条件付き確率からはじめるベイジアンネットワーク
～その技術と実社会への応用～

2024年9月27日

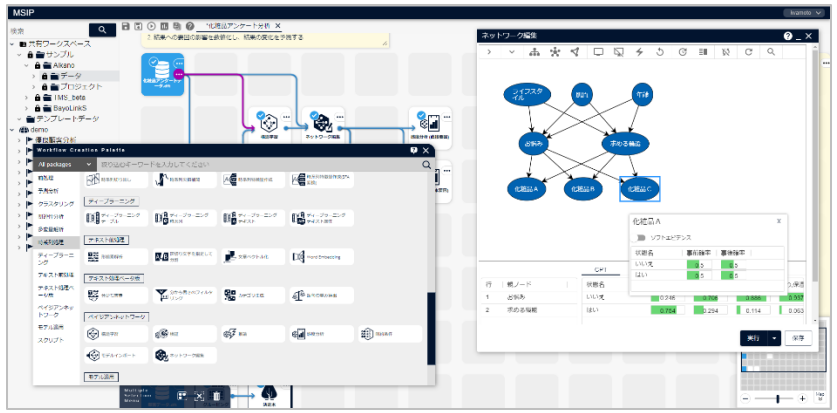
AITeC データマイニングWG / 株式会社NTTデータ数理システム
岩本 圭介

© 2024 NTT DATA Mathematical Systems Inc.

自己紹介

- 岩本 圭介
- 産業技術総合研究所 人工知能技術コンソーシアム（AITeC）
  データマイニングWG リーダー
- (株)NTTデータ数理システム にて、次のような業務に携わっています
  - データ分析ツールの開発
  - AI・機械学習を用いた、お客様の課題解決のための分析伴走サポート

Alkano



社屋：在 東京都新宿区信濃町

人工知能技術コンソーシアム データマイニングWG

- 企業におけるデータ活用を促進し、オープンイベントの開催等を通じてビジネス課題を解決するための情報や事例の共有を行います。
- 特に、データマイニングや機械学習といった、大量データに対するデータ分析技術に焦点を当てています。
  - キーワード：ベイジアンネットワーク、機械学習、テキストマイニング

ステークホルダー

保有データの活用をお考えの企業・研究機関

データ活用の経験や事例をお持ちの企業・研究機関

分析ツールベンダー

解決したい問題

- ご自身の問題に適した分析手段やツールは？
- 問題解決のためどのようなデータを準備すべきか？

アウトプット

- データ分析で問題解決を目指す契機に
- 事例・分析ユースケースの共有

© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

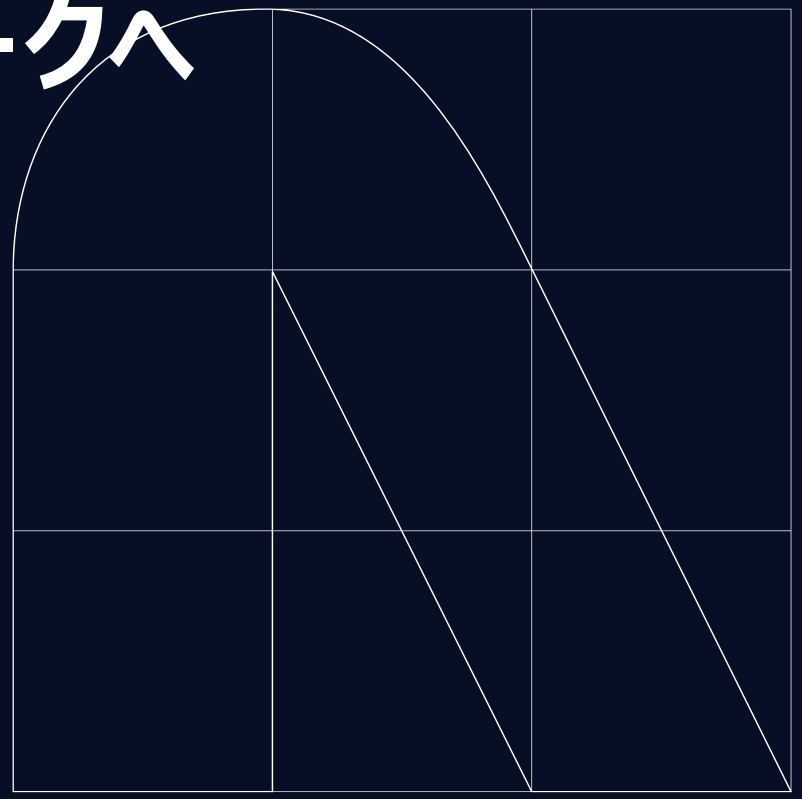
目次

1. 条件付き確率からベイジアンネットワークへ
2. ベイジアンネットワークの3大主要機能
3. ベイジアンネットワークの適用事例

© 2024 NTT DATA Mathematical Systems Inc.

01

条件付き確率からベイジアンネットワークへ



© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

条件付き確率

条件付き確率

■ ある事象 B が起こるという条件の下での、別の事象 A の発生確率

確件付き確率 $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

AとBが同時に発生する確率

Bが単独に発生する確率

■ 前提条件を置くということは、世の中の**不確定性を減らして物事をとらえる**ということ

■ その条件のもとで **世界がどう変わるか？** ということを記述する根本になるものが **条件付き確率**

ターゲットを絞る

特定の事象に注目する

新たな観測結果が得られる

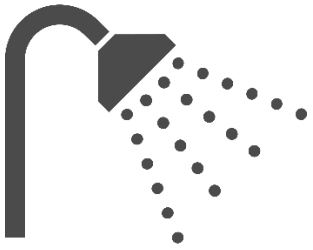
……などなど

© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

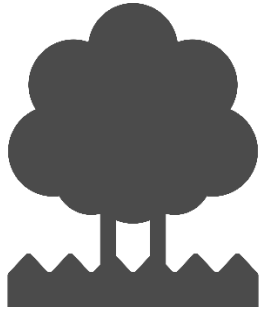
物事の発生を、条件付き確率で記述する

	確率
動作している S_T	0.3
動作していない S_F	0.7



スプリンクラーS

因果関係



庭の芝生W

	スプリンクラー	
	動作している S_T	動作していない S_F
濡れている W_T	0.945	0.45
濡れていない W_F	0.055	0.55

	確率
動作している S_T	$P(S_T) = 0.3$
動作していない S_F	$P(S_F) = 0.7$

	スプリンクラー	
	動作している S_T	動作していない S_F
濡れている W_T	$P(W_T \mid S_T) = 0.945$	$P(W_T \mid S_F) = 0.45$
濡れていない W_F	$P(W_F \mid S_T) = 0.055$	$P(W_F \mid S_F) = 0.55$

