

テンソル分解による訪問地予測

POI Prediction by Tensor Factorization

中辻 真 戸田 浩之 小池 義昌
 Makoto Nakatsuji Hiroyuki Toda Yoshimasa Koike

NTT サービスエボリューション研究所
 NTT Service Evolution Laboratories

訪問地予測は、ユーザの不慣れな地域での POI(Point of Interests) 推薦における鍵となる技術である。テンソルは、訪問ログ上の複数のオブジェクトの関係 (例えば、ユーザ、POI、POI への訪問時間帯からなる関係など) を取扱う事が出来る。そのため、テンソル分解による訪問地予測が期待されている。しかし、現在のテンソル分解手法は、(1) ユーザが不慣れな地域には訪問ログが無い、(2) 訪問した POI 間の遷移関係を分解時に利用できない、という課題を持つ。本稿で提案する手法は、個別の訪問場所のみでなく、訪問場所の背景となるセマンティクスを用いる。結果として、ユーザが習熟した地域の訪問履歴のセマンティクスを用い、ユーザが不慣れな地域の訪問 POI を高精度に予測できる。また、提案法は、訪問ログ内にある、訪問 POI の遷移関係をテンソル分解時に参照させる。結果として、訪問 POI が属する地域の訪問に関する特性を反映する事ができる。

1. はじめに

スマートデバイスの普及や位置計測技術の多様化など、近年のユビキタス計算環境の普及に従い、ユーザにとって便利または興味のある場所として定義される Point of Interests (POI) を、ユーザの訪問ログから予測し推薦する技術が重要になってきている。商用的観点からは、例えば、位置ベースのソーシャル・ネットワークサービス (location-based social networks, LBSN) を提供する Foursquare^{*1} は、本技術を用い、ユーザに、周辺にある場所の情報を提示し、当該場所へのユーザの習熟度を向上させようとしている。研究的観点からは、Foursquare や Jiepan^{*2}、Facebook Place^{*3} 等の LBSN を通じて大規模な訪問ログが Web 上に出現するようになってきた事から、POI 予測や推薦に関わる研究が多く見られるようになってきている。例えば、ユーザと POI からなる行列を分解し、POI 予測を行う研究では、ユーザの移動行動に関する時間的属性という補助情報を行列分解時に追加することで、時刻属性を伴った予測を行う [Gao et al.2013, Yuan et al.2013]。ユーザの訪問ログは、ユーザ、訪問 POI、訪問時間帯といった複数オブジェクト関係により表現でき、また、そうした関係はテンソルとして表現 (例えば、ユーザ、訪問 POI、訪問時間帯からなるテンソル) できるため、テンソル分解による POI 予測は、潜在的な価値がある。情報推薦に関する研究 [Karatzoglou et al.2010] によると、テンソル分解は、行列分解に基づく行動予測よりも精度が高い事が指摘されている。また、ユーザは、高精度に予測された結果を信頼し利用する傾向があるため [McNee et al.2006]、高精度に予測を行う手法を開発することは、POI 推薦の技術としての有用である。

しかし、テンソル分解手法を POI 予測に適用するには、以下の 2 つの問題がある。(1) 一つ目は、ユーザが不慣れな場所において、ユーザの訪問ログが無いという問題である。ユーザは不慣れな地域を訪問した時にこそ、魅力的な POI を推薦されたいと感じるものであるにも関わらず、ログ不足により予測精度が劣化するため、満足度の高い推薦は期待できない。図 1-(a) は、この問題を示している。図には、ユーザ、訪問 POI、

時間帯 (日中と夜間) といった 3 つのオブジェクト間の関係が示されている。図において、ユーザ u_{m1} は、ボストンにおいて、“イザベラ・スチュワート・ガードナー美術館” という POI v_{n1} を“昼間”に訪問しており、“ネプチューン・オイスター” という POI v_{n2} を“夜間”に訪問をしている。一方、ユーザ u_{m2} は、マンハッタンにおいて、“フリッカーコレクション” という POI v_{n3} を“昼間”に訪問しており、“ウルフギャング・ステーキハウス” という POI v_{n4} を“夜間”に訪問をしている。ユーザ u_{m1} はマンハッタンに訪問ログを持たないため、現在の POI 推薦の手法では、 u_{m1} に対してマンハッタンにおいてボストンでの訪問と類似する POI を予測できない。結果として、予測精度が劣化する。(2) 二つ目の問題は、テンソル分解において、訪問 POI 間のユーザ全体での遷移の依存関係を組込む事ができない事である。これは、テンソル分解では、同一のオブジェクトタイプに属するオブジェクト間 (例えば、ユーザ間、POI 間、訪問時間帯間) では、オブジェクトは独立して存在するという仮定が置かれているためである。この、訪問 POI 間の遷移関係は、各地域の特性を反映していると考えられるため、不慣れな場所を訪問する旅行者にとっては、次の訪問先を決定する際に有用である。例えば、図 1-(b) は、マンハッタンを訪問した旅行者は、“フリッカーコレクション”で芸術を嗜んだ後、“ウルフギャング・ステーキハウス”で夕食を楽しむという傾向がある事を図示している。こうした訪問 POI の遷移関係を現在のテンソル分解手法は利用する事ができないため、予測精度が低く、推薦の満足度も下げてしまう。

