

DB2 intelliJ Minerを使ったPOSデータ分析システム

大変お待たせしましたが、intelliJ Minerを利用した商品の購入の因果関係を分析する仕組みにトライしています。  
 今回はお客様のデータをお借りして、どのような因果関係があるかを実験してみました。サンプルデータの量が少ない為、十分な状態とはいえませんが面白い結果が出ましたので画面サンプルと分析内容をレポートさせていただきました。  
 是非今後の参考にしていただければ幸いです。

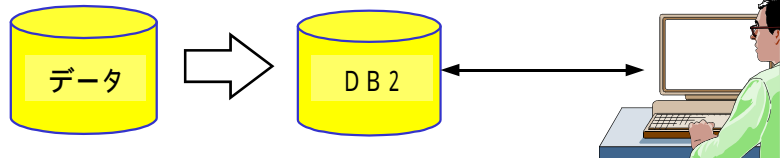
データマイニングを理解する上でのキーワードは下記のものがあります。

- 1. サポート値 全体の売上の中での組合せ購入の確率(支持率)を示しています。
- 2. 確信度 下の例でいうとソース付焼そばを買った人がもやしを同時に購入する確率を言います。
- 3. リフト 下の例でいうと単にもやしを買うお客様に比べ、ソース付焼そばを購入する確率を言います。(5.1倍購入する確率が高いことになります。)

