

Precision Medicine (精密医療) (総説) New Engl J Med Oct.11,2018

西伊豆早朝カンファランス

「僻地で世界最先端！」西伊豆健育会病院 H30.10 仲田和正

Classification, Ontology, and Precision Medicine (Review Article)

著者

Melissa A. Haendel, Ph.D. オレゴン健康科学大学、Linus Pauling 研究所

Christopher G. Chute, M.D. Dr.P.H. Johns Hopkins University Schools of
Medicine

Peter N. Robinson, M.D. コネチカット大学、Jackson Laboratory for Genomic
Medicine

New Engl J Med の 2018 年 10 月 11 日号に Precision Medicine (精密医療) の総説
がありました。

2015 年、オバマ大統領が一般教書演説でこのテーマを取り上げ、今後米国は
Precision medicine に注力することを宣言しました。

2016 年には米国で 2.15 億ドル (1 ドル 112 円として 241 億円) をこの研究に
投ずるといのです。

このオバマ大統領の演説がバンドワゴン (パレードの先頭の楽隊) 効果となり
precision medicine (精密医療) に突然関心がもたれるようになりました。
従来の治療は one-size-fits-all で平均的な患者のための治療でした。

Precision medicine とは、ヒトゲノムと個人の詳細情報を基にした医療のことです。
例えば乳がん患者でハーセプチン (trastuzumab) は遺伝子の HER2 ドメイン
(細胞成長と 増殖に関与する tyrosine kinase) を標的とするモノクローナル抗体
です。

もし患者のガン細胞遺伝子が HER2 陽性であればハーセプチンの良い適応になります。

また乳癌の 5 から 10%は遺伝性であり、BRCA1 と BRCA2 遺伝子 が主に関与し、
70 歳までの乳がん発症リスクは何と 80 から 85%になります。

ブラッド・ピットの奥さんのアンジェリーナ・ジョリーは映画「トゥームレイダー」
の主演女優ですが BRCA1 遺伝子陽性であったことから 2013 年に予防的に自分の乳
房切除を行いました。

このように先制的 (preemptive) に乳房切除を行ったことは世界に衝撃を与え
Angelina effect と呼ばれました。

また BRCA1 は卵巣がんも起こし得ることから彼女は 2015 年には卵巣、卵管切除も
行っています。

これから世界の医療は一体どうなっていくのか小生、興味を引かれこの総説を読んでみました。

しかしこの総説は Precision Medicine の各論ではなく、その方法論の解説でした。

NEJM 総説「Precision medicine」最重要点は下記7点です。

- ・ Precision medicine のゴールは患者を層別化して類似性を発見すること。
- ・ オントロジーによりデータを計算可能 (computable)かつ分類可能 (classifiable)とせよ。
- ・ オントロジーで概念の関係づけができて初めて論理的推論、機械学習が可能となる。
- ・ データは最終的に structure (構造) と semantics(意味)に分解される。
- ・ SNOMED-CT は記述論理による疾病分類のオントロジー、日本未加盟。
- ・ 同じ概念に複数の用語があると病院間のデータ比較はできぬ。用語の明言が必要。
- ・ 知識データは可視化、部分化、関連、俯瞰、連想、創造、支援の7つを目指せ！

1. Precision medicine のゴールは患者を層別化して類似性を発見すること。

この総説の結論は「Precision medicine のゴールは診断、治療改善のために患者を層別化することだ」と言うのです。

当たり前で、どうでもいいことを言うもんだなあと小生、最初は思ったのですが、Precision medicine (精密医療) を行う為には膨大な、異なる臨床情報、科学的情報の統合が必要であり、そのツールがなくては始まらないと言うのです。

この共通プラットフォームであるツールを作ることこそが precision medicine の一丁目一番地だと言うのです。ツールさえ作れば、あとは急速に発達します。

保存されず再利用できないようなデータは知識ではありません。

死蔵されている知識は価値を生まないのです。

また再利用の度に付加価値がつかないような知識システムもシステムと言えません。

知識融合による新知識の生成は極めて一般的な知識生成方法です。

しかし単に知識を収集して統合すれば新知識が生成される訳でもありません。

知識の構造化が必要なのです。

一般に知識の供給側はできるだけ知識を細分化しようとします。

医学学会の限りない細分化を見ればよくわかります。

しかし需要側は融合、統合された知識を欲しています。

現在、医学知識は爆発的に急膨張し、もはやその全貌を俯瞰することが困難になってきました。しかし知識の構造化により知識を統合することができます。異質データの知識、概念を統合、分析するためのツールを情報工学では「オントロジー (ontology)」と言います。

