1Q-05

小売りに関するオントロジーを用いたグラフデータベースによる商品推薦

今野　琢裕†
法政大学情報科学部コンピュータ科学科†
Email: takahiro.konno.8h@stu.hosei.ac.jp†

黄　潤和‡
法政大学情報科学部コンピュータ科学科‡
Email: rhuang@hosei.ac.jp‡

1. まえがき

従来，データ管理に用いられてきた RDB（Relational Database）はデータ関連性の管理が困難である．データ関連性のモデル化や管理はシステムを急激に複雑化させる．また近年 IoT（Internet of Things）によりユーザー，システム，センサーにより生み出されるデータは指数的に増加している．これに対応するために新たなデータタイプや関連性を追加すると，スキーマの再設計・再構築が必要となる．そこで，NoSQL（Not only SQL）の一つである GDB（Graph Database）が注目されている．

本研究では，GDB である Neo4j を選択し，その構成要素の一つであるリレーションシップを用いた新たな知識表現に焦点を当てる．オントロジーに基づき，小売りに関するデータを Neo4j に Concept Layer と Instance Layer の 2 層構造で表現する．GDB のデータ探索の実行時間を測定し，データ属性に適用したオントロジーを評価し考察する．

2. Neo4j と Jess を用いた推薦システム

本研究の Neo4j と Jess を用いた推薦システムの構成を，図 1 に示す．

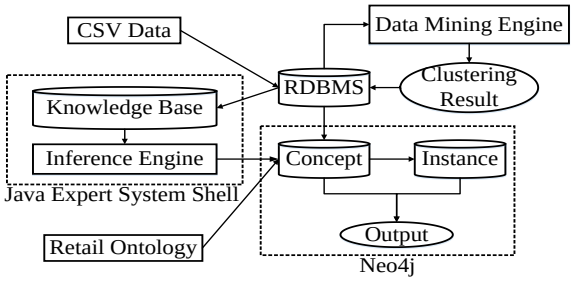

図 1　Neo4j と Jess を用いた推薦システムの構成．

本研究は，図 1 において，csv ファイルを保持する RDBMS，RDBMS から得たデータをクラスタリングする Data Mining Engine，RDBMS のデータを格納する Neo4j，商品の推薦を行う Jess から構成される．

2.1. Two-Layers Structure

本研究は，2 層構造による knowledge store の小売事業への適用を目的としている．2 層構造による knowledge store は，Concept Layer と Instance Layer の 2 つの異なる階層から構成される．上位層である Concept Layer は概念を格納する．下位層である Instance Layer は各概念に対応するデータを格納する．Concept Layer と Instance Layer の一部を図 2 に示す．

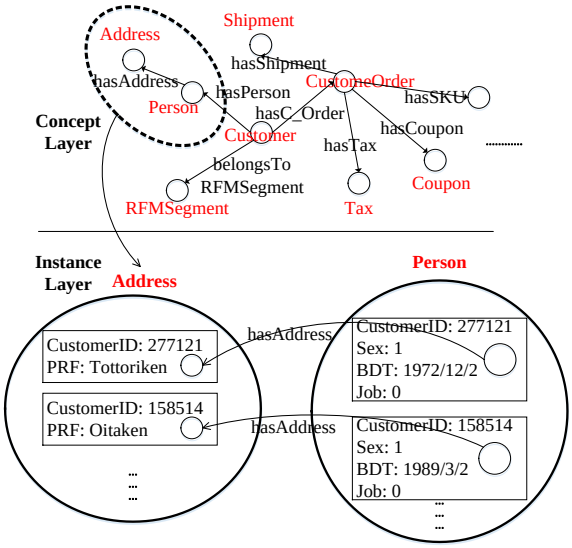

図 2　小売に関する Two Layers Structure の一部．

2.2. Rule-base による商品探索

RDBMS から顧客 ID，RFM 値，セグメント ID，購入商品番号のデータを取り出す．ルールベースにあるルールと照らし合わせ、商品を推薦する．推薦する商品は 5 つにクラスタリングされたセグメントの売上げ上位 20 商品のうちの 1 つである．

3. 実装

本研究で提案するシステムの実装には，Java を使用し，開発とシミュレーションには Windows 上の Eclipse の環境を使用した．

3.1. RDBMS とクラスタリングタスク

顧客情報に関するデータと購入商品に関するデータは 1 年以上の期間にわたる 20000 名の顧客と 400000 個の商品がある． RFM 分析モデルを用い，各顧客データから RFM 値を取り出し，それに基づいて 5 等分し，セグメントを降順に並び替え，各顧客を当てはまるセグメントに割り振る．

3.2. graph database の構築

小売りに関する semantic graph の構築過程を以下 3 つの手順で示す．

[1]：小売りに関するオントロジーから概念を抽出する．
[2]：小売りに関するオントロジーからリレーションシップを抽出する．

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.

