

モデルのライフサイクルを考慮した大量予測モデル管理手法の検討

A Study on Massive Predictive Model Maintenance based on Their Lifecycle

谷本 啓 本橋 洋介

Akira Tanimoto Yosuke Motohashi

NEC 情報・ナレッジ研究所

Knowledge Discovery Research Laboratories, NEC Corporation

In the real operation of batch based predictive analysis, the decay of predictive models caused by the change of phenomena over time is one of big problems. We studied on predictive model maintenance methods to keep accuracy efficiently by checking model accuracies and selecting the timing of re-learning.

1. 背景

近年、機械学習による予測システムが実社会で広く利用され始めている。例えば、電力需要量予測に基づく蓄電最適化、小売店での売り上げ予測に基づく発注最適化、中古商品の価格予測に基づく値付け最適化などである。

これらの長期にわたって運用される予測システムでは、消費行動の変化などの影響で傾向が変化することによる精度低下や、データ量が増えることによる精度向上の可能性があるため、予測モデルの更新が必要となる。我々は実際に、適用対象によって予測精度の劣化がどの程度の速度で進行するか確かめ、定期的にモデルを更新するシステムを構築してきた。

しかしながら、こうした定期更新には問題があることがわかってきた。予測モデルは電力ならばビルごと、小売りならば店舗×商品ごとなど、大量の予測器を運用する場合が多い。このとき、それぞれの予測対象ごとに予測劣化傾向は異なるが、一律に定期更新を行うと、更新不要なモデルまで更新してしまうことにより、予測モデルの変化を確認するなどの人的コストが膨大になってしまうという問題である。

この問題を解決するため、我々はモデルのライフサイクル(図1)を考慮した管理手法を検討し、本稿にてこの問題の整理と簡単な実証実験を行った結果を述べる。

2. 関連研究

本研究の他にモデルライフサイクル管理に取り組んでいる事例と我々の立場について述べる。SAS社のSAS Model Manager[SAS]では、各モデルの予測精度の劣化状況を可視化することで運用者に対してモデルの更新を促す機能を提供している。またCrankshawらは運用の完全自動化を目指し、学習の分散実行や自動更新フローをサポートするソフトウェア基盤の開発について発表している。[Crankshaw15] Crankshawらのシステムでは、図1と同様のモデルライフサイクルを仮定し精度劣化状況の監視に基づく更新を意図している点では類似しているが、予測の目的として推薦システムを念頭に置いていることから、予測モデル自体の人間による解釈を前提としない。

一方、我々は人間による最終決定を支援する大量予測器の効率的運用を目指す立場から、自動化された更新システムであって、その更新ルールは人間にとって解釈可能であり、更新頻度

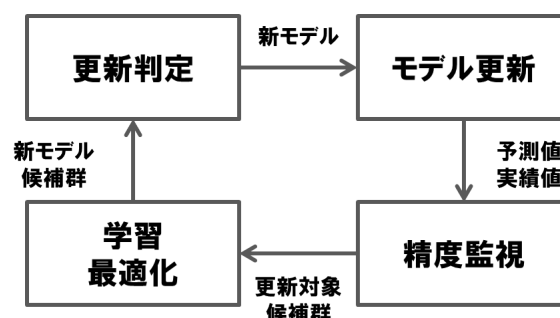


図1: 予測モデルのライフサイクル

の低さと精度を両立することが重要と考えている。このような利用目的に沿った問題設定について次節で述べる。

3. 問題設定

我々の対象とするシステムは社会インフラとして人間と協調し人間による意志決定を支援するコンポーネントとしての予測システムである。このような領域では、最終的な意志決定を人が行う必要があるため、多くの場合予測モデルの解釈性が求められ、時々刻々とモデルが変化する逐次学習のような方法はなじまない。むしろ、モデルを運用するにつれて人間が予測モデルの傾向を学習し、それに最適化した運用を始めるなど、予測モデルと人間の感覚の密な連携によって意志決定が行われる。モデルの更新は即ち人間による内容確認や予測モデルの傾向の再学習を意味するため、非常に高コストであり、低更新頻度で予測精度が維持されることが望ましい。