オン (ov) とはギリシャ語の「存在」のことで ontology はもともと形而上学のことです。「神は存在するのか?」「時間の始まり、終わりはあるのか?」「最小単位はあるのか?」などの哲学をすることが ontology (形而上学) でした。

これらの問題はカントが二律背反問題 (Antinomie) として純粋理性批判で「証明不能である」ことを証明しました。

「神の存在の証明は不可能であるし、また存在しないことの証明も不可能」なのです。考えるだけ無駄なので最近の哲学ではこんな問題はもう考えません。

小生の外来にレヴィー小体型認知症の方が通っています。

長谷川のテストでは点数は低くないのですが、人、動物、虫などの幻視があります。ベッドの上で毛布がモコモコしているとその一つ一つが人の顔に見えると言うのです。またテレビ画面の中でさえ幻視が現れます。

外来で初めてこの話を伺ったとき、これってフッサールの現象学の問題そのものだよなと興奮しました。

現象学では「一見リアルに見えている世界の風景の一切を意識に生じた表象にすぎないもの」と見なします。客観があってそれを見て主観が生ずるのでなく人間の主観、つまり意識が世界を秩序づけると逆に考えるのです。

小生が見ている景色が本物なのか、それともこの患者さんが見ているものが本物なのか誰にも証明できません。

以前、奈良長谷寺の境内で次男がビデオカメラを落としました。

後でその映像を見たところ画面にところどころモザイクがかかるのです。

その時、自分が実際に見ているものは本当だろうかと妙にギクリとしました。

「形而上」と言うと小生、日露戦争の時の「連合艦隊解散の辞」を思い出します。「二十閱 (えつ：過ぎ去る) 月の征戦已 (すで) に往事と過ぎ、我が連合艦隊は今や其の隊務を結了して茲 (ここ) に解散する事となれり」で始まる東郷平八郎の名文です。

この文にはルーズベルト大統領が感激し英訳されて米国の将兵に配布されました。

特に感動するのは次の下りです。

「而 (しか) して武力なる物は艦船兵器等のみにあらずして、之を活用する無形の実力にあり。百発百中の一砲能 (よ) く百発一中の敵砲百門に対抗し得るを覚らば我ら軍人は主として武力を形而上に求めざるべからず」

当、西伊豆健育会病院のような小病院も、実力は病院のハードでなく無形の実力である「形而上」の知識に求めざるを得ません。

そして「連合艦隊解散の辞」は次のように終わります。

「神明（神）は唯平素の鍛錬に努め戦わずして既に勝てる者に勝利の栄冠を授けると同時に、一勝に満足し治平に安んずる者より直ちにこれを奪う。

古人曰く勝って兜の緒を締めよと。

明治 38 年 12 月 21 日 連合艦隊司令長官 東郷平八郎」

電子カルテ（EHR、electronic health records）により臨床情報の収集は大変容易となりました。しかし電子カルテ情報と言っても、記載が不十分、不正確、出所不明、PDF でアクセス不能だったりします。また電子カルテは色々な会社のものがあり、システムが違えば相互の参照ができません。

このため情報の類似性がわからず分類ができないと言うのです。

例えば電子カルテで「心筋梗塞」を検索した場合、その患者自身のことでなく「心筋梗塞の家族歴」なのかもしれないし「心筋梗塞の既往はない」という文脈かもしれません（contextual meaning）。

また遺伝子情報は膨大な処理量のデータ（high-throughput data）です。患者を層別化して疾患間の類似性を発見することが precision medicine のゴールです。

ひとつの遺伝疾患でも多様な表現型（phenotype）があり見ただけではわからぬ（fuzzy match）かもしれないのです。

特に年齢が異なれば同じ遺伝疾患でも表現型が随分異なります。

2. データを処理可能（computable）かつ分類可能（classifiable）とせよ。

異なる分野のフォーマット、構造を統合しコンピューター処理可能（computable）とし分類可能（classifiable）とする必要があるのです。

これを行うプラットフォームを情報工学ではオントロジーと言います。

Computable かつ classifiable がキーワードです。

オントロジーにより基礎科学データ、遺伝情報、電子カルテ間のやり取りが初めて可能となるのです。

今まで小生気にしたことはなかったのですが、用語の統一は極めて重要だと言うのです。

Pubmed 検索の際の MeSH(Medical Subject Headings)のように術語 (terminology) は名称の標準化、類義語リスト、相互参照 (cross reference) に必須です。

