ポスター

[PO-9～16、P-17～33] ポスター立会

2019年6月8日(土) 14:30 ～ 15:30 ポスター会場 (熊本市民会館　2F　ホワイエ)

[P-19] 深層学習によるレセプトデータからの疾患予測

中田 智史 （株式会社地球快適化インスティテュート）

©一般社団法人 日本医療情報学会

深層学習によるレセプトデータからの疾患予測

中田 智史[*1], 佐藤 尚哉[*1], 満武 巨裕[*2]
[*1] 株式会社地球快適化インスティテュート フューチャーデザイン室,
[*2] 一般財団法人 医療経済研究・社会保険福祉協会 医療経済研究機構 研究部

Predicting disease risk based on medical claim data by deep learning

Satoshi Nakata[*1], Naoya Sato[*1], Naohiro Mitsutake[*2]
[*1] Future Design Division, The KAITEKI Institute, Inc.
[*2] Research Department, Institute for Health Economics and Policy

抄録：増大し続ける医療費の適正化のため，早期に疾患リスクを捉えることで、予防活動につなげていくことが重要だと考える．そこで本研究では，レセプトデータを活用して，将来の疾患リスクを予測することを試行する．個人単位で保険診療行為を把握できるレセプトデータは，医療機関を跨いた横断的な情報を得ることができる。経年推移の追跡も可能なため，予測に用いることができると考えられる．今回，複雑な疾患の相互依存を抽出するために，患者ごとの疾患履歴から深層学習を用いた疾患リスクの予測，及び診療行為や患者の属性情報等のレセプトが持つ多様な情報も合わせた疾患予測の検討を行う．

キーワード レセプト，深層学習，ディープラーニング，疾患予測，予防医学

1. はじめに

我が国の急速に増大する医療費を適正化し，社会保険システムを維持していくためには，「予防や早期治療」への対応が重要である．そのため，疾患の発症リスクを早期に捉える疾病予測を試行する．

効果的な予防のためには疾患の相互依存を理解する必要があるが，その関連性は複雑である．しかし，患者間の共通したリスク要因は同様の結果をもたらすと想定して，患者ごとの疾患履歴，および診療行為や属性情報等のレセプト情報から深層学習を用いた疾患リスクの予測の検討を行った．

2. 方法

1) レセプトに基づく疾患予測

患者間における関係性に基づく予測を生成するために，協調フィルタリングを用いる．協調フィルタリングは，あるユーザと嗜好の類似した他のユーザの情報を用いて推論を行う方法論である．ユーザの嗜好の代わりに疾患（病名）の履歴を用い，レセプトから取得する．

(1) 深層学習の適用

協調フィルタリングを実装する統計モデルとして，ニューラルネットの一つであるVAE(Variational Autoencoders) [1]を用いた．VAE は入力よりも低い次元の中間層を持ち，入力を再現するように学習することにより，低次元の潜在変数でデータの関係を効果的に捉えることができる．そのため，データに潜む関係性から推論を行う協調フィルタリングにおいても有効性が期待できる[2]．

各疾患の有無を 0,1 で表したベクトルで患者を表現し，入力とする(Fig. 1)．出力は入力と同次元の実数値ベクトルで，各疾患の罹り易さを表す．

	疾患1	疾患2	疾患3	疾患4	…	疾患n
患者1	0	1	1	0	…	0
患者2	1	0	0	0	…	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
患者N	0	1	0	1	…	1

Fig.1 ニューラルネットの入力

(2) 周辺情報の活用

疾患名に加えて，患者の属性情報として年齢・性別を用いる予測及び診療行為を用いる予測を行う．これらの情報を扱うため，前者は CVAE (Conditional VAE) [3]，後者は JMVAE (Joint Multimodal VAE) [4]を用いた．

2) データ

三重県地域保険（国民健康保険および後期高齢者医療広域連合）の入院外レセプトデータ（約 70 万人）を，患者ごとに 1 年単位で患った疾

患を集計したものを疾患履歴と見なした. 年齢は 5 歳を一区切りとして階層化し, 疾患は ICD10 の中分類による分類を用いる.