さらに、予測対象が大量である場合、学習にも時間とマシンリソースを要するため、学習回数も少ない方が望ましい。ここで、図1に示したように学習した後に新モデルの性能を評価し更新すべきかどうか判断することができるため、学習と更新は必ずしもセットである必要はない。一般的には、学習自体は低コストであるため、比較的多くの予測対象、及び1つの予測対象に対して複数の学習方法を試行し、更新判定部で学習した中の一部のモデルのみ更新することで、更新に伴う人的コストを最小限に抑えつつ精度を維持することを目指す。

さらに、一度の学習であれば人力によって最適な前処理方法や学習方法を選定することができるが、これらの最適性も経時変化する可能性があるため、再学習時に最適化されることが望

ましい。
よって、主要な課題は以下の3点である。

1. 精度の維持, 改善
2. マシン及び人的リソースの抑制
(学習回数及びモデル更新回数の抑制)
3. 学習パラメタ, 学習データの選択による精度の最適化

とくに更新頻度が低く制限される場合, 1と2はトレードオフの関係にある。即ち, 更新頻度が十分高くないとき, 更新頻度が下がると平均精度は低下する。

上記課題の1と2について第5.4節で, また3について第5.5節でそれぞれ検討する。

4. システム構成と処理フロー

第3.章で述べた問題を解決する具体的な更新ルールは, 対象とするデータの性質及び予測の利用目的によって異なる。このような予測案件では実運用の前に価値実証フェーズがあり, その際に過去のデータからモデルの更新ルールも最適化できるものと考えられる。価値実証フェーズにおける更新ルールのシミュレーションのフローを示したのが図2である。

図2のように, 予測器の性能を逐次評価し, 予め定められた更新ルールをもとに再学習起動判定, 再学習, 更新判定を行い, 全期間運用した結果として平均品質指標を得る。平均品質指標は問題によって様々であるが, 本稿では誤差率の平均値と更新回数の平均値とする。

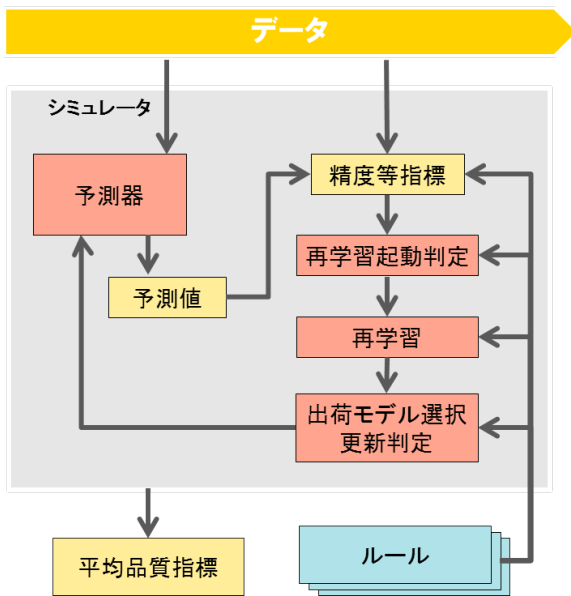


図2: モデル更新則のシミュレーション及び評価フロー

5. 実験

5.1 使用するデータ

今回の実験では, POS データ*1を用いたビールの売り上げ数の予測問題について実験を行った。ビールの売り上げは新商品のリリースや気候などの影響で傾向変化が著しい。売り上

げ数推移の一例を図3に示す。用いたデータの詳細は表1に示す。

表1: データ概要

予測対象	小売り店商品売り上げ数量
予測対象数	80 (10 店舗*8 商品)
期間	3 年間 (2012/1~2014/12)

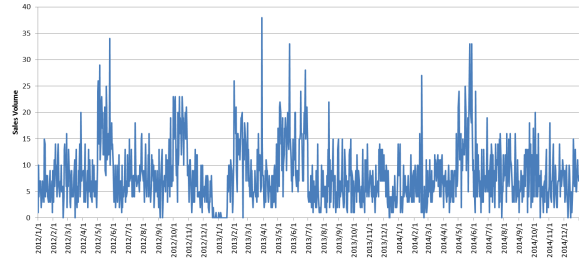


図3: 関東 A 店におけるビール X の売り上げ数推移

5.2 学習器

このデータに対し, 学習器としては解釈性と精度の高さから区分スパース線形モデル [Eto 14]を用いた。今回のデータに適用した場合の予測モデルの一例を図4に示す。