医学用語には今まで ICD (国際疾病分類、International Classification of Diseases) が使われてきました。

この第 1 版は 1893 年に発行され 2003 年に ICD-10 が発行されました。

例えば ICD-10 では連鎖球菌性敗血症は次のように分類されます。

【A40 連鎖球菌性敗血症】

A40.0 A 群連鎖球菌による敗血症

A40.1 B 群連鎖球菌による敗血症

A40.2 D 群連鎖球菌による敗血症

A40.3 肺炎連鎖球菌による敗血症

A40.8 その他の連鎖球菌性敗血症

A40.9 連鎖球菌性敗血症、詳細不明

これを見てわかるように ICD 分類は単一系統 (monohierarchy) で網羅的 (exhaustive) でかつ排他性(exclusiveness)が保たれそれぞれのコードが 1 病名、単一系統になっています。

しかしこのため逆に、同一疾患なのに他の表現型 (phenotypes) である場合、分析がうまくできないと言うのです。

また ICD のアルゴリズムは医療費算定目的で使われており鑑別診断が目的ではありません。

3. オントロジーで概念の関係づけができて初めて論理的推論、機械学習が可能となる。

一方、オントロジーは概念とその類義語、そして概念同士の関係を定義づける記述論理 (description-logic) です。

これにより多くの領域の膨大な概念間の論理的な一致が可能となります。

またビッグデータ内の潜在的知識をコード化 (encode) し処理可能 (computable) かつ意味あるもの (semantics) となり機械学習 (machine learning) が可能となるのです。

遺伝疾患では同じ疾患であってもその表現型は様々です。年齢が異なればそれだけで表現型は大きく異なります。オントロジーにより異なる病院間で、異なる表現型を持つ遺伝疾患を相互参照できるようになるのです。

また遺伝子バンクとの相互乗り入れも可能となります。

現在「**Human Phenotype Ontology**」と言うサイトがあり遺伝疾患の様々な症状（表現型）を入力すると候補疾患を挙げてくれます。
ネットですぐ出てきますが登録が必要です。

オントロジーには色々な論理記号が使われます。例えば下記のようなものです。

- ¬A 否定 (A ではない)
- A∧B 論理積 (A かつ B)
- A∨B 論理和 (A または B)
- ∀ x 全称量化 (全ての x)
- ∃ x 存在量化 (x が存在する)

4. データは最終的に **structure** (構造) と **semantics** (意味) に分解される。

データ規格 (**data standards**) は最終的に二つに落ち着きます。

Structure(構造)と **Semantics**(意味論)です。**Semantic** とは複数の概念、それらの間での関連性を言います。ただし **structure** と **semantics** は全く別の概念ではなく、交互に絡み合っています。

例えばオントロジーが「ウイルス」を感染性物質 (**infectious agent**) と定義し「感染性髄膜炎」を感染性物質による髄膜炎と定義したなら、「ウイルス性髄膜炎」は「感染性髄膜炎」の下の亜系 (**subclass**) と結論付けられます。

5. **SNOMED-CT** は記述論理による疾病分類のオントロジー、日本未加盟。

国際疾病分類(**ICD**)は単なる術語集 (**terminology**) ですが、これに対し **SNOMED-CT** (**Systematized Nomenclature of Medicine-Clinical Terms**) という疾病分類があります。
これは記述論理 (**description logic**) を使用したオントロジーです。

例えば「急性穿孔性虫垂炎、**acute perforated appendicitis**」は「**acute inflammation**」、「**perforation**」、「**appendix structure**」の概念を組み合わせて定義します。
これにより一貫した (**consistent**)、かつコンピューター処理可能な (**computable**) 概念となります。

また SNOMED-CT は異なるシステム上で様々なユーザーが使用でき同じ概念 (concept) を扱うことができます。
例えば肝臓がんは日本では肝がん、肝癌、HCC、hepatoma など表記は様々です。Hepatoma の「概念」に数字の識別子を付け、その概念の「表記」 (description) として複数の用語 (肝癌、肝がん、HCC、hapatoma) を関連付けるのです。

1980 年代、「見栄講座・ミーハーのための戦略と展開」という本がありました。人との会話では極力「コンセプト (概念)、ストラテジー (戦略)」の単語をさりげなく入れろとのことでした。そうすると知的に、格好よく見えます。

6. 同じ概念に複数の用語があると病院間のデータ比較はできぬ。用語の明言が必要。

Hepatoma の「概念」に数字の識別子を付け、その概念の「表記」 (description) として複数の用語 (肝癌、肝がん、HCC、hapatoma) を関連付けることにより、一つの概念に関連付けられた用語は同義語となります。