2. 提案手法

本研究は、POI の背景となるタクソノミで表現されるセマンティクスと、POI 間の遷移関係をテンソル分解に入込み、上述した問題を解決する、新たな手法を開発する。提案手法は、以下の 2 つのアイデアに基づく。

(1) 提案手法は [Nakatsuji and Fujiwara2014] でアイデアの原型が提示され、[Nakatsuji et al.2014] でテンソルに応用されたセマンティクスペースのテンソル分解手法を拡張して、ユーザにとって不慣れな場所の訪問行動を正確に予測する。それは、テンソル分解において、タクソノミにより表現されるセマンティクスを用いる。提案手法は、まず、ユーザ、POI、時間帯からなるテンソル (元テンソル) に対し、ユーザ、POI

連絡先: 〒 239-0847 神奈川県横須賀市光の丘 1-1

*1 <https://ja.foursquare.com/>

*2 <http://jiepan.com>

*3 <https://www.facebook.com/places/>

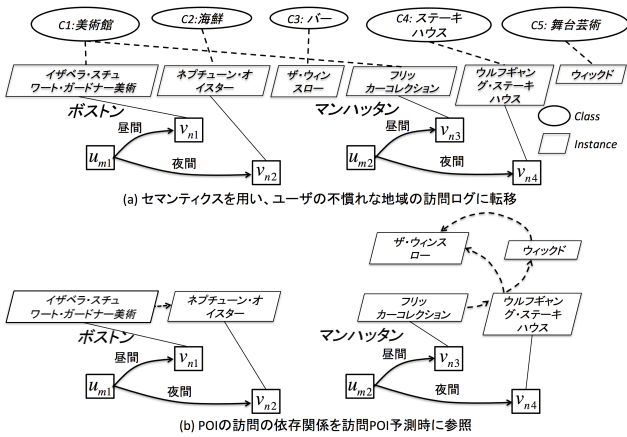


図 1: 提案手法のアイデア; 点線が提案手法のアイデアを示す。

の所属クラス、時間帯からなる複数オブジェクト関係を追加する事で、増大テンソルを作成する。次に、提案手法は、元テンソルと増大テンソルを同時に分解する。これにより、増大テンソルを通じ、ユーザの習熟した地域における訪問 POI の知識を、ユーザが不慣れな地域の知識として、転移する事ができるようになる。これは、ユーザの訪問の傾向は、ユーザが不慣れな地域と、ユーザが習熟した地域で同じ傾向があるという我々の知見に基づく。なお、[Nakatsuji et al.2014]とは異なり、提案手法は、セマンティクスを習熟した地域の訪問ログを不慣れな地域の訪問ログに転移する、という利用の仕方をしており、その点で分解技術が異なっている事を注意されたい。図 1-(a)では、提案手法は、ユーザ u_{m1} の訪問の傾向を意味的に把握している。それにより、ボストンにおける u_{m1} の訪問ログを用いて、マンハッタンにおける訪問を予測できる。例えば、提案手法は、 v_{n1} と v_{n3} が同じクラス“美術館”に属しているという知識を用いる事で、 u_{m1} , v_{n1} 、“daytime”からなる複数オブジェクト関係は、 u_{m2} , v_{n3} 、“daytime”と類似しているという事を理解する事ができる。