5.3 精度指標

精度指標としては, 回帰問題であれば二乗誤差や絶対誤差など問題設定に合わせて様々な指標をとりうるが, 今回は式1に示される平均誤差率を用いた。すなわち, 更新を行うかどうかの判断を1ヶ月に1回行うものとして, 最新1ヶ月間をバリデーション区間にとり, 平均誤差率を算出している。

$$AverageErrorRate := \frac{\sum_{i \in validation} |\hat{y}_i - y_i|}{\sum_{i \in validation} |y_i|} \quad (1)$$

ただし, y は実測値を, \hat{y} は予測値をそれぞれ表す。

このとき, 予測器を更新しない場合の精度推移の一例を図5に示す。図5aでは一時的な変動もありつつ, 全体的に誤差率は上昇傾向にある。このような精度劣化の傾向を捉えてモデルの再学習及び更新を行うのが望ましい。なお, 今回の予測は店舗ごと, 商品ごとに行っているが, 例えば商品の分類種ごとなど, 予測範囲を変えることも考えられる。より広範囲の予測対象とした場合, ノイズに対してバイアスが相対的に高まり, より変化傾向が精度変化として顕著に表れやすいためと考えられる。

図5bの例では, 一時的な要因で急激に誤差率が悪化したその後回復している。このような一時的な変化を捉えたと, 無駄に更新してしまうばかりでなく, 一時的な傾向を捉えた予測器を生成してしまい誤差率の悪化にも繋がるため, 監視する精度指標の設計は重要である。

5.4 実験1: 学習及び更新対象の選定

図5のような誤差率の系列が観測されたとき, 一般的には誤差率が悪化したら予測モデルの有効期限切れと考え更新するべきであろう。今回は更新直後からの誤差率の差分 [%] を監視して再学習を起動するルールを設定した。これに対し, 全てのモデルを定期的に再学習するルールを比較対象とした。

また, 更新判定ルールには, 再学習を行ったあと必ず更新するルールと, 誤差率がバリデーションデータで改善していたら更新するルールを用意した。

*1 データ出典元: KSP-SP 社 [KSP-SP]

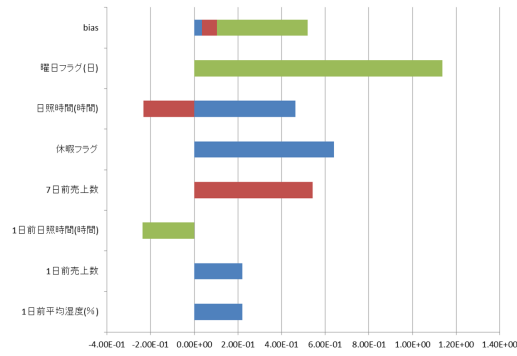
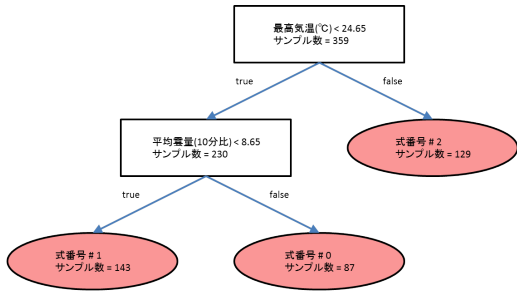


図4: 予測モデルの一例。データを場合分けする門関数(上)とスパースな線形モデル(下)からなる。例えば上記モデルの式番号0は、最高気温が24.65℃より低く平均雲量が8.65より高いときに下側の予測式0で示される回帰係数をもつ線形予測器で予測されることを示している。線形予測器の係数値は目的変数及び説明変数を標準化した状態の値を示している。