同じ概念に複数の用語がある時、複数病院間のデータ比較はできません。ソフトウェアシステムでは用語の明言 (assertion) が必要なのです。コンピューター言語は最終的には 0 と 1 でありその中間はありません。あいまいな言い方は許されないのです。

SNOMED-CT は 30 万 8,000 の概念、77 万 7,000 用語から成る、医療分野において最も大きな用語集の 1 つです。
英語とスペイン語版が公式に更新されており、半年に 1 回改訂されています。現在は 9 か国からなる IHTSDO (International Health Terminology Standards Development Organization) という国際機関によって管理されており、年間運営費は 6 億円を超えます。日本は非加盟です。
加盟国内では、商用・非商用を問わずに無償で利用することができます。

IHTSDO は、その運営費を各国の国内総生産 (GDP) に応じて、負担します。2007 年の IHTSDO の年間予算は、総額で 601 万 185 米ドルであり、その 72%を米国、12%を英国、5%をカナダが負担しています。
日本は加盟していませんが、もし日本が加盟した場合、全体予算の約 22%、年間 130 万米ドルを負担することになります。

しかし、SNOMED-CT の英語病名と日本語病名は 1 対 1 対応しませんから、これを日本国内で運用するのは難しいことだろうなあとつくづく思いました。言語の壁は非常に大きいと思います。

JAMA の Feb.23、2016 は一冊丸ごと敗血症の特集でした。

この中で敗血症性ショックの定義は「mBP \geq 60 を保つに昇圧剤を要し、かつ乳酸 18 mg/dl 以上」とされました。これは低血圧で循環不全を代表させ、乳酸値で細胞障害を代表させたのです。

この敗血症性ショック定義の「血圧、乳酸、昇圧剤使用」の 3 変数 (variables) の選択は、実に米国の感染症 318 万 4,340 例の膨大な電子カルテの分析から導かれたものです。わずか 3 変数の選択にこれだけの手間をかけたのです。

この大規模スタディは、全米の電子カルテを統合する共通プラットフォーム、オントロジーがあつてのことだったのだなあとおつくづく感心しました。

7. 知識データは可視化、部分化、関連、俯瞰、連想、創造、支援の 7 つを目指せ！

小生、この New Engl J Med の総説で、知識の層別化に興味を持ち、ネットで調べてみたところ元東大総長小宮山宏先生の「知識の構造化、Open Knowledge 社、2004 年」という本を見つけ読んでみました。

もう 30 年も前からこの先生は「知識の構造化」に興味を持っていたと言うのです。

専門家と非専門家を区別する内的条件は、関連付け能力にあると言います。専門家は知識間の関連付けが論理的です。論理的なため話の展開が広く深くなります。非専門家は断片的知識は持っていますが専門家のような関連付けができません。毎日ご飯を食べていても非専門家は、うまい、まずいしか言えません。

しかし専門家が集まっても専門家間の意見も分かれます。結論が異なるのは、どこかで知識間の関連付けが異なるためだと言うのです。どれが正しいというわけでもありません。

この本によると知識の構造化により、知識データで次の 7 つのことを可能とすべきとのことです。

この本に日本の各企業、日立、三菱重工、住友化学、NEC、花王、日揮、トヨタ、三井物産、三井住友銀行の知識管理システムの例が示されていましたが、下記 7 つの全ての項目で合格点の企業は一つもありませんでした。

顧客のクレーム対応には関連、支援ができていても創造、全体の俯瞰ができないといった具合です。

【知識の構造化により次の7つを目指せ】

- ① 可視化
- ② 部分化
- ③ 関連
- ④ 俯瞰
- ⑤ 連想
- ⑥ 創造
- ⑦ 支援

皆様！どなたか日本の医学の共通プラットフォーム、オントロジーを研究して頂けませんかねえ。

それでは NEJM 総説「Precision medicine」最重要点7つの怒涛の反復です。

- ・ Precision medicine のゴールは患者を層別化して類似性を発見すること。
- ・ オントロジーによりデータを計算可能 (computable) かつ分類可能(classifiable)とせよ。
- ・ オントロジーで概念の関係づけができて初めて論理的推論、機械学習が可能となる。
- ・ データは最終的に structure (構造) と semantics(意味)に分解される。
- ・ SNOMED-CT は記述論理による疾病分類のオントロジー、日本未加盟。
- ・ 同じ概念に複数の用語があると病院間のデータ比較はできぬ。用語の明言が必要。
- ・ 知識データは可視化、部分化、関連、俯瞰、連想、創造、支援の7つを目指せ！