スプリンクラーが動作していれば、
高い確率で芝生は濡れるはず

動作していない場合も、それ以外の原因
(降雨など)で芝生は濡れることがあるはず

© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

結果から原因にさかのぼって考える

芝生が濡れている原因はスプリンクラーによるものか否か？

芝生が濡れているときに、スプリンクラーが
動作していた確率はどの程度だろう？

ベイズの法則

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

よって $$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

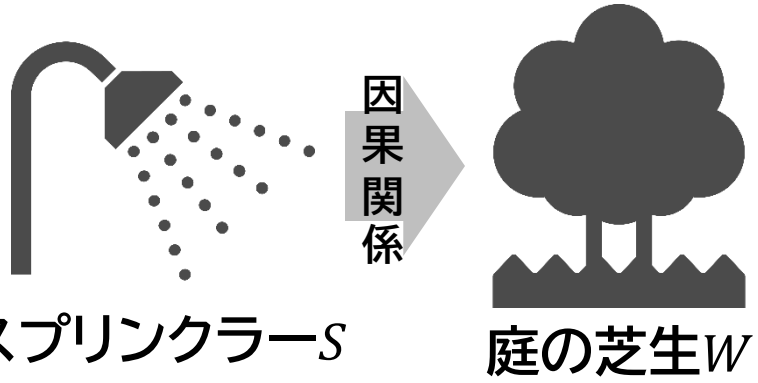
条件付き確率の式
と
そのAとBを入れ替えたもの

「B が起きたときの A の条件付き確率」$P(A|B)$ を利用して、
「A が起きた時の B の条件付き確率」$P(B|A)$ も求められる！

© 2024 NTT DATA Mathematical Systems Inc.

芝生が濡れているときに、スプリンクラーが動作していた確率は？

	確率
動作している S_T	$P(S_T) = 0.3$
動作していない S_F	$P(S_F) = 0.7$



	スプリンクラー	
	動作している S_T	動作していない S_F
濡れている W_T	$P(W_T\|S_T) = 0.945$	$P(W_T\|S_F) = 0.45$
濡れていない W_F	$P(W_F\|S_T) = 0.055$	$P(W_F\|S_F) = 0.55$

$$P(S_T|W_T) = \frac{P(W_T|S_T)P(S_T)}{P(W_T)}$$
$$= 0.474$$

芝生が濡れている確率 $P(W_T)$ は、
「スプリンクラーが動作し、かつ芝生が濡れた」確率と
「スプリンクラーが動作せず、かつ芝生が濡れた確率」の和

$$P(W_T) = P(W_T|S_T)P(S_T) + P(W_T|S_F)P(S_F)$$

事前確率 $P(S_T) = 0.3$ と比較して、芝生が濡れているという条件での事後確率は $P(S_T|W_T) = 0.474$

$$\frac{0.474}{0.3} = 16\% \text{ UP!}$$

© 2024 NTT DATA Mathematical Systems Inc.

要因の影響が多段階になることもある

	確率
雲が出ている C_T	$P(C_T) = 0.5$
雲が出ていない C_F	$P(C_F) = 0.5$

	雲の状態	
	雲が出ている C_T	雲が出ていない C_F
動作している S_T	$P(S_T \mid C_T) = 0.1$	$P(S_T \mid C_F) = 0.5$
動作していない S_F	$P(S_F \mid C_T) = 0.9$	$P(S_F \mid C_F) = 0.5$

	スプリンクラー	
	動作している S_T	動作していない S_F
濡れている W_T	$P(W_T \mid S_T) = 0.945$	$P(W_T \mid S_F) = 0.45$
濡れていない W_F	$P(W_F \mid S_T) = 0.055$	$P(W_F \mid S_F) = 0.55$

© 2024 NTT DATA Mathematical Systems Inc.

結果から原因にさかのぼって考える：3段階の例

芝生が濡れているときに、
雲が出ていた確率はどの程度だろう？

$$P(W_T|\,C_T) = P(W_T|\,S_T)\,P(S_T|\,C_T) + P(W_T|\,S_F)\,P(S_F|\,C_T)$$

$$P(C_T|\,W_T) = \frac{P(\,W_T|C_T)P(C_T)}{P(\,W_T)}$$

$$P(\,W_T) = P(W_T|S_T)P(S_T|C_T)P(C_T) + P(W_T|S_F)P(S_F|C_T)P(C_T) + P(W_T|S_T)P(S_T|C_F)P(C_F) + P(W_T|S_F)P(S_F|C_F)P(C_F)$$

$$= 0.417$$

事前確率 $P(C_T) = 0.5$ と比較して、芝生が濡れている
という条件での事後確率は $P(C_T|\,W_T) = 0.417$

$\frac{0.417}{0.5} = 17\%$ DOWN!

© 2024 NTT DATA Mathematical Systems Inc.

