

ケアプラン作成支援システムのための トピックモデルに基づく特徴空間の対応付け

Mapping feature space based on the topic model for systems to support the creation of care plans

兵頭 幸起[†] 伊藤 豪[‡] 石原 憲之[‡] 濱上 知樹[†]
寺尾 勇一[§] 林 慧子[§] 佐野 貴洋[§] 竹林 奈々子[§]
Koki HYODO Go ITO Noriyuki ISHIHARA Tomoki HAMAGAMI
Yuichi TERAOKA Chieko HAYASHI Takahiro SANO Nanako TAKEBAYASHI

1. はじめに

日本は超高齢化社会であり、労働集約性の高い介護分野におけるケアマネジメントが重要視されている。現状ではケアマネージャーがアセスメントからケアプランを作成しているが、多くの時間と労力が必要であり主観や経験に依存している。

本研究では過去の介護記録を基に適したケアプランを推薦する、ケアプラン作成支援システムを実現する。代表的な推薦手法として協調フィルタリング (Collaborative Filtering)[1] が挙げられるが、構造が複雑なケアプランへの適用は困難である。そこで、文書構造に基づいてケアプランを分割し問題の簡略化を図る。ケアプランに対してトピックモデルにより複数の分類基準を作成し、アセスメントを異なる視点で評価する。評価を加味することで協調フィルタリングの改良が達成され文書構造に基づく提案システムの有効性が明らかとなった。

2. ケアマネジメントデータ

2.1 アセスメント

アセスメントはヒアリングにより得られる利用者情報であり、選択項目と記述項目から構成されている。選択項目は客観的情報を表し、記述項目は利用者特有の情報を表す。

2.2 ケアプラン

ケアプランはアセスメントを基に作成される介護計画書であり、ニーズ、長期目標、短期目標、サービス内容により階層構造を成している。本研究では 1 組のニーズ、長期目標、短期目標、サービス内容のセットをコンテンツとして定義する。

3. 協調フィルタリング

代表的な推薦手法。利用者間の類似度から評価値を予測する。類似度として式 (1) のピアソン積率相関係数を用いる。

$$sim(\mathbf{q}, \mathbf{d}) = pear_{\mathbf{q}, \mathbf{d}} = \frac{(\mathbf{q} - \bar{\mathbf{q}}) \cdot (\mathbf{d} - \bar{\mathbf{d}})}{|\mathbf{q} - \bar{\mathbf{q}}| \cdot |\mathbf{d} - \bar{\mathbf{d}}|} \quad (1)$$

\mathbf{q} は活動利用者ベクトル、 \mathbf{d} は標本利用者ベクトルを表す。

評価値予測には式 (2) の単純重み付き和を用いる。

$$p_{a,i} = \frac{\sum_{u \in U} r_{u,i} s_{a,u}}{\sum_{u \in U} |s_{a,u}|} \quad (2)$$

U は学習データ数、 $r_{u,i}$ は標本利用者 u のアイテム i の評価値、 $s_{a,u}$ は利用者 a , u 間の類似度を表す。

本研究で扱うアセスメントの記述項目とケアプランは自然言語である。文書ベクトルに対して適用することで文書構造を大きく崩すことなく予測を行うことが可能になると考えら

表 1 分類詳細

ケアプラン	分類基準作成方法	空間表現方法
上位層	LDA	トピック該当確率
下位層	LDA	トピック該当確率
大域レコメンデーション	k-means++	IDF

れる。その一方で次のような課題も存在する。

- 入出力間に非線形な対応関係が存在する場合に適用困難
- 出力結果の解釈が困難

4. 提案システムの概要

本研究ではケアプラン構造に基づくレコメンデーションを提案する。提案システムは以下の要素から構成される。

- 上位層ケアプラン空間
 - Latent Dirichlet Allocation(LDA)[2] によるコンテンツのトピック抽出
 - ランダムフォレスト (Random Forest)[3] によるマルチラベル分類、類似度比較
- 下位層ケアプラン空間
 - LDA によるケアプランのトピック抽出
 - ランダムフォレストによるラベル分類、類似度比較
- ケアプラン空間
 - 大域レコメンデーション
 - * k-means++ 法 [5] によるクラスタリング
 - * ランダムフォレストによるラベル分類
 - * クラスタ重心間類似度比較
 - 局所レコメンデーション
 - * 協調フィルタリングによるケアプラン予測
 - * ケアプラン空間における類似度比較