全体の中でソース付焼そばを購入する人がもやしを購入する確率は0.7%だが、ソース付焼そばを購入する人がもやしを購入する確信度は更に46.%と大きい。

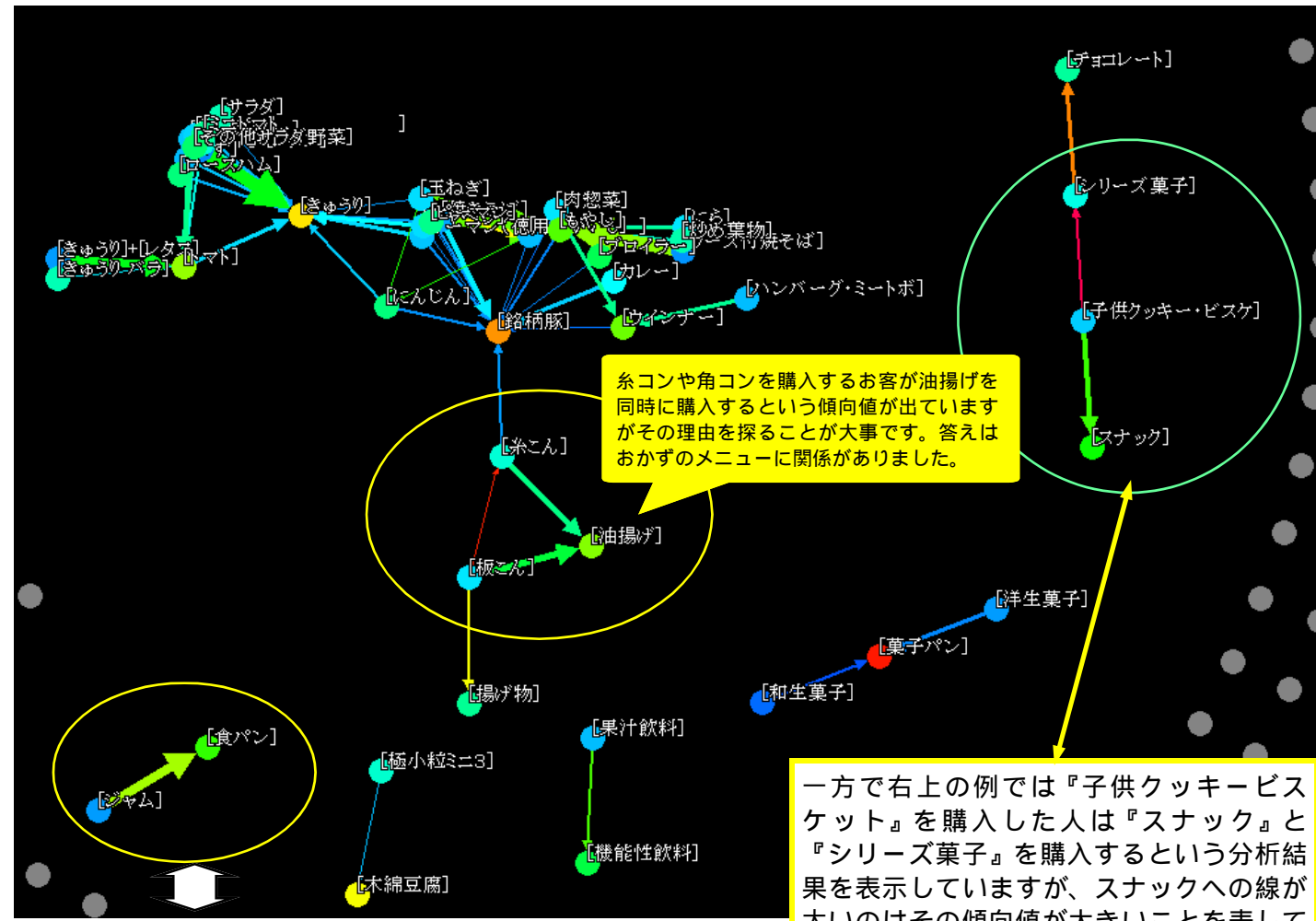
サポート値 (%)	確信度 (%)	タイプ	リフト	ルール・ボディ	ルール・ヘッド
0.7078	46.0800 +		5.1300	[ソース付焼そば]	==> [もやし]
0.6251	45.9700 +		4.2300	[トマト]+[レタス]	==> [きゅうり]
0.8046	44.0000 +		4.8900	[焼そば]	==> [もやし]
0.6583	39.5300 +		5.2300	[ジャム]	==> [食パン]
0.6251	36.9800 +		4.1600	[きゅうり]+[レタス]	==> [トマト]
1.3145	35.0300 +		3.2200	[レタス]	==> [きゅうり]
0.9251	34.4800 +		3.9700	[肉まん]	==> [揚げ物]
1.3124	33.4100 +		3.0700	[大根]	==> [きゅうり]
0.6089	33.2900 +		2.7900	[焼そば]	==> [銘柄豚]
1.6932	32.5500 +		2.9900	[レタス]	==> [きゅうり]
0.7401	32.2400 +		4.6600	[子供クッキー・ビスケット]	==> [スナック]
0.6777	31.3700 +		3.5300	[ブロッコリー]	==> [トマト]
1.1467	31.1000 +		3.5800	[茶こん]	==> [油揚げ]
0.7810	31.0500 +		2.8600	[キャベツ]	==> [きゅうり]
0.6885	30.9500 +		2.8500	[大根]	==> [きゅうり]
0.9660	30.8400 +		2.5800	[カレー]	==> [銘柄豚]
0.6570	30.1600 +		2.2000	[洋生菓子]	==> [菓子パン]
0.9165	30.1300 +		3.3900	[ブロッコリー]	==> [トマト]
0.8154	29.7300 +		3.3100	[くら]	==> [もやし]
0.8003	29.6400 +		3.5500	[肉惣菜]	==> [ウインナー]
0.6562	29.2100 +		2.6900	[ピーマン(徳用)]	==> [きゅうり]
0.6110	29.0400 +		3.4800	[ハンバーグ・ミートポ]	==> [ウインナー]
2.5796	29.0100 +		2.6700	[トマト]	==> [きゅうり]
0.9961	28.6900 +		6.4400	[シリーズ菓子]	==> [チョコレート]
0.6928	28.5000 +		2.6200	[アスパラ]	==> [きゅうり]
1.4178	28.3000 +		2.6000	[こんじん]	==> [きゅうり]
0.6842	28.1400 +		3.1600	[アスパラ]	==> [トマト]
1.2091	28.1300 +		3.1300	[炒め惣菜]	==> [もやし]
1.6330	28.1000 +		2.5900	[こんじん]	==> [揚げ物]
1.0456	28.0800 +		2.5800	[キャベツ]	==> [菓子パン]
1.0994	27.7700 +		2.5600	[なす]	==> [きゅうり]
1.0241	27.7700 +		2.3200	[茶こん]	==> [揚げ物]
0.7573	27.6100 +		3.4700	[くら]	==> [銘柄豚]
0.7035	27.5900 +		2.3100	[玉ねぎ]	==> [こんじん]
1.4136	27.4200 +		2.6200	[その他野菜]	==> [きゅうり]
0.9897	27.3800 +		4.7700	[果汁飲料]	==> [機能性飲料]
0.6089	27.1100 +		2.2700	[ピーマン(徳用)]	==> [銘柄豚]
1.3523	27.0100 +		2.5800	[こんじん]	==> [揚げ物]
1.0112	26.9500 +		3.0300	[レタス]	==> [トマト]
1.2565	26.9200 +		2.4800	[ミニトマト]	==> [きゅうり]
1.3382	26.8200 +		2.4700	[ロースハム]	==> [きゅうり]
0.6110	26.6200 +		7.6700	[子供クッキー・ビスケット]	==> [シリーズ菓子]
2.3817	26.5000 +		2.2200	[もやし]	==> [銘柄豚]
0.7100	26.4600 +		5.7100	[板こん]	==> [揚げ物]
0.5333	26.3500 +		1.9200	[肉惣菜]	==> [菓子パン]
1.2435	26.3400 +		2.4200	[ピーマン]	==> [きゅうり]
0.6713	26.3300 +		5.7800	[玉ねぎ]	==> [じゃがいも]
0.9783	26.2900 +		2.3000	[キャベツ]	==> [銘柄豚]
1.4873	26.2700 +		3.3900	[ブロッコリー]	==> [銘柄豚]
0.6691	26.2400 +		4.5200	[玉ねぎ]	==> [こんじん]