複数の要因の影響を受けることもある

雲の状態C

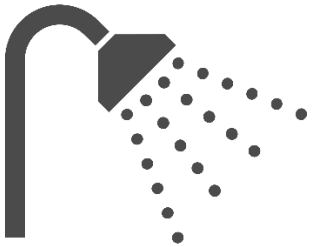
	確率
C_T	0.5
C_F	0.5

因果関係

因果関係

降雨R

	雲の状態	
	C_T	C_F
R_T	0.8	0.2
R_F	0.2	0.8

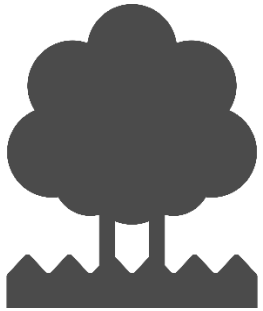


スプリンクラーS

	雲の状態	
	C_T	C_F
S_T	0.1	0.5
S_F	0.9	0.5

因果関係

因果関係



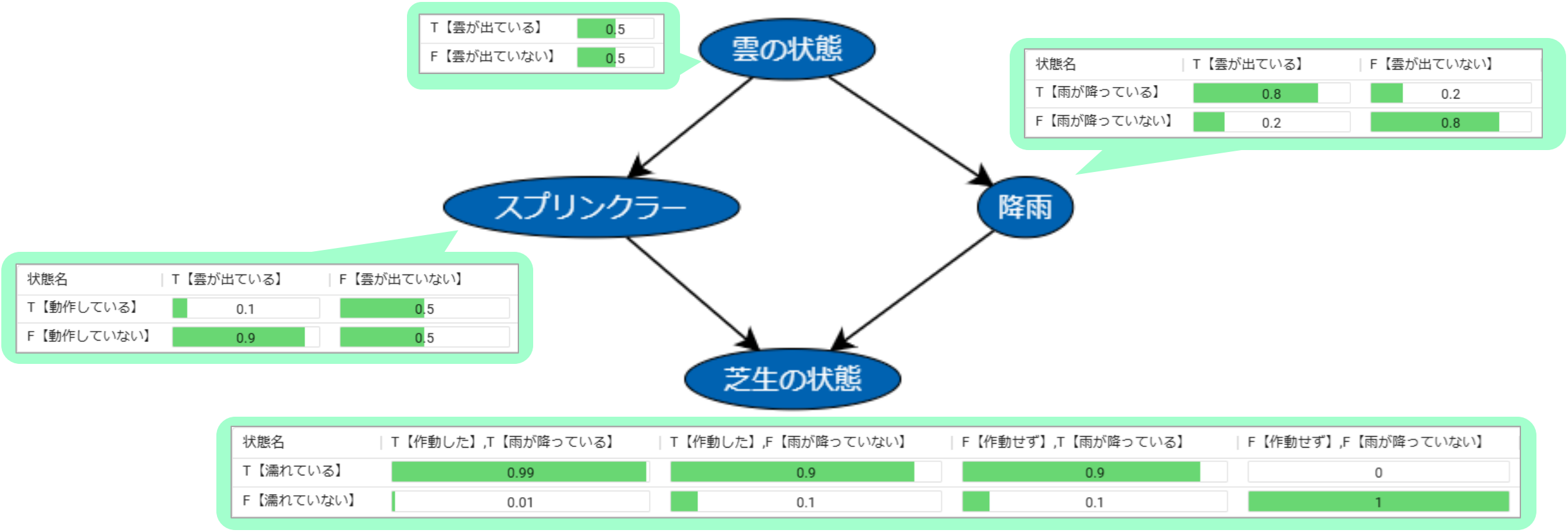
庭の芝生W

	降雨			
	R_T	R_T	R_F	R_F
	スプリンクラー			
	S_T	S_F	S_T	S_F
W_T	0.99	0.9	0.9	0.0
W_F	0.01	0.1	0.1	1.0

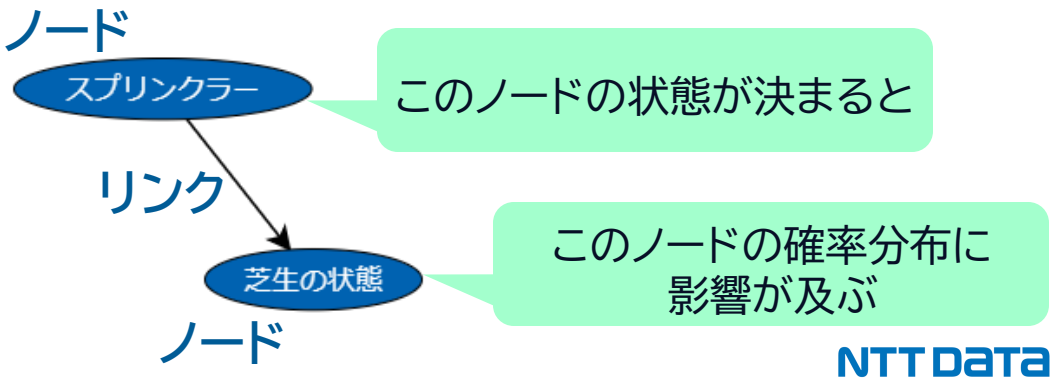
© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

ベイジアンネットワーク



- 要因(＝ 変数)間の関係を 条件付き確率 で表し、ネットワーク構造で表現したモデル
  - ネットワーク構造：ノードとリンクからなる構造



© 2024 NTT DATA Mathematical Systems Inc.

02

ベイジアンネットワークの3大主要機能

© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

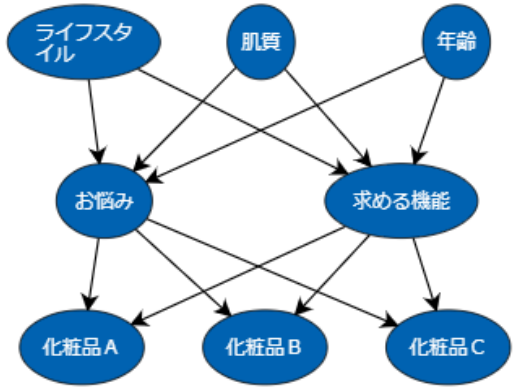
ベイジアンネットワークの構築から利用まで

データ

No.	年齢 Category	ライフスタイル Category	肌質 Category	お悩み Category	求める機能 Category	化粧品A Category	化粧品B Category	化粧品C Category
1	60代	アウトドア	普通肌	しわ	保湿	はい	いいえ	はい
2	30代-50代	インドア	普通肌	毛穴	UV	はい	いいえ	はい
3	60代	インドア	脂性肌	肌のくすみ	美白	はい	はい	はい
4	60代	インドア	脂性肌	肌のくすみ	美白	はい	いいえ	いいえ
5	60代	アウトドア	[illegible]	肌のくすみ	保湿	はい	いいえ	はい
6	60代	アウトドア	乾燥肌	肌のくすみ	保湿	はい	いいえ	はい
7	60代	インドア	[illegible]	[illegible]	[illegible]	はい	はい	はい
8	60代	アウトドア	普通肌	肌のくすみ	UV	はい	はい	はい
9	60代	アウトドア	[illegible]	肌のくすみ	UV	はい	はい	はい
10	60代	インドア	普通肌	肌のくすみ	保湿	はい	いいえ	いいえ
11	30代-50代	インドア	乾燥肌	肌のくすみ	保湿	はい	いいえ	いいえ
12	20代	アウトドア	普通肌	ニキビ	保湿	いいえ	はい	いいえ
13	20代	インドア	混合肌	ニキビ	アンチエイジング	いいえ	いいえ	はい
14	20代	アウトドア	混合肌	ニキビ	美白	いいえ	はい	はい
15	20代	アウトドア	混合肌	毛穴	美白	いいえ	はい	いいえ