(2) 提案手法は、テンソル分解における訪問 POI 間の遷移関係も利用する。この遷移関係は、その訪問 POI が存在する地域の訪問の特性を暗黙的に反映している。そして、そうした特性は、ユーザが不慣れな地域に訪問した際にこそ、有益である。提案手法は、まず、テンソル分解の前に、全ユーザの全訪問ログから訪問 POI 間の遷移関係を計算する。次に、テンソル分解の学習フェーズにおいて、逐次的に、各 POI の特徴ベクトルを、その POI と遷移関係のある POI の特徴ベクトルを反映しながら、計算していく。結果として、本処理は、テンソル分解実行時に、POI 間の遷移関係に潜む地域特性を反映することができる。一方、現在のテンソルベースの手法 [Nakatsuji et al.2014, Xiong et al.2010] では、そうした遷移関係をテンソル分解時に考慮することが出来ない。図 1-(b)は、クラス“美術館”に属している“フリッカーコレクション”は、良く“ウルフギャング・ステーキハウス”の前に訪問されている事を示している。また、クラス“舞台芸術”に属する POI “ウィックド”は、良く“ウルフギャング・ステーキハウス”の後に訪問され、また、クラス“バー”に属する POI “ザ・ウィンスロー”の前に訪問されているとする。提案手法は、POI 間の遷移関係を取扱う事ができる。それにより、“フリッカーコレクション”の属性ベクトルを、それと依存関係のある POI (例えば、“ウルフギャング・ステーキハウス”) の属性ベクトルからバイアスを受けつつ、依存関係のない POI (例えば、“ザ・ウィンスロー”や“ネブチューン・オイスター”) の属性ベ

クトルからは独立して、計算することが可能となる。

上記 2つのアイデアを組合せる事により、提案手法は、ユーザが訪問した事のない地域であっても、ユーザにとって潜在的に興味があり、かつ、地域の特性を反映した POI を推薦する事ができる。例えば、ユーザ u_{m1} は“イザベラ・スチュワート・ガードナー美術館”に過去に行っており、“イザベラ・スチュワート・ガードナー美術館”は、クラス“美術館”に属しているため、同じクラス“美術館”に属している“フリッカーコレクション”は彼の興味に沿う可能性があり推薦するに有益である。また、マンハッタンの旅行の特徴にもとづき、彼は“フリッカーコレクション”で芸術を楽しんだ後、“ウルフギャング・ステーキハウス”にち寄る予定を考える事もできる。

我々は、本手法を Bayesian Probabilistic Tensor Factorization (BPTF) [Xiong et al.2010] フレームワーク上に実装を行った。BPTF は、ページアンに基づかない手法よりも高精度であるとされる。また、大規模データセットに対しても高速に適用でき、パラメータ設定も容易である [Xiong et al.2010]。

我々はまた、ユーザの将来の訪問は、行く・行かないのバイナリ値で表現されるため、BPTF フレームワークにロジスティック回帰 [Bishop2006] を適用した。これにより、提案手法は、テンソル分解時にベルヌイ分布を考慮し予測を計算できるようになる。

3. 結論

本提案は、ユーザが不慣れな地域を訪問した際に、魅力的な訪問候補となる POI を、時間帯などの細やかな特性も考慮しながら予測し推薦する事を可能とする。大きな貢献としては、(1) ユーザが不慣れな場所であっても、セマンティクスを用い、ユーザが習熟した地域での訪問ログを転移する事、(2) 訪問地域での POI の訪問の遷移関係を考慮する事、という 2 点をテンソル分解のフレームワークに入れ込んだ事である。今後、実データを用い、実験をし、結果を示していきたい。

参考文献

[Bishop2006] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.

[Gao et al.2013] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. Exploring temporal effects for location recommendation on location-based social networks. In *Proc. RecSys'13*, pages 93–100, 2013.

[Karatzoglou et al.2010] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proc. RecSys'10*, pages 79–86, 2010.

[McNee et al.2006] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *Proc. CHI'06*, pages 1097–1101, 2006.

[Nakatsuji and Fujiwara2014] Makoto Nakatsuji and Yasuhiro Fujiwara. Linked taxonomies to capture users' subjective assessments of items to facilitate accurate collaborative filtering. *Artif. Intell.*, 207:52–68, 2014.

[Nakatsuji et al.2014] Makoto Nakatsuji, Yasuhiro Fujiwara, Hiroyuki Toda, Hiroshi Sawada, Jin Zheng, and James A. Hendler. Semantic data representation for improving tensor factorization. In *Proc. AAAI'14*, 2014.

[Xiong et al.2010] Liang Xiong, Xi Chen, Tzu-Kuo Huang, Jeff G. Schneider, and Jaime G. Carbonell. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *Proc. SDM'10*, pages 211–222, 2010.

[Yuan et al.2013] Quan Yuan, Gao Cong, Zongyang Ma, Aixun Sun, and Nadia Magnenat Thalmann. Time-aware point-of-interest recommendation. In *Proc SIGIR'13*, pages 363–372, 2013.