入力アセスメントの選択項目と記述項目であり、出力はケアプランである。レコメンデーションの流れを図 1 に示す。

上位層、下位層、大域レコメンデーションにおける分類器としてランダムフォレストを採用することで解釈性の向上を図り、局所レコメンデーションでは分類結果を加味した協調フィルタリングによるケアプラン予測を行う。

分類の詳細を表 1 に示す。LDA はトピックモデルの一種であり、文書内に複数の潜在トピックの存在を仮定してトピックの該当確率によりベクトル化を行う。また、k-means++ 法は、k-means 法において初期のセントロイドを互いに離れた位置に配置するクラスタリング方法である。

ケアプラン空間においてケアプランは IDF (Inverse Document Frequency)[4] によってベクトル化を行う。IDF は文書群全体における単語の重要度を表す。

局所レコメンデーションでは協調フィルタリングを用いる。アセスメント間類似度は式 (1) より算出し、そこに上位層ケアプラン空間 ($\mathbf{q}_{up}, \mathbf{d}_{up,u}$)、下位層ケアプラン空間 ($\mathbf{q}_{low}, \mathbf{d}_{low,u}$)、大域レコメンデーション ($\mathbf{q}_{glob}, \mathbf{d}_{glob,u}$) で

[†] 横浜国立大学大学院理工学部

[‡] 横浜国立大学未来情報通信医療社会基盤センター

[§] リゾートトラスト株式会社メディカル本部シニアライフ運営戦略部

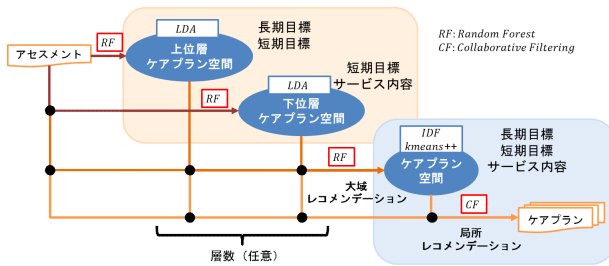


図1 提案システムにおける推薦の流れ

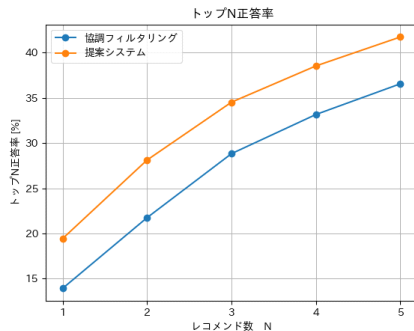


図2 トップN正答率 1プラン当たりの正解数: 5

の類似度を掛け合わせることで式 (4) のように利用者間類似度を表す。この時、ユークリッド距離から類似度への変換は式 (3) に従う。そして利用者間類似度を基に式 (5) に従い予測ケアプランの生成を行う。その後、ケアプラン空間において学習ケアプランと予測ケアプランの類似度をコサイン類似度により算出し、類似度の高い複数のケアプランを推薦する。

$$\text{sim}(\mathbf{q}, \mathbf{d}) = \text{euc}_{\mathbf{q}, \mathbf{d}} = \exp\{-\|\mathbf{q}, \mathbf{d}\|\} \quad (3)$$

$$\text{score}_u = \text{pear}_{\mathbf{q}_{\text{sel}}, \mathbf{d}_{\text{sel}}, u} \cdot \text{pear}_{\mathbf{q}_{\text{des}}, \mathbf{d}_{\text{des}}, u} \cdot \text{euc}_{\mathbf{q}_{\text{up}}, \mathbf{d}_{\text{up}}, u} \cdot \text{euc}_{\mathbf{q}_{\text{low}}, \mathbf{d}_{\text{low}}, u} \cdot \cos\theta_{\mathbf{q}_{\text{glob}}, \mathbf{d}_{\text{glob}}, u} \quad (4)$$

$$\mathbf{p} = \frac{\sum_{u \in U} \text{score}_u \cdot \mathbf{p}_u}{\sum_{u \in U} \text{score}_u} \quad (5)$$