構造学習

ベイジアンネットワーク



ベイジアンネットワークを「作る」ステップ

ベイジアンネットワークを「使う」ステップ

確率推論

感度分析

© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

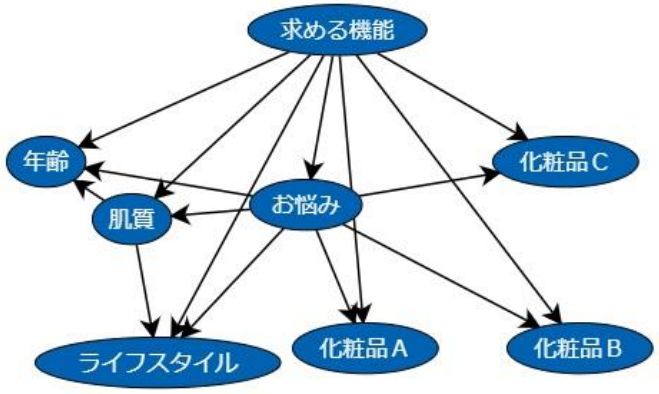
構造学習 = ネットワークの形を決定する

■ データから **構造学習** によってネットワークの構造を決定します。

No.	年齢 Category	ライフスタイル Category	肌質 Category	お悩み Category	求める機能 Category	化粧品A Category	化粧品B Category	化粧品C Category
1	60代	アウトドア	普通肌	しわ	保湿	はい	いいえ	はい
2	30代-50代	インドア	普通肌	毛穴	UV	はい	いいえ	はい
3	60代	インドア	脂性肌	肌のくすみ	美白	はい	はい	はい
4	60代	インドア	脂性肌	肌のくすみ	美白	はい	いいえ	いいえ
5	60代	アウトドア	脂性肌	肌のくすみ	保湿	はい	いいえ	はい
6	60代	アウトドア	乾燥肌	肌のくすみ	保湿	はい	いいえ	はい
7	60代	インドア	普通肌	肌のくすみ	UV	はい	はい	はい
8	60代	アウトドア	普通肌	肌のくすみ	UV	はい	はい	はい
9	60代	アウトドア	普通肌	肌のくすみ	UV	はい	はい	はい
10	60代	インドア	普通肌	肌のくすみ	保湿	はい	いいえ	いいえ
11	30代-50代	インドア	乾燥肌	肌のくすみ	保湿	はい	いいえ	いいえ
12	20代	アウトドア	普通肌	ニキビ	保湿	いいえ	はい	いいえ
13	20代	インドア	混合肌	ニキビ	アンチエイジング	いいえ	いいえ	はい
14	20代	アウトドア	混合肌	ニキビ	美白	いいえ	はい	はい
15	20代	アウトドア	混合肌	毛穴	美白	いいえ	はい	いいえ

基本情報　内面　行動

特に
制約条件を
設定せずに
学習



・・・・？ となることも

- アンケート項目は「基本情報」「内面」「行動」の3カテゴリに分けられると仮定する
- 「基本情報」が「内面」に影響し、「内面」が「行動」に影響すると仮定する

仮説を入れる

© 2024 NTT DATA Mathematical Systems Inc.

構造学習における制約条件

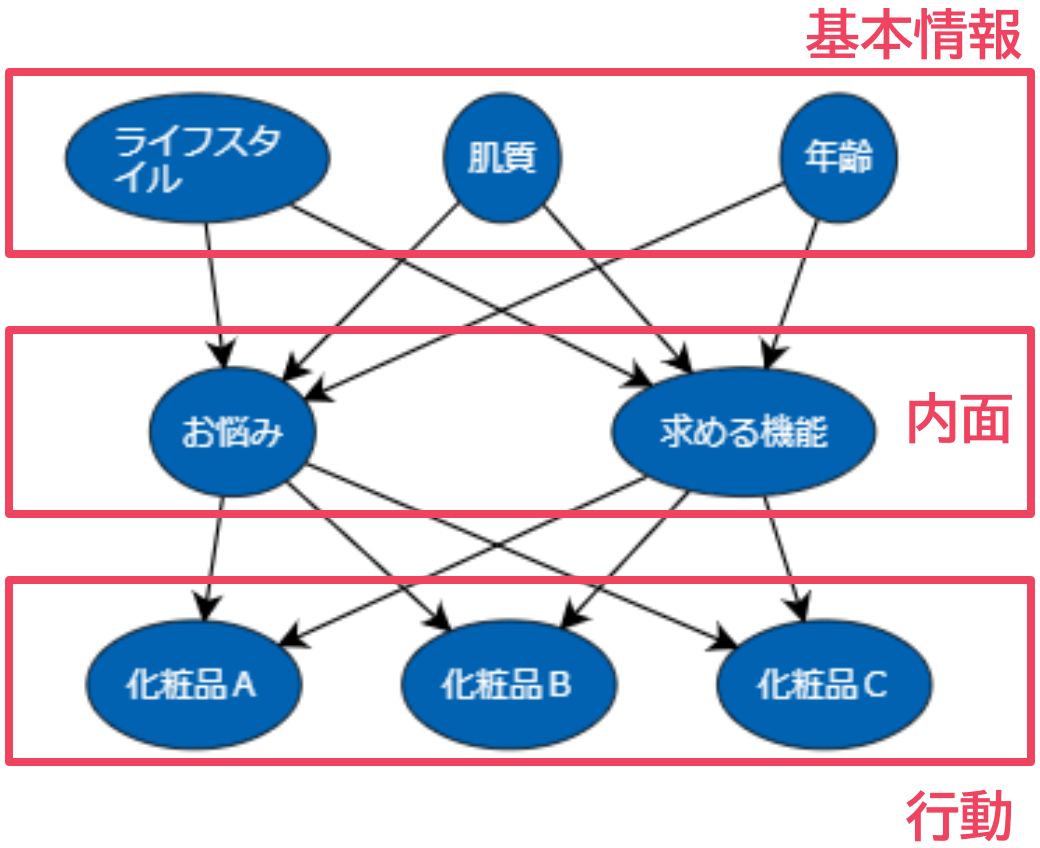
■ 制約条件を設定することで、より**実世界の物事の発生機序を反映した、人にとってより理解しやすい**構造が得られます。

制約条件

- アンケート項目は「基本情報」「内面」「行動」の3カテゴリに分けられると仮定する
- 「基本情報」が「内面」に影響し、「内面」が「行動」に影響すると仮定する

この制約条件は、実世界での知識・知見 や 仮説 をもとに考えます。

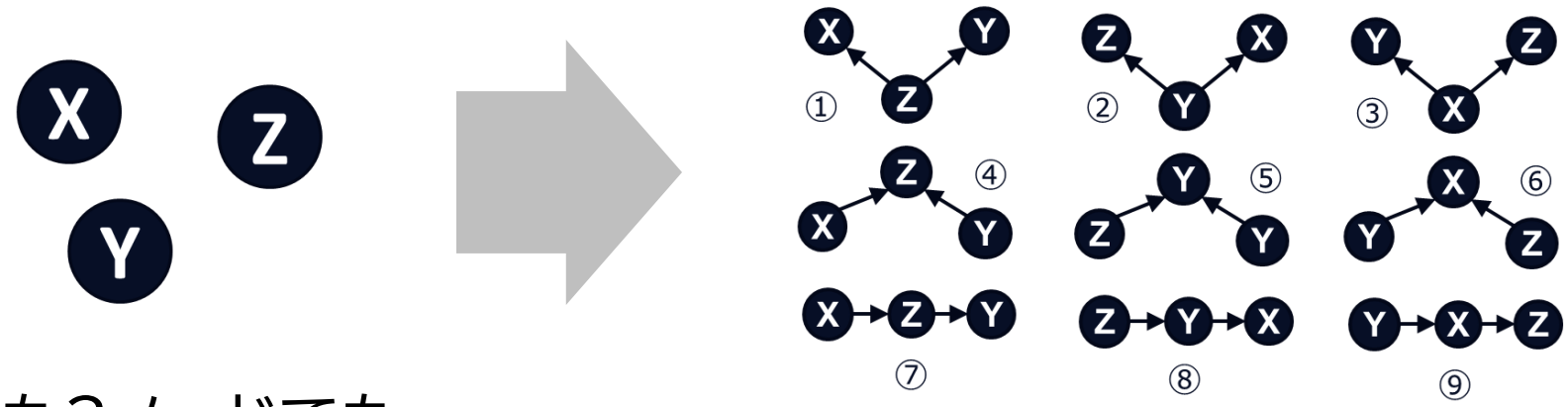
制約条件を設定して学習



© 2024 NTT DATA Mathematical Systems Inc.

