 正答率では Isabel と同率で第 3 位、TOP3 正答率では Ask MD に次ぐ第 4 位であった。

表2 検証症例の診断⁽⁶⁾および推論結果一覧。(数字:推論結果の疾患リスト内の順位、-:疾患リストの5位以内に含まれなかった、×:知識ベースに未収載の疾患)

No	診断	医師 1	医師 2	医師 3
緊急				
1	急性肝不全（アセトアミノフェン中毒も許容）	-	3	-
2	虫垂炎	4	2	3
3	喘息（または急性喘息、喘息状態、喘息の悪化）	2	1	1
4	COPD の悪化（重度）	3	4	2
5	深部静脈血栓症	-	-	1
6	心筋梗塞	-	1	1
7	溶血性尿毒症症候群	-	-	-
8	尿管結石	2	2	1
9	マラリア	-	-	4
10	髄膜炎	4	4	1
11	肺炎	1	2	1
12	肺塞栓症	2	2	3
13	ロッキー山紅斑熱	1	1	1
14	脳血管障害	-	-	-
15	破傷風	1	1	1
非緊急				
16	急性中耳炎（上気道炎またはウイルス感染も許容）	2	5	-
17	扁桃炎（または急性咽頭炎、咽頭感染症、上気道炎）	1	1	1
18	急性咽頭炎	1	1	2
19	急性副鼻腔炎（または副鼻腔炎、慢性副鼻腔炎、慢性副鼻腔炎）	2	2	2
20	腰椎捻挫（または急性腰椎捻挫、急性腰椎捻挫、腰椎捻挫、腰椎損傷、椎間板ヘルニア、坐骨神経痛）	2	3	-
21	蜂窩織炎	-	-	-
22	COPD の悪化（軽度）	3	1	2
23	インフルエンザ	2	1	2
24	伝染性単核球症	-	1	1
25	消化性潰瘍疾患（または GERD、十二指腸潰瘍、胃潰瘍）	5	1	1
26	肺炎	1	1	1

27	サルモネラ	1	1	1
28	帯状疱疹	1	1	1
29	膀胱炎	1	1	1
30	良性発作性頭位めまい症	2	1	1
セルフケア				
31	急性気管支炎（気管支炎または上気道炎も許容）	1	5	1
32	急性気管支炎	-	-	-
33	結膜炎	1	1	1
34	急性咽頭炎	1	1	1
35	アレルギー性鼻炎	2	1	1
36	急性腰痛症	2	3	2
37	アナフィラキシーのないハチ刺傷	-	-	-
38	アフタ性潰瘍（口内炎）	1	1	1
39	カンジダ	2	2	2
40	便秘	×	×	×
41	湿疹	×	×	×
42	麦粒腫	-	1	-
43	ウイルス性上気道炎	1	1	1
44	ウイルス性上気道炎	1	1	1
45	嘔吐	×	×	×

表3 推論プログラム(ポケット問診)の正答率

		医師 1		医師 2		医師 3		平均		
		正 答 数	%	正 答 数	%	正 答 数	%	正 答 数	%	(95%CI)
全 体	TOP1	15 /45	33	22 /45	49	23 /45	51	60 /135	44	(36-52)
	TOP3	28 /45	62	31 /45	69	32 /45	71	91 /135	67	(59-75)
緊 急	TOP1	3 /15	20	4 /15	27	8 /15	53	15 /45	33	(19-47)
	TOP3	7 /15	47	9 /15	60	11 /15	73	27 /45	60	(46-74)
非 緊 急	TOP1	6 /15	40	11 /15	73	8 /15	53	25 /45	56	(41-71)
	TOP3	12 /15	80	13 /15	87	12 /15	80	37 /45	82	(71-93)
セ ル フ ケ ア	TOP1	6 /15	40	7 /15	47	7 /15	47	20 /45	44	(29-59)
	TOP3	9 /15	60	9 /15	60	9 /15	60	27 /45	60	(46-74)