ここで、 \mathbf{q}_{sel} と \mathbf{q}_{des} は活動利用者アセスメントの選択項目と記述項目、 $\mathbf{d}_{\text{sel}, u}$ と $\mathbf{d}_{\text{des}, u}$ は u 番目の標本利用者アセスメントの選択項目と記述項目、 \mathbf{p}_u は u 番目の標本利用者ケアプランベクトルを表す。

5. レコメンデーション実験

5.1 実験設定

実際のケアマネジメントデータ（対象ケース数：1333）を用いて実験を行った。評価方法として正解ケアプランと予測ケアプランのコサイン類似度に加えてトップN正答率を用いた。

5.2 実験結果

実験結果を図2と表2に示す。図2よりNの値によらず、提案システムのトップN正答率は協調フィルタリングを上回っていることが確認され提案システムの有効性が示された。表2より最小類似度を除き、提案システムの類似度が高い結果となった。最小類似度が協調フィルタリングを下回ってしまった原因を推察する。提案システムでは上位層、下位層での予測結果を用いて大域推薦と局所推薦と局所推薦を行っているため、各階層における予測が正確でなかった結果誤りを助長してしまったと考えられる。

表2 予測ケアプランと正解ケアプランの類似度比較

類似度	提案システム	協調フィルタリング
平均 ± 標準偏差	0.4814 ± 0.0969	0.4749 ± 0.0876
最大値	0.9533	0.9131
最小値	0.2028	0.2128

表3 ランダムフォレストを用いたアセスメントの重要属性

重要度	トピック1	トピック2	トピック3	トピック4	トピック5	トピック6
1	食事(重度) 入浴形態	排泄 座位	食事(軽度) 食事区分	体調、生活 薬剤管理 外用薬 自己管理	援助(声掛け) 視力	リハビリ 物理療法
2	副食形態	意思伝達	異食行動 がある	薬剤管理 外用薬 ホーム管理	幻視・幻聴 がある	上肢筋力 訓練
3	義歯装着	紙おむつ	髪剃り	洗髪	入浴形態	関節可動域 訓練
4	水分形態	意思伝達方法 会話	上着着脱	主食形態	不快な音 をたてる	電話利用
5	処置内容 疼痛	処置内容 カテーテル	寝返り	上肢筋力 訓練	服薬方法	社会活動 への参加

6. アセスメント重要属性

ランダムフォレストを用いた重要属性の確認を行った。ここでは解釈が容易であった下位層の分類について表3に示す。トピックに関連があると思われる属性（赤字）が重要であると判断されていることから、影響力の高い属性を提示することでケアマネージャーの支援に繋がると考えられる。

7. おわりに

提案システムではケアプラン構造に基づく推薦によってケアプラン作成支援システムを実現した。ランダムフォレストを用いることで解釈性の向上を図り、ケアプランを分割し部分的な予測を行うことで複雑なケアプラン構造に対応した。推薦実験を行い協調フィルタリングと比較した結果、提案システムの有効性が確認された。

今後の展望として単語間の共起関係を加味することでより柔軟な推薦が可能になると考えられる。

参考文献

- [1] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 175–186. ACM, (1994).
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, (2003).
- [3] Leo Breiman. Random forests. *Machine learning*, Vol. 45, No. 1, pp. 5–32, (2001).
- [4] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, Vol. 24, No. 5, pp. 513–523, (1988).
- [5] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027–1035. Society for Industrial and Applied Mathematics, (2007).