ネットワークの形をどのように決定するのか？

- ネットワークの構造と、データの一致度を**スコア化**します。
  - より「**データを良く表している**」**ような構造**を見つけることを目標にします。
- 最もよいスコアのネットワーク構造を得るためには、原理的には考えられる全てのパターンのスコアを計算する必要がありますが、それは現実的ではありません。
  - ノードが 10個 だと 4×10^{18} 個以上のパターンが存在します！

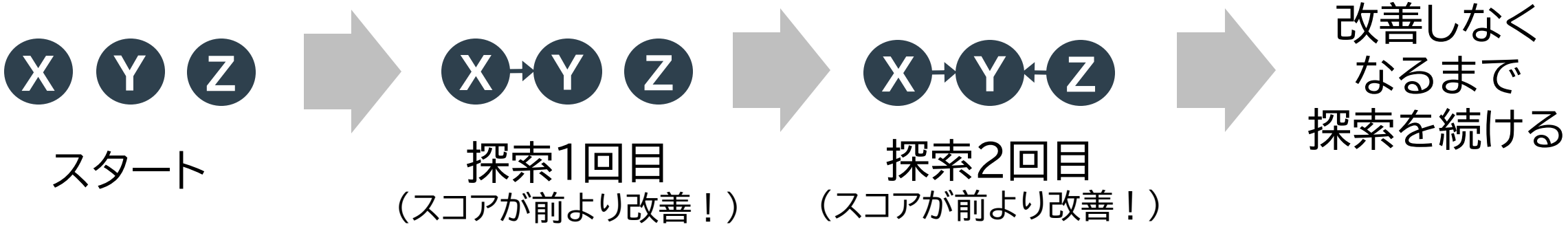


たった3ノードでも・・・・

パターンは9通りもある！

構造学習アルゴリズム

- 全通り計算せず、「**最も良いスコアのネットワーク構造**」**に近いネットワーク**を何とか見つけよう、というアプローチで解決します。
- 基本的なアプローチは、「**スコアが改善しなくなるまでネットワークの変更を行っていく**」という方法です。



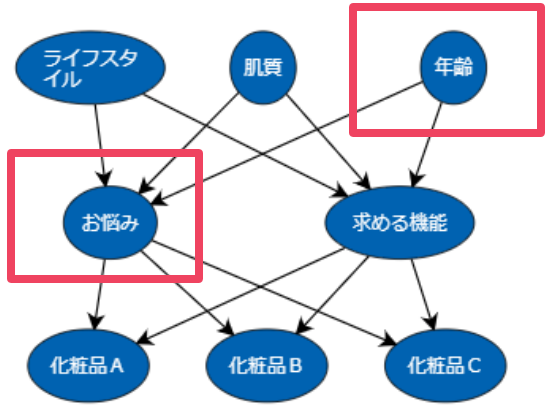
- スコアとして、**AIC**、**BIC** といった量がよく使用されます。
- ネットワークの変更方針として、**Greedy法**、**ヒルクライム法**、といった手法が知られています。

© 2024 NTT DATA Mathematical Systems Inc.

確率推論

■ 確率推論では、前提条件を与えて確率推論を行うことで、前提条件を置いた場合のシミュレーションが可能です。

肌のくすみ に悩む 30～50代の方 は、どういった製品を購入する？



前提条件を指定

お悩み

ソフトエビデンス

状態名	事前確率	事後確率
しわ	0.305	0
ニキビ	0.147	0
肌のくすみ	0.319	1

年齢

ソフトエビデンス

状態名	事前確率	事後確率
20代	0.331	0
30代-50代	0.347	1
60代	0.322	0

推論！

化粧品A

Aの購入確率は増加

状態名	事前確率	事後確率
いいえ	0.516	0.304
はい	0.484	0.696

化粧品B

Bの購入確率は減少

状態名	事前確率	事後確率
いいえ	0.571	0.667
はい	0.429	0.333

化粧品C

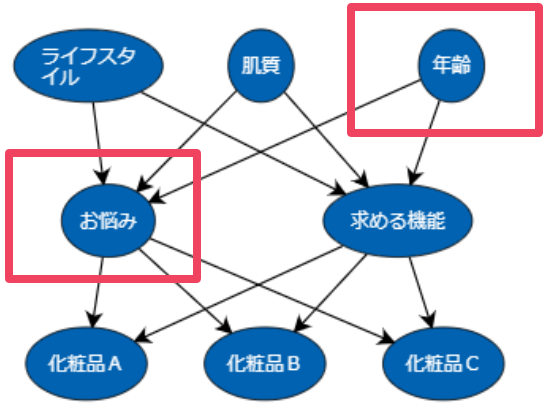
Cの購入確率は減少

状態名	事前確率	事後確率
いいえ	0.59	0.593
はい	0.41	0.407

© 2024 NTT DATA Mathematical Systems Inc.

確率推論 ＝ 条件付き確率を求めること

肌のくすみ に悩む 30～50代の方 は、どういった製品を購入する？



前提条件を指定

お悩み

ソフトエビデンス

状態名	事前確率	事後確率
しわ	0.305	0
ニキビ	0.147	0
肌のくすみ	0.319	1

年齢

ソフトエビデンス

状態名	事前確率	事後確率
20代	0.331	0
30代-50代	0.347	1
60代	0.322	0

推論！

化粧品A

Aの購入確率は増加

状態名	事前確率	事後確率
いいえ	0.516	0.304
はい	0.484	0.696

化粧品B

Bの購入確率は減少

状態名	事前確率	事後確率
いいえ	0.571	0.667
はい	0.429	0.333

化粧品C

Cの購入確率は減少

状態名	事前確率	事後確率
いいえ	0.59	0.593
はい	0.41	0.407

条件付き確率 $P(\text{製品}A|30\sim50\text{代},\text{くすみ})$, $P(\text{製品}B|30\sim50\text{代},\text{くすみ})$, $P(\text{製品}C|30\sim50\text{代},\text{くすみ})$ を求めることに相当します

© 2024 NTT DATA Mathematical Systems Inc.