下の例では果汁飲料を購入する人が機能性飲料を購入する確率が高いことを示しています。



実験方法は単品のジャーナル・データをIBMのRDBMSのDB2へコンバートしたものを直接にintelliJ Minerで検索・分析しています。

下の画面では商品のクラス単位での購入傾向を見ています。



系コンや角コンを購入するお客が油揚げを同時に購入するという傾向値が出ていますがその理由を探ることが大事です。答えはおかずのメニューに関係がありました。

一方で右上の例では『子供クッキー・ビスケット』を購入した人は『スナック』と『シリーズ菓子』を購入するという分析結果を表示していますが、スナックへの線が太いのはその傾向値が大きいことを表しています。

ジャムから食パンに矢印が走っています。ジャムを購入する人が食パンを購入する人がジャムを購入する確率より高いために表示されています。(こんにやくと油揚げの関係と同じです。)

今回はまだ実験段階ですが、レタスやトマトを購入する場合に何が一緒に購入されているのか?という検証をしたところ『きゅうり』が一番高い確率が出ていました。このことから野菜売り場での陳列レイアウトの参考データとしても利用できると思われます。

また『カップヌードル』を購入した人は『シーフードヌードル』を購入する比率がその逆より大きいという結果が出ました。シーフードヌードルを購入する顧客は価格が安いからといって標準のカップヌードルを買うわけではないということになります。こうしたことから特売設定でも**カップ及びシーフード・ヌードルを(118円)とそれ以外のシリーズ商品の単価を(88円)**と設定しても、ひょっとすると売上の比率は変わらないかも知れません。一度実験されてはいかがでしょうか?(ちなみにカレー味などは上記の商品の半分程度の売上比率です。)