[3]：小売りに関するデータと照らし合わせ，各概念にデータ項目を対応付けする．要素層に顧客 ID や明細 ID を含めることでリレーションシップの整合性を保つ．図 3 は semantic graph 構築の流れをモデル化したものである．

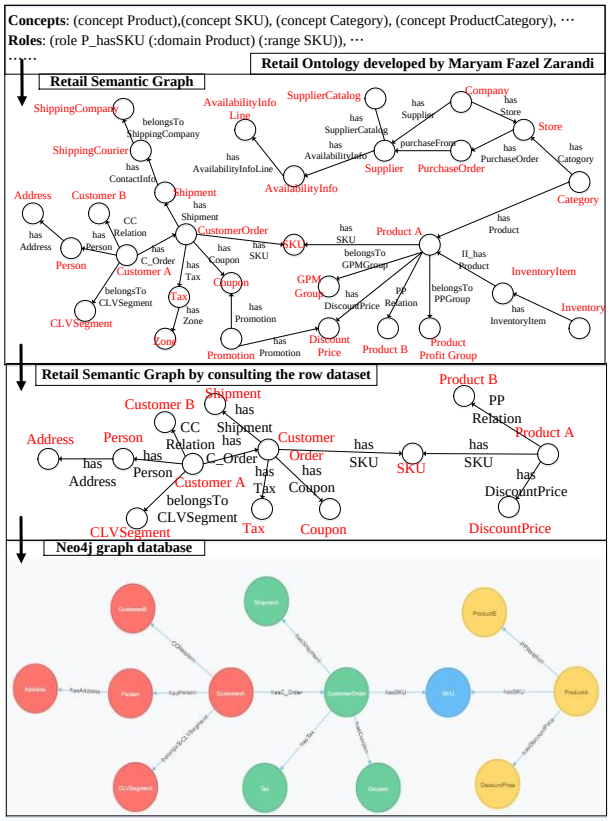

図 3　retail semantic graph 構築の手順.

3.3. graph database と Jess による商品推薦

3.1 節から 3.2 節にかけて説明した商品推薦を行うために必要な処理の流れは以下 8 点である．

①MySQL に csv ファイルを格納する．
②Weka でクラスタリングを行い，その結果を格納する．
③Neo4j にノード sku 以外のノードを作成し（ノード sku に推薦した商品を格納するため），MySQL からデータを取得し格納する．
④ノード sku 以外のリレーションシップを作成する．
⑤Jess による商品推薦を行う．
⑥その結果を txt ファイルで出力し，reclist に格納する．
⑦reclist から Neo4j のノード sku にデータを格納する．
⑧ノード sku に関するリレーションシップを作成する．

4. 評価実験

本研究における評価は 2 点ある．構築したオントロジーの妥当性と，データ探索の実行時間である．

構築したオントロジーの妥当性について，[1]をもとにする．評価項目として，class と class names の割合，properties と property names の割合，*RR*（relationship richness）を用いる．class/class names ratio は式（1），properties/property names ratio は式（2）に示す．

$$R_C(O) = \frac{|C_N(O)|}{|C(O)|} \tag{1}$$

$$R_p(O) = \frac{|P_N(O)|}{|P(O)|} \tag{2}$$

RR は式（3）に示す．本研究では，$|P|$ はプロパティの数を定義し，$|SC|$ はサブクラスの包摂数を定義する．

$$RR = \frac{|P|}{|SC| + |P|} \tag{3}$$

5. 実験結果

構築した小売りに関するオントロジーについて，式（1）と式（2）ともに 1 となった．また *RR* は 0.8659 となった．次に，データ探索の結果を図 4 に示す．

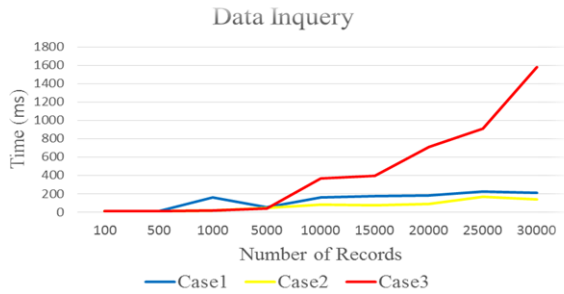

図 4　Neo4j におけるデータ探索時間.

3 つの場合でデータを探索した．Case1 は顧客情報を探索した．Case2 は購買情報に基づいた商品を探索した．Case3 はリレーションシップを探索した．

6. 考察

[2]に基づき，式（4）を用いて推薦リストの新規性を求める．*U* は商品 *i* を購入した利用者のリスト，*R* は推薦リストである[2]．本研究の推薦リストの新規性は 2.2746 であった．[2]の *MIUF* と比較すると，本研究の推薦リストの新規性は十分高いといえる．

$$MIUF = -\frac{1}{|R|}\sum_{i \epsilon R} \log_2 \frac{|U(i)|}{|U|} \tag{4}$$

また式（1）と式（2）の結果より，オントロジーにおける要素の補填に問題はない．式（3）はオントロジーを互いに比較する際に用いる．図 3 最上部のグラフでは，*RR* は 0.7433 となる．よって本研究で用いたオントロジーの方が与えられた計量を正確に捉えることができている．データ探索に関して，Case3 に最も時間を要した．これは該当するノード・リレーションシップ数の違いと，それらの結果の表示に起因していると考えられる．

文　献

[1] Denny Vrandečić，”Ontology Evaluation”，June，2010.
[2] 園田隆志，出雲英剛，佐藤政寛，”食料品小売業の購買傾向に基づく商品推薦指標の提案と評価”，人工知能学会，4J1-2

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.