確率推論アルゴリズム

■ 条件付き確率の厳密値を求めるには、一般に莫大な計算量が必要です。

➢ 現実的な時間で厳密解を得ることが難しいため、LoopyBP といった**近似解法**のための確率推論アルゴリズムが用いられます。

■ 3ノードだと・・・・

結果から原因にさかのぼって考える：3段階の例

芝生が濡れているときに、
雲が出ていた確率はどの程度だろう？

$$P(W_T|C_T) = P(W_T|S_T)\,P(S_T|C_T) + P(W_T|S_F)\,P(S_F|C_T)$$

$$P(C_T|W_T) = \frac{P(W_T|C_T)P(C_T)}{P(W_T)}$$

$$P(W_T) = P(W_T|S_T)P(S_T|C_T)P(C_T) + P(W_T|S_F)P(S_F|C_T)P(C_T) + P(W_T|S_T)P(S_T|C_F)P(C_F) + P(W_T|S_F)P(S_F|C_F)P(C_F)$$

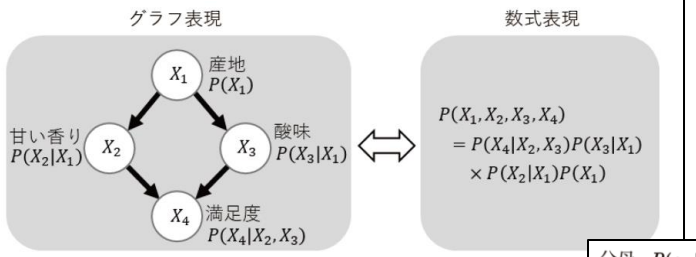
$$= 0.417$$

事前確率 $P(C_T) = 0.5$ と比較して、芝生が濡れている
という条件での事後確率は $P(C_T|W_T) = 0.417$

$\frac{0.417}{0.5} = 17\%$ DOWN!

© 2024 NTT DATA Mathematical Systems Inc.

■ 4ノードだと・・・・



$$P(X_2 \mid e_4) = \frac{P(e_4, X_2)}{P(e_4)}$$

分母 $= P(e_4 \mid x_2 = なし, X_3 = なし)P(x_3 = なし \mid x_1 = A)P(x_2 = なし \mid x_1 = A)P(x_1 = A)$
$+ P(e_4 \mid x_2 = なし, x_3 = あり)P(x_3 = あり \mid x_1 = A)P(x_2 = なし \mid x_1 = A)P(x_1 = A)$
$+ P(e_4 \mid x_2 = あり, x_3 = なし)P(x_3 = なし \mid x_1 = A)P(x_2 = あり \mid x_1 = A)P(x_1 = A)$
$+ P(e_4 \mid x_2 = あり, x_3 = あり)P(x_3 = あり \mid x_1 = A)P(x_2 = あり \mid x_1 = A)P(x_1 = A)$
$+ P(e_4 \mid x_2 = なし, x_3 = なし)P(x_3 = なし \mid x_1 = B)P(x_2 = なし \mid x_1 = B)P(x_1 = B)$
$+ P(e_4 \mid x_2 = なし, x_3 = あり)P(x_3 = あり \mid x_1 = B)P(x_2 = なし \mid x_1 = B)P(x_1 = B)$
$+ P(e_4 \mid x_2 = あり, x_3 = なし)P(x_3 = なし \mid x_1 = B)P(x_2 = あり \mid x_1 = B)P(x_1 = B)$
$+ P(e_4 \mid x_2 = あり, x_3 = あり)P(x_3 = あり \mid x_1 = B)P(x_2 = あり \mid x_1 = B)P(x_1 = B)$
$= 0.15 \times 0.35 \times 0.70 \times 0.75 + 0.35 \times 0.65 \times 0.70 \times 0.75$
$+ 0.64 \times 0.35 \times 0.30 \times 0.75 + 0.95 \times 0.65 \times 0.30 \times 0.75$
$+ 0.15 \times 0.67 \times 0.10 \times 0.25 + 0.35 \times 0.33 \times 0.10 \times 0.25$
$+ 0.64 \times 0.67 \times 0.90 \times 0.25 + 0.95 \times 0.33 \times 0.90 \times 0.25$
$= 0.508 \cdots$

© 2024 NTT DATA Mathematical Systems Inc.

感度分析

■ ターゲットを固定して、関連性の強い要因を一覧でリストアップすることができます。

ターゲット
「化粧品A を 購入する」に固定

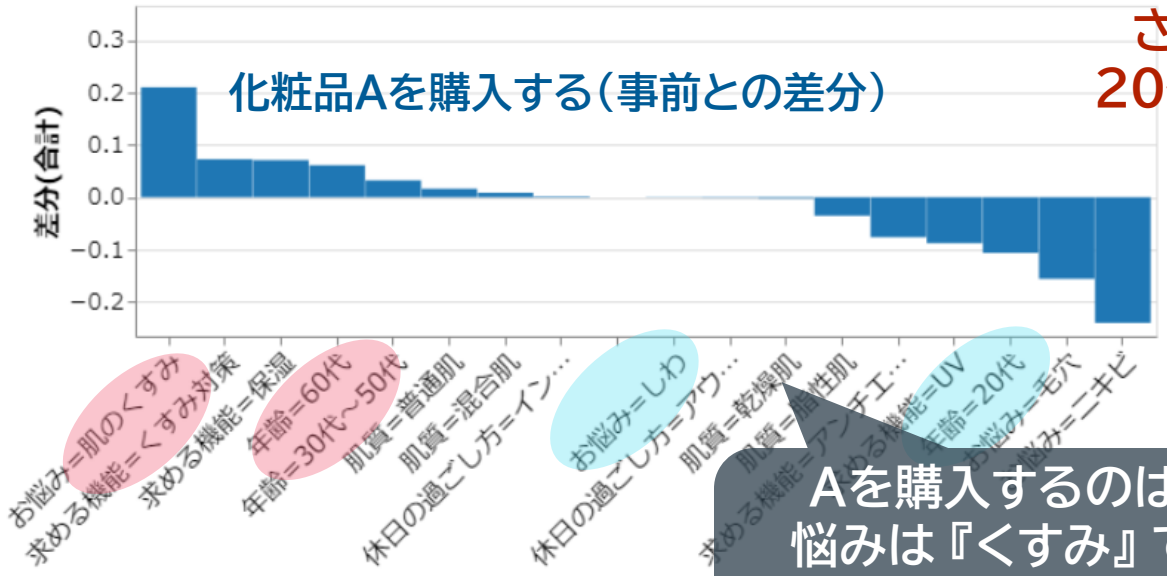
その条件のもとで、
確率がどれくらい up するか？

その条件は？

No.	目的変数 Category	1 値 Cate	事前確率 Float	事後確率 Float	2 差分 Float	リフト Float	年齢 Category	ライフスタイル Category	肌質 Category
1	化粧品 A	はい	0.484313	0.566671	0.082357	1.170050	60代	アウトドア	普通肌
2	化粧品 A	はい	0.484313	0.562949	0.078636	1.162366	60代	インドア	脂性肌
3	化粧品 A	はい	0.484313	0.561552	0.077239	1.159482	60代	アウトドア	混合肌
4	化粧品 A	はい	0.484313	0.558268	0.073955	1.152701	60代		脂性肌
5	化粧品 A	はい	0.484313	0.558029	0.073716	1.152207	60代		普通肌
6	化粧品 A	はい	0.484313	0.554636	0.070323	1.145202	60代	アウトドア	脂性肌
7	化粧品 A	はい	0.484313	0.549454	0.065140	1.134501	60代	アウトドア	
8	化粧品 A	はい	0.484313	0.548767	0.064454	1.133084	60代	インドア	普通肌
9	化粧品 A	はい	0.484313	0.545319	0.061006	1.125964	60代		
10	化粧品 A	はい	0.484313	0.542767	0.058454	1.120695	60代	インドア	乾燥肌
11	化粧品 A	はい	0.484313	0.541214	0.056901	1.117488	30代-50代	インドア	普通肌
12	化粧品 A	はい	0.484313	0.540883	0.056570	1.116805	60代	インドア	
13	化粧品 A	はい	0.484313	0.538063	0.053750	1.110981	60代		混合肌
14	化粧品 A	はい	0.484313	0.534852	0.050539	1.104352	30代-50代	インドア	混合肌
15	化粧品 A	はい	0.484313	0.531223	0.046910	1.096858	30代-50代	インドア	