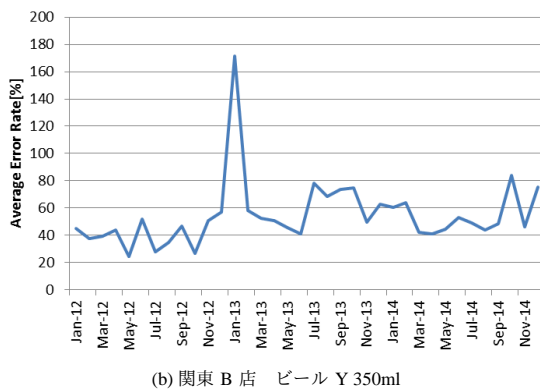
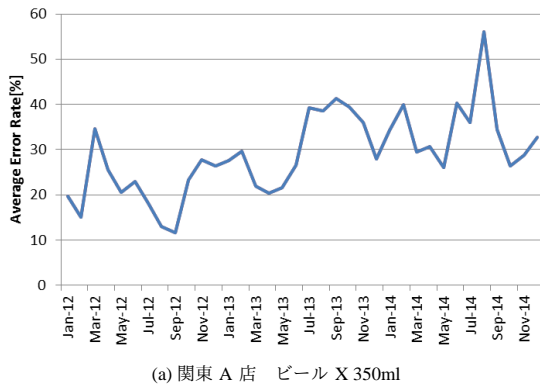


図5: 再学習しない場合の精度推移の一例

表2: 試行するルール

再学習対象選択	定期: 予め決めたタイミングで再学習(間隔は複数パターン) 誤差率監視: 誤差率の変化度が閾値を超えたら再学習(閾値は複数パターン)
更新対象選択	全更新: 再学習したモデル全て更新 改善更新: バリデーション誤差が改善したら更新
学習区間	前月までの6ヶ月間
バリデーション区間	最新1ヶ月間(学習区間との重複なし)

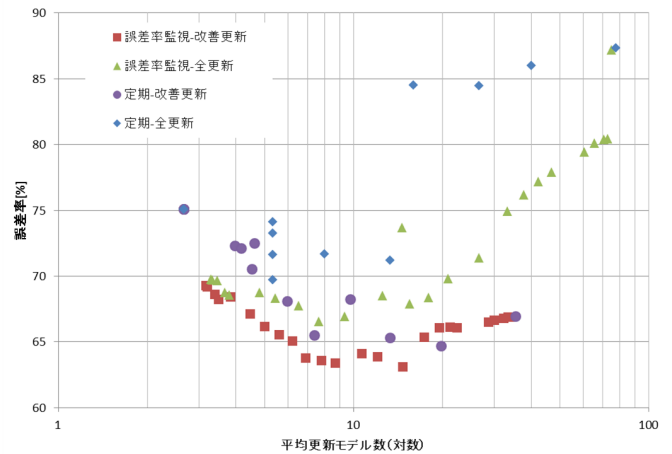


図6: 各更新ルールを閾値を変えてシミュレーションしたときの平均更新モデル数(平均更新頻度)と平均誤差率。平均更新モデル数は1度の更新判断において更新されたモデル数の全期間における平均値であり、予測対象数である80が最大である。

それぞれのルールについて様々な閾値でシミュレーションを行ったとき、全予測対象・全区間での平均誤差率及び平均更新モデル数(最大80)を平均品質指標として評価した結果が図6である。

この結果から、とくに低更新頻度では誤差率と更新頻度がトレードオフの関係にあること、また定期再学習よりも誤差率の悪化度合いが大きい予測対象から再学習を行った方がよいことが分かる。

さらに、更新頻度が高いと精度が悪化することが示されている。これは、販促キャンペーンなどの影響で一時的に増加した誤差率から判断して再学習及び更新した結果、一時的な傾向を捉えたモデルを生成し、その後のテストデータに対する予測性能が悪くなっていると考えられる。

このような不安定性に対して更新判定でバリデーション誤差が改善したかどうかを検証するルールが一定程度有効であることが分かる。ただし、今回の更新判定基準はバリデーション誤差が0%以上改善しているかどうかのみであるが、更新判定を行ってもなお高更新頻度で多少の悪化が見られることから、最適な更新頻度となるよう更新判定の閾値に関してもパラメタ探索による改善の可能性があると考えられる。

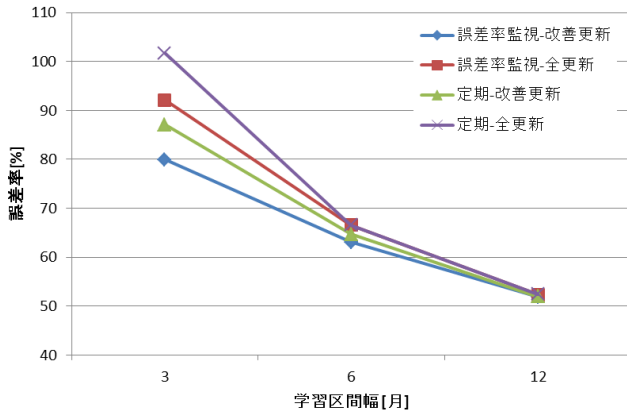


図 7: 学習区間の幅と誤差率の関係

5.5 実験 2: 学習方法の最適化

次に、再学習に用いるデータ量についての実験を行った。実験 1 の学習区間を 3ヶ月、6ヶ月、12ヶ月のそれぞれに変更して試行し、最も平均誤差率の低い閾値を採用したときの平均誤差率を学習区間幅ごとにプロットしたのが図 7 である。

