2E-5

多属性データの照合を短時間で実現する差分探索方式の提案と評価

山崎　健太郎† 小林 佑嗣† 喜田　弘司†
日本電気株式会社　クラウドシステム研究所†

1. はじめに

近年，ビッグデータを活用したリアルタイムのレコメンドサービスが注目を集めている。例えば，スマートフォンで収集される位置情報などから，適切な情報推薦を行うレコメンドサービス等がある．

レコメンドサービスでは，処理のリアルタイム性と推薦精度が重要となる．推薦精度を高めるためには，多くの状況情報を用いて推薦対象の絞り込みを行う必要がある．例えば，位置だけで情報推薦を行うよりも，購入履歴を加えて推薦対象を絞り込んだ方が，精度が向上する．

このようなビッグデータをリアルタイムに処理する技術として複合イベント処理技術“CEP（CEP：Complex Event Processing）”があるが，上記のようにサービスの推薦精度を高めるためには CEP においても，多数の状況情報を扱う必要がある．

本稿では，CEP にて多数の状況情報をリアルタイムに処理する方式について述べる．

2. 従来研究と課題

CEP を活用したレコメンドシステムは，登録処理と照合処理で構成される．登録処理は，情報推薦の条件をリアルタイムに探索可能なデータ構造で記憶する．条件は，属性と呼ばれる対象の情報（位置，動作等）の組で記述される．照合処理は，推薦対象の属性が変化した際に行われる．推薦対象の属性と記憶する条件を照合し，全属性が一致する条件を抽出し，対応する情報推薦を行う．

レコメンドシステムの条件には任意値や範囲値が含まれる場合があるため，条件を扱うデータ構造も任意値，範囲値へ対応する必要がある．任意値は*で表現される，いかなる値にも合致する値である．範囲値は 1～10 といった数値の区間を示す値である．任意値，範囲値へ対応するデータ構造の従来研究として，トライ木を基にした方式[1]や区分木を基にした方式[2]が挙げられる．しかし，このいずれの方式も，多数の属性を扱う場合に，照合を短時間で行うことができないため，レコメンドシステムの条件を記憶するデータ構造として利用することができない．上記方式の計算時間はそれぞれ，線形時間，指数時間となる．また，拡張トライ木方式[3]では，計算時間を線形時間以下とすることを実現しているが，効果は特定の属性のみが変化する場合に限定される．

Proposition and Evaluation of difference search for Numerous Attributes
†Kentaro Yamasaki, Yuji Kobayashi, Koji Kida
†Cloud System Laboratories, NEC Corporation

そこで本稿では，任意値，範囲値へ対応するデータ構造にて，属性が多様に変化する場合にも，多数の属性の照合を短時間で実現することを課題とする．

3. 条件間距離を利用した差分探索方式

本稿で提案する差分探索方式では，前回照合時の結果である条件のデータ構造内での記憶位置と，変化した属性の前回値と現在値から照合処理を行う．多くの場合，変化した属性の数は照合毎に数個程度である．これは推薦対象の状況変化に対してリアルタイムに情報推薦するために，短い間隔で推薦対象の属性がシステムに入力されていることに起因する．

以下，提案方式の登録処理手順と照合処理手順を説明する．登録処理は以下，A～D の手順で行う．

A) 条件間距離算出表の更新

提案方式では，データ構造から高速に条件を探索するため，条件間の距離を計算可能にする条件間距離算出表を管理する(図 1 参照).

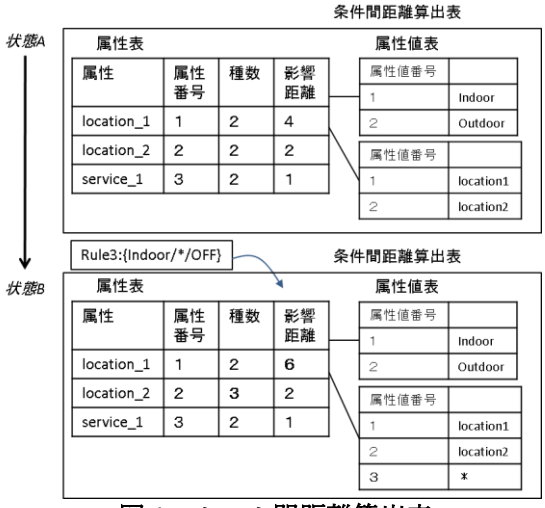

図 1　ルール間距離算出表

入力された条件に新出の属性値がある場合，条件間距離算出表を更新する．更新では，対応する属性値表に属性値を追加し，属性表の種数と影響距離を更新する．属性 i の影響距離 Di は $Di = \begin{cases} \prod_{j=i-1}^{N-1} Vj (0 < i < N-1) \\ 1 (i = N-1) \end{cases}$ で求める．Vjは属性j の種数である．図 1 の状態 A に示す条件間距離算出表に Rule3 を登録する場合，location2 の属性値表に*が追加され，属性表の location2 の種数が 2，location1 の影響距離が 6 となる(図 1 状態 B 参照).