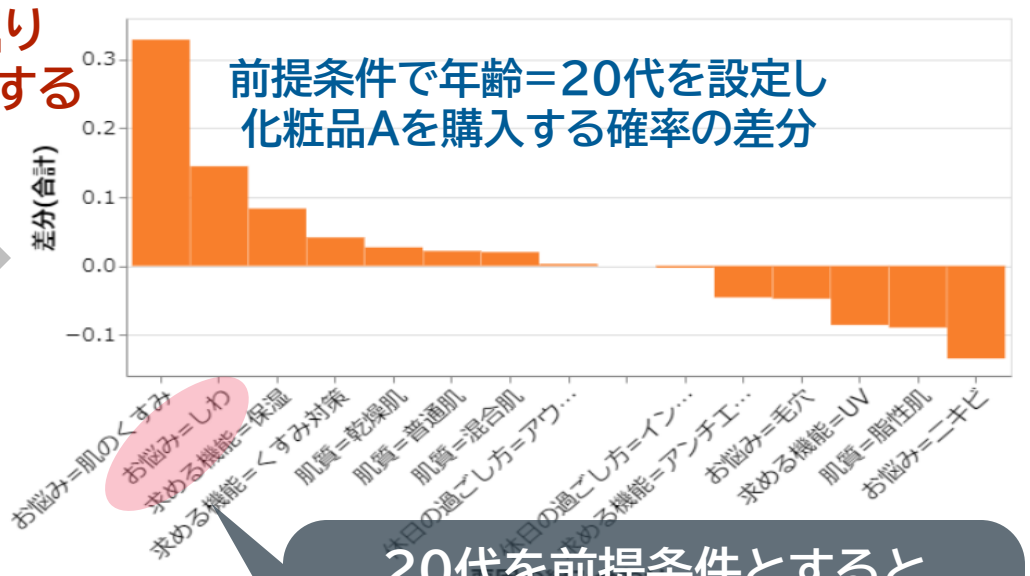
© 2024 NTT DATA Mathematical Systems Inc.

感度分析の活用

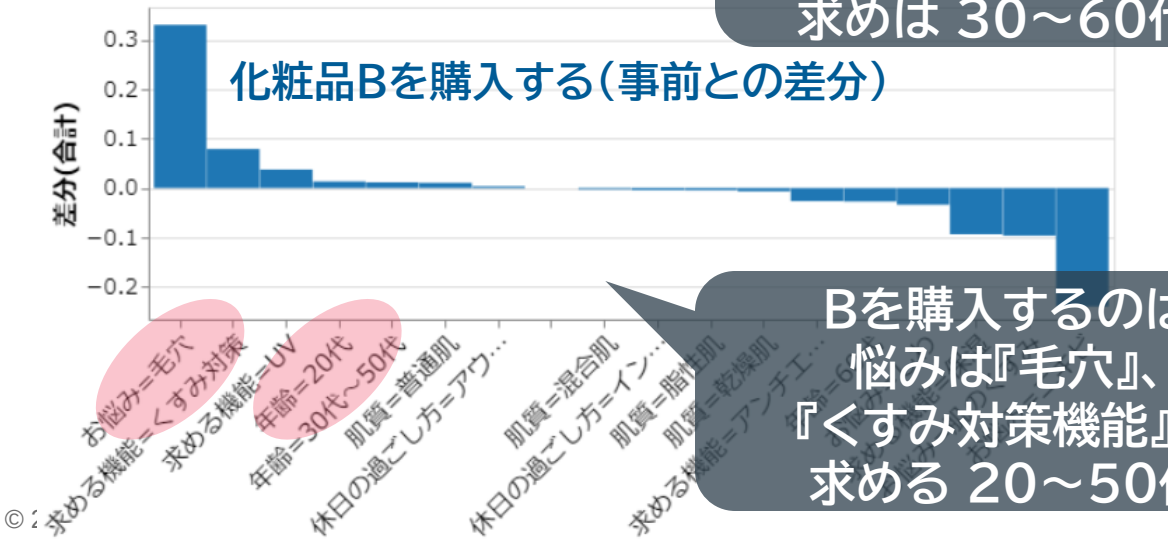


Aを購入するのは
悩みは『くすみ』で
『くすみ対策機能』を
求めは 30～60代

さらに深堀り
20代が購入する
条件は？



20代を前提条件とすると
悩み『しわ』が上位に浮上
→『しわ』改善をアピールすると
20代の購入確率が上がる!



Bを購入するのは
悩みは『毛穴』、
『くすみ対策機能』を
求める 20～50代

感度分析の利用で
「化粧品の購入確率を上昇させる条件」を
効率的かつ多角的に
分析することができます。

NTT DATA

ベイジアンネットワークの特徴

1 依存関係をネットワークで可視化

変数間の依存関係をネットワークで表現します。直接的・間接的に影響を与える要因を直感的に把握できます。**人が理解しやすい表現形式なので、これまで気づかなかった要因に新たに気付くことも。**

2 答えを確率で表現

変数間の関係を条件付き確率で表します。そこに既知の情報（エビデンス）を与え、未知の知りたい答えを確率計算によって推論します。そのため、**不確実性を含む問題に対して自然な予測**を行うことができます。確率的なアプローチにより、同じ条件でも異なる結果が出るような問題を自然に扱うことができます。