学習区間が長い方が精度劣化のスピードが遅いと考えられることから、更新頻度制限が厳しいときにデータ量を大きくとるのが有利であると考えられるが、今回の結果では更新頻度制限がない場合でも学習区間は長い方がよいことを示している。

モデルの表現力が十分高ければ、学習データが多いほど精度が高まる傾向がある。しかし、例えば 48ヶ月などにデータ点数を増やした場合、精度は飽和すると考えられる。さらに、例えば電力消費量など季節性や一時的傾向変化の影響を強く受ける現象を予測する場合などは、データ量を削減しても更新時の傾向に近いデータを適切に選択することで精度が向上する可能性がある。

また、今回は学習に使用した CPU 時間を評価していないが、マシンリソースが限られている、ないしデータ量が多く学習コストが高い場合、精度と学習回数及び使用データ量との間のトレードオフを考慮して最適なデータ量や再学習試行パターン数などを調整する必要がある可能性がある。

6. 結論

本稿では、大量予測モデルを運用する際に新たに発生する課題を述べ、その解決方針を示し、簡単な実証実験を行った。

まず、更新頻度と精度の間にトレードオフ関係があることを確認し、更新頻度が制限された下では精度劣化状況を監視することで定期更新よりも高精度を達成できることを示した。

また、学習後にバリデーション誤差を確認して更新可否を判断することで精度が高まることを示した。このことから、学習を行うかどうかの判断とは別途、実際に更新を行うかどうかの判断が必要であると考えられる。

学習の最適化に関しては、今回の検討ではデータ量が多いほど精度が高いという結果が得られた。ただし、データの性質によってはデータ量を減らしてより直近のデータを重視することにより精度が高まる場合もあると考えられる。例えば、季節性の影響が強い場合、同じ季節のデータを用いた方が精度が高まる場合がある。さらに、データ量を減らして特定時期に特化すると精度劣化が速まるなど、更新頻度も加味して最適化を行

う必要があり、更新頻度上限が明に与えられない場合、多目的最適化問題となる。

今回の検討では、再学習起動判断/学習方法/更新判断のルールの候補を予め与えた上で検討を行った。しかし、学習データやパラメタ、再学習起動や更新の条件をより複雑なものへと拡張する場合、精度や更新頻度など運用価値指標に対するルールの多目的最適化問題として定式化することが望ましい。複雑な更新ルールの表現方法とその最適化は今後の課題として挙げられる。

また、今回の実験は個別の商品に対して予測を行ったが、データの SN 比の問題から、より広い予測対象の予測値を利用する方が結果的に予測性能が高まる場合も考えられる。例えば、小売りのような予測対象となる商品の入れ替えが頻繁に発生する場合、新商品については学習データが不足することから、分類種全体の予測モデルによる代替や、転移学習などの方法が考えられる。これらをデータ量などに応じて適切に最適化することも今後の重要な課題である。

参考文献

[Crankshaw 15] Dan Crankshaw, Peter Bailis, Joseph E. Gonzalez, Haoyuan Li, Zhao Zhang, Michael J. Franklin, Ali Ghodsi, and Michael I. Jordan. "The Missing Piece in Complex Analytics: Low Latency, Scalable Model Management and Serving with Velox." In CIDR, 2015.

[Eto 14] Eto, Riki, Ryohei Fujimaki, Satoshi Morinaga and Hiroshi Tamano. "Fully-Automatic Bayesian Piecewise Sparse Linear Models." Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. 2014.

[KSP-SP] <http://www.ksp-sp.com/> (2015 年 3 月現在)

[SAS] SAS Model Manager:
http://www.sas.com/en_us/software/analytics/manager.html
(2015 年 3 月現在)