【考察】

本研究では原始的な知識ベースに、病歴推論を行う上で臨床的に特に重要と考えられる情報を厳選して登録した。推論プログラムについても入力する情報はごくわずかで、かつ、単純な計算アルゴリズムを用いている。それらによって算出された推論結果は、より高度に成熟していると考えられる他の既存ツールと比して、同等の正答率を達成した。このことは、大局的にみれば **illness script** として包含されるような病歴情報が推論上極めて重要であることを示唆している。

DDSS の発展において近年は大量のデータによる機械学習に力点が置かれているが、機械学習と知識ベースの両者の長所を組み合わせたアプローチの重要性も指摘されている⁷⁾。あらゆる病歴情報を全て扱おうとすれば、ほぼ無限の域に達し、手作業での整備は当然実現せず、機械的に処理するにしてもそれぞれの情報の重要性に差異を生じるまでに膨大なデータ量と計算量が必要になる。したがって、いずれにしても扱う病歴情報は優先順位をつけて処理すべきである。その筆頭候補として本研究で扱った「年齢、性別、key feature (主訴となり得る症状、発症様式、受診までの期間、症状推移、その他の主たる付随情報)」を提案する。

また、既存の DDSS は臨床現場でのワークフローと整合性が取れておらず、冗長な入力がそれらの利用の妨げになっているとの指摘がある⁸⁾。本研究で取り上げた病歴情報は問診票で聴取できる程度の情報量であり、推論時にユーザーが入力する上でも少ない労力で済む。少なくとも推論の初期にこれらを活用することによって効率的に鑑別疾患を絞り込むことが可能になると推察される。

正答率は入力者によってばらつきがあり、TOP1正答率では 33%から 51%、TOP3正答率では 62%から 71%と幅があった。これは各医師の入力値のばらつきに起因している。各症例で入力する項目や内容は、症例ヴィネットから情報を読み取って各医師に判断してもらったが、発症様式や受診までの期間などの情報がヴィネットに明記されていないものもあり、ばらつきの一因となっていた。ポケット問診は少ない入力で推論結果を出力するプログラムであるため、推論結果は入力値の影響を大きく受けやすかった。

知識ベースの拡充や更新には継続的な管理が必要になる。そのうえ、推論を少ない情報量で実行するため、その推論精度は知識ベースに格納されるデータの個々の質に依存する度合いが大きくなる。本研究では病歴推論のエキスパートがそれらの情報の重み付けを行ったが、持続可能性が今後の課題である。

【本研究の限界】

検証に用いる症例が異なれば推論結果も変わるため、正答率に影響を及ぼす可能性がある。また、知識ベースの収載疾患数を拡充させれば同一主訴に対する候補疾患が増えるため、精度が悪化する可能性がある。

【結論】

Illness script を構成要素とした知識ベースにより、問診票相当の病歴情報量でも、既存ツールと同等の診断精度が得られる。

文献

- 1) Uehara T, Ikusaka M, Ohira Y, Ohta M, Noda K, Tsukamoto T, et al. Accuracy of diagnoses predicted from a simple patient questionnaire stratified by the duration of general ambulatory

training: an observational study. *Int J Gen Med*. 2013;7:13-9.

- 2) Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ*. 2007;41:1178-84.
- 3) Bowen JL. Educational Strategies to Promote Clinical Diagnostic Reasoning. *N Engl J Med*. 2006 Nov 23;355(21):2217-25.
- 4) Schmidt HG, Rikers RMJP. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ*. 2007 Nov 14;41:1133-9.
- 5) Illness Scripts Overview. *J Gen Intern Med* [Internet]. [https://www.sgim.org/web-only/clinical-reasoning-exercises/illness-scripts-overview# (cited 2021-Aug-21)]
- 6) Semigran HL, Linder J a, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ*. 2015;351:h3480.
- 7) Müller L, Gangadharaiah R, Klein SC, Perry J, Bernstein G, Nurkse D, et al. An open access medical knowledge base for community driven diagnostic decision support system development. *BMC Med Inform Decis Mak*. 2019;19(1).
- 8) Cahan A, Cimino JJ. A Learning Health Care System Using Computer-Aided Diagnosis. *J Med Internet Res*. 2017;19(3):e54.