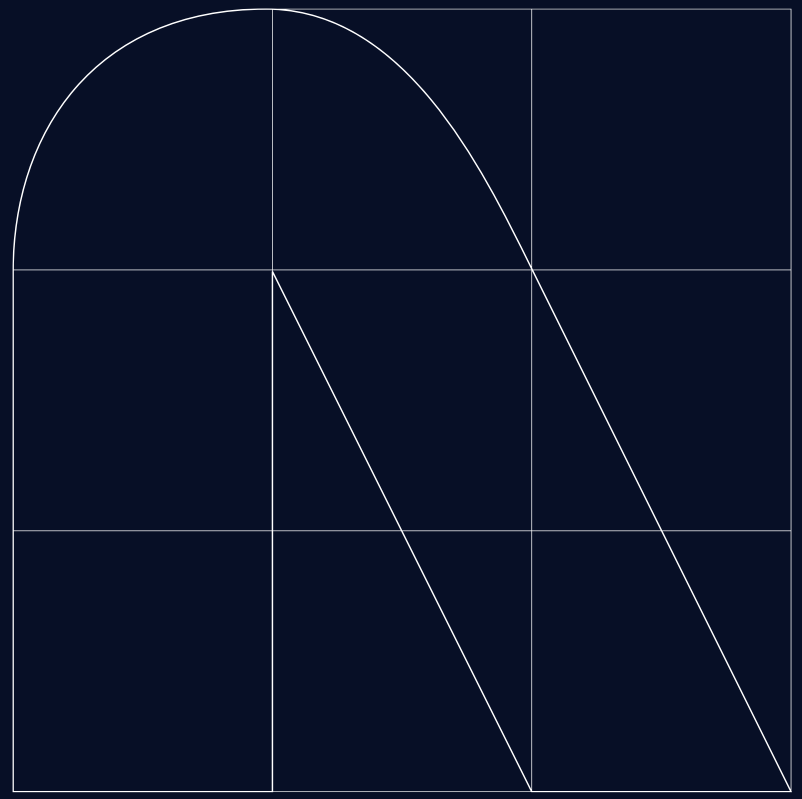
3 双方向の要因分析が可能

複数の既知の情報から知りたい結果を**「順方向に推論」**できます。反対に、結果からそれが起こるための要因を導き出す**「逆方向の推論」**もできます。これにより効率的に要因分析を進められます。

© 2024 NTT DATA Mathematical Systems Inc.

03

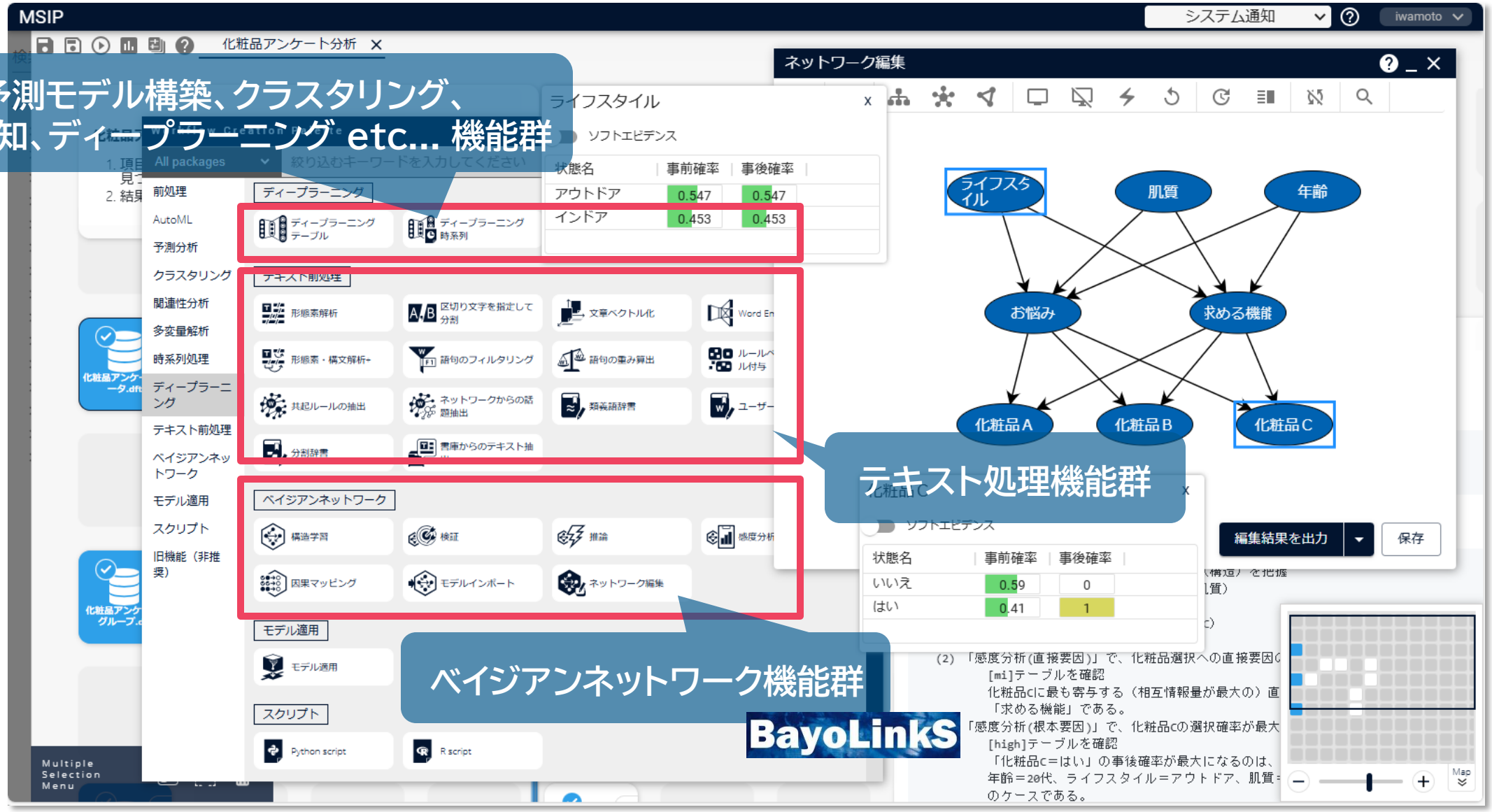
ベイジアンネットワークの適用事例



© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA

データ分析プラットフォーム Alkano



予測モデル構築、クラスタリング、
異常検知、ディープラーニング etc... 機能群

テキスト処理機能群

ベイジアンネットワーク機能群

BayoLinkS

ベイジアンネットワーク活用事例のご紹介

下記サイトで紹介を行っています。ぜひともご参照ください。
https://www.msiism.jp/case/bayolinks/

© 2021 NTT DATA Mathematical Systems Inc.

ベイジアンネットワークの活用は様々な場面で広がっています

■ ベイジアンネットワークの基本的な知識から、実務の現場でベイジアンネットワークが使われている実践例までを幅広く紹介した書籍が出版されました。

BayoLinkSで実践するベイジアンネットワーク
著者：本村陽一 編著
発行元：オーム社

章題例

- ベイジアンネットワークによる製造情報論の実現
- 因果連鎖分析とベイジアンネットワーク
- 思考力を拡張させるベイジアンネットワーク
- ベイジアンネットワークをBayoLinkS で体験しよう
- ……

多くの BayoLinkS ユーザー様より
ご寄稿を頂きました。是非お手にとってみてください。

https://www.ohmsha.co.jp/book/9784274230486/

© 2024 NTT DATA Mathematical Systems Inc.

お問い合わせ先

■ AITeC データマイニングWG に関して

- WGリーダー NTTデータ数理システム 岩本：iwamoto@msi.co.jp

■ データ分析プラットフォーム Alkano に関して

- 株式会社 NTTデータ数理システム：sales@ml.msi.co.jp

© 2024 NTT DATA Mathematical Systems Inc.

NTT DATA