B) データ構造の更新

提案方式では，任意値リンクを付与したデータ構造で条件の記憶を行う(図 2 参照)．任意値リンクとは，同意となる条件同士の紐付情報である．

初めに，新出の属性値と既存の属性値の組み合わせでできる条件(空条件)を作成する．この空条件はべき集合により容易に求めることができる．次に，空条件の記憶位置を算出し，エントリの挿入を行う．空条件の記憶位置 P は，$P = \sum_{i=0}^{N-1}(Ii - Di)$で求める．Iiは属性iの属性値番号，Diは属性iの影響距離である．最後に，入力された条件の記憶位置を同様に求めて記憶する．図 2 の状態 A に示す構造データの更新を行う場合，まず新出の属性値*から空条件{Indoor/*/ON}{Indoor/*/OFF}{Outdoor/*/ON}{Outdoor/*/OFF}を作成する．次に，この空条件の記憶位置 4，5，10，11 を求め，エントリを挿入する．最後に，Rule3{Indoor/*/OFF}の記憶位置 5 を求め，記憶する(図 2 状態 B 参照)．

状態A → 状態B

データ構造

記憶位置	記憶位置に対応する条件	登録条件	任意値リンク
0	Indoor/location1/ON	rule1	
1	Indoor/location1/OFF		
2	Indoor/location2/ON		
3	Indoor/location2/OFF		
4	Outdoor/location1/ON		
5	Outdoor/location1/OFF		
6	Outdoor/location2/ON		
7	Outdoor/location2/OFF	rule2	

データ構造

記憶位置	記憶位置に対応する条件	登録条件	任意値リンク
0	Indoor/location1/ON	rule1	/4
1	Indoor/location1/OFF		5/
2	Indoor/location2/ON		/5
3	Indoor/location2/OFF		5/
4	**Indoor/*/ON**		
5	**Indoor/*/OFF**	**rule3**	
6	Outdoor/location1/ON		/10
7	Outdoor/location1/OFF		/11
8	Outdoor/location2/ON		/10
9	Outdoor/location2/OFF	rule2	/11
10	**Outdoor/*/ON**		
11	**Outdoor/*/OFF**		

図 2　提案手法のデータ構造

C)　任意値リンクの更新

新出の属性値が*であった場合，記憶した空条件と同意となる条件の記憶位置を同エントリの任意値リンクに記憶する．同意となる条件の記憶位置の集合Lは$L = \{P - Di \times n\}:(1 \leq n \leq Vi-1)$で求める．Pはルールの記憶位置，Diは*である属性iの影響距離，Viは属性iの種数である．図 2 の状態 A に示す構造データの任意値リンクを更新する場合，まず{Indoor/*/ON}の同意となる条件の記憶位置の集合{2，0}を求め，構造データの 2，0 のエントリの任意値リンクに 4 を追加する．以降，同様に他の空条件についても処理を行う(図 2 状態 B 参照)．

D)　A～C を新出の属性値がなくなるまで実施する．

次に，照合処理の手順について説明を行う．初めに，前回の探索結果である条件の記憶位置を取得する．次に，変化した属性の前回値と現在値から，今回の照合結果となるべき条件と前回の探索結果との距離を，条件間距離算出表を用いて算出する．この距離 d は $d = \sum\big((Vn[i] - Vp[i]) \times Di\big):\{i \in$ 変化属性の属性番号$\}$で求める．Vn[i]，Vp[i]はそれぞれ，変化前，変化後の属性値番号，Diは影響距離である．そして，今回の照合結果となるべき条件の記憶位置を，前回の探索結果と算出した距離の和により求める．最後に，同エントリの任意値リンクを辿り，任意値で一致する条件を抽出する．

以上の手順で処理を行うことにより，提案方式では，前回結果と変化した属性の前回値と現在値のみで照合を行う．そのため，照合処理に要する時間が属性数に影響されず，多数の属性を扱う場合においてもデータの照合を短時間で行うことが可能となる．

4.　評価

提案方式による照合処理が，属性数に依存せず，短時間で行えることを検証する．

提案方式と従来方式について，属性数を変化させたときの 10,000 回の照合時間を計測した．計測条件は，照合毎の変化属性数を 10，各属性の種数を 20，条件数を 1000 とした．図 3 に結果を示す．

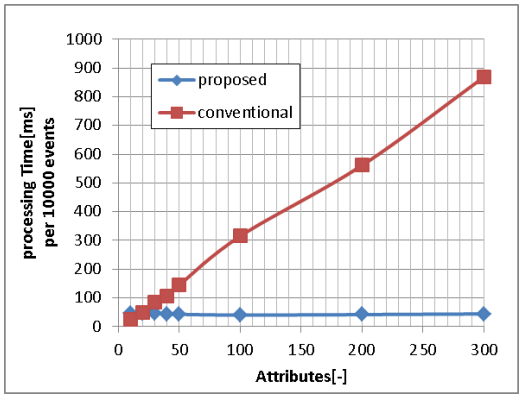

図 3　提案手法と従来手法の照合時間の比較

属性数の増加に対する照合時間が，従来手法では線形時間であるのに対して，提案手法では定数時間である．属性数が 10 種(全属性が変化する場合)を除いて，提案手法は従来手法よりも短時間での照合を実現している．属性数が 100 種の場合においては，従来手法に比べ照合時間を約 87%削減している．

5.　まとめ

本稿では，ルール間距離を利用した差分探索方式の提案，評価を行い，その有効性を検証した．本方式により，対象の各属性が多様に変化する場合にも，多数の属性をリアルタイムに処理することが可能となる．これによりリアルタイムかつ推薦精度の高いレコメンドシステムを実現することが可能となる．

謝辞

本研究の一部は，総務省の「ネットワーク仮想化基盤技術の研究開発」ならびに「ネットワーク仮想化統合技術の研究開発」による委託を受けて実施しています．

参考文献

[1]Marcos K. Aguilera, et al. Matching events in a content-based subscription system, PODC '99, Pages 53-61, May , 1999 .

[2] 佐藤ら，イベントパターンマッチングのための Vaishnavi アルゴリズムの拡張，電子情報通信学会総合大会講演論文集 2008(通信_2),p585,May, 2008.

[3] 小林ら,多属性の照合を実現する拡張トライ木方式の提案と評価,電子情報通信学会ソサイエティ大会講演論文集 2012(通信_2),p365,August, 2012