将来のリスクを予測するため, 時系列方向にデータを分け, 平成 25 年のデータで学習する. 各患者の平成 25 年の状態から, 未罹患の疾患のうち出力値の上位 20 疾患を予測とし, 平成 26 年のデータで予測された疾患を罹ったかを確認する.

3. 結果

評価指標は再現率（マクロ平均, マイクロ平均）, 及び予測順位も考慮した NDCG (Normalized Discounted Cumulative Gain)を用いた. その結果を Table.1 に示した. 患者数 N, 予測のうち患者 i が実際に罹った疾患数m_i, 患者 i が罹った全疾患数M_iとすると, 各再現率は以下とした.

$$\text{マクロ平均} = \frac{1}{N}\sum_{i=1}^{N}\frac{m_i}{M_i},\ \text{マイクロ平均} = \frac{\sum_{i=1}^{N} m_i}{\sum_{i=1}^{N} M_i}$$

また, 患者 i が予測順位 j の疾患に罹ったかどうかを$r_{i,j}$ (0: 罹らなかった, 1 : 罹った）とすると, NDCG は以下とした.

$$DCG_i = \sum_{j=1}^{20}\frac{r_{i,j}}{log_2(j+1)},\ NDCG = \frac{1}{N}\sum_{i=1}^{N}\frac{DCG_i}{DCG_{i,max}}$$

$DCG_{i,max}$: 予測の上位M_i番目までが患者 i が罹った全M_i個の疾患であるときの DCG, 即ちDCG_iの取り得る最大値

Table.1 予測精度

	疾患のみ	疾患＋年齢・性別	疾患＋診療行為
マクロ平均	0.331	0.343	0.327
マイクロ平均	0.280	0.290	0.297
NDCG	0.218	0.225	0.205

4. 考察

1) 年齢・性別の効果

年齢・性別という条件が与えられたと考えられ, 条件付き分布をモデル化する CVAE を用いた. 後期高齢者のみのデータで同様の実験を行ったところ, 疾患のみに比べて各指標は最大でも 1% 程度の変化であった. 一方, 全年代のデータ (Table.1)では, 再現率のマクロ平均が約 4%向上しており, 年齢が有効な情報だと推定できる.

2) 診療行為の効果

疾患と診療行為という異なる種類の情報の同時分布を単一の潜在変数によってモデル化するため, JMVAE を用いた.

この場合, 再現率のマイクロ平均では向上が見られたのに対して, その他の指標ではわずかながら低下する結果となった.

疾患ごとに見てみると, 疾患のみの場合よりも診療行為も合わせたほうが, 狭心症(I20), 心不全(I50)、脳梗塞(I63)等を含む 26 疾患で予測できた患者数は向上した. 前述以外の 23 疾患は A09,B35,E11,E14,E78,G47,H10,H16,H26,H52,I10,J02,J06,J20,J30,K21,K25,K29,K76,L30,M54,M81,Z96 であり, 紙面の都合上, ICD10 コードのみの記載とした. これらの疾患については診療行為が有効な情報だと言えるが, 何れも患者数が多い疾患のため, これらに過剰に適合して, 他の疾患では精度が低下した可能性も考えられる.

今後, 診療行為を精度向上に繋げるモデル化と調剤レセプトの活用を検討する予定である.

5. 結語

本研究ではレセプト上の情報から, 患者ごとの将来の罹患リスクの評価手法を試行した. ここで評価しているのはレセプト上に疾患名が付く確率であり, 実際の病態を表現するとは限らないが, 予防に一貫としての情報提供としては有効ではないかと考えられる.

参 考 文 献

[1] D. P Kingma, M. Welling: Auto-Encoding Variational Bayes, 2013

[2] D. Liang, R. G. Krishnan, M. D. Hoffman, et al: Variational Autoencoders for Collaborative Filtering, WWW '18 Proceedings of the 2018 World Wide Web Conference, 689-698, 2018

[3]S. Kihyuk, L. Honglak, Y. Xinchen: Learning Structured Output Representation using Deep Conditional Generative Models, Advances in Neural Information Processing Systems 28, 3483-3491, 2015

[4] M. Suzuki, K. Nakayama, Y. Matsuo, Joint Multimodal Learning with Deep Generative Models, ICLR 2017 workshop, 2017.