

テキストマイニングとベイジアンネット を用いた日常生活行動モデリング ～傷害予防や認知症支援を例にして～

(独)産業技術総合研究所
サービス工学研究センター
大規模データモデリング研究チーム&
デジタルヒューマン研究センター
本村 陽一
y.motomura@aist.go.jp

生活中的履歴データなどの大規模 データから再利用できる知識へ

従来の統計(母集団依存性大)

- 各事例の保存と検索(データベース)
- 単独の頻度、集計、統計
- 複数の変数の間の関係(相関、独立性)
- ある変数を説明するルール、予測モデル
- 変数群の依存関係の条件付確率(確率的知識)
- 因果構造のネットワーク(因果的知識)

有用な知識(再利用性大)

集めたデータの説明から、そのデータ以外にも成り立つ知識、
すなわち過去の経験から、『次』に活かせる知識へ!!

表したい知識とモデルの性質

- 因果的關係(ネットワーク構造モデル)
- 不確実性と多様性(ベイズ)
- 日常生活の行動(センサ)と心理(テキスト)



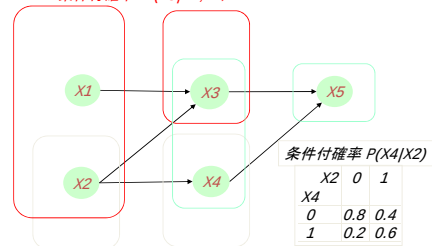
ベイジアンネットワークによるモデリングと
テキストマイニングの事例

ベイジアンネット

和文書籍

ベイジアンネットワーク技術: 東京電機大学出版局(本村・岩崎)
ベイジアンネットワーク概説: 培風館(繁例・植野・本村)

条件付確率: $P(\text{目的変数} | \text{説明変数})$
条件付確率 $P(X3|X1, X2)$



ベイジアンネットの歴史

- ベイジアンネットワーク
主に人工知能分野の中で発展(1980年代～)
- モデル: 不確実性を表す命題論理の確率化
国際会議: Uncertainty in AI (1985～)
- アルゴリズム(主に1990年～)
 - 確率推論(モデルが表す確率分布の効率的計算方法)
 - 統計的学習(モデルの構造とパラメータの統計的推定)
- 応用(情報機器の小型化・個人化とも関連)

ベイジアンネットの学習と確率推論

統計データ → 知識表現・モデル → 予測・制御・意思決定

統計的学習
統計的推論

人工知能、エキスパート
システムでの推論・意思決定

ベイジアンネットの学習

ベイジアンネットの確率推論

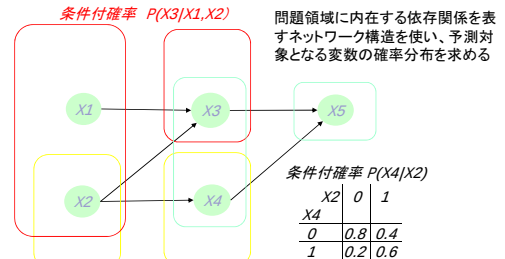
ベイジアンネットの ベイズ的世界観

- ドメインを全て確率変数として表現
 - ドメインの状態は各確率変数の同時分布
 - 観測可能、決定的な場合を含む
 - 末端の変数は事前確率分布を仮定
-
- 目的: 知りたい対象の変数の事後確率分布を計算
 - この時、主要な依存関係のみをグラフとして表し、モデル化することで記述量、計算量を劇的に削減。

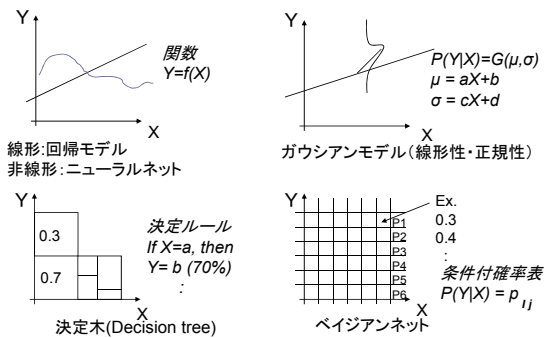
$$P(X_1, X_2, \dots, X_n) = P(X_1|X_2)P(X_2|\dots)P(X_n)$$

確率モデルとしての特徴

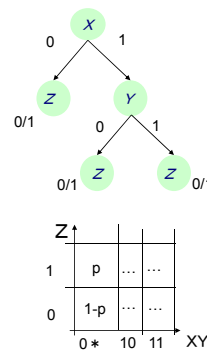
条件付確率: $P(\text{目的変数}|\text{説明変数})$



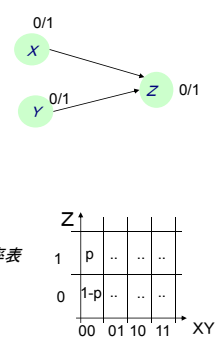
各種モデルの変数間の共変関係



決定木

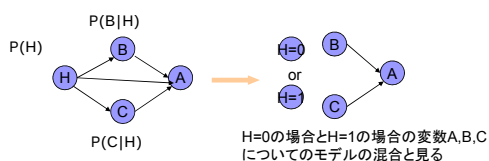


ベイジアンネット

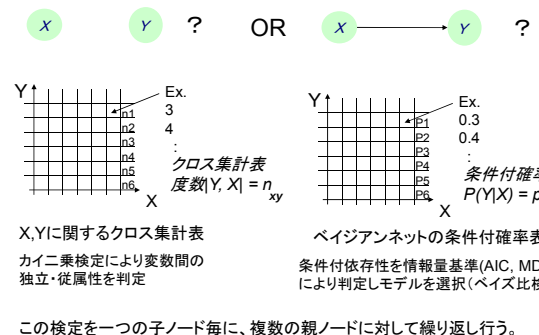


ベイジアンネットの中の ベイズ的世界観

- 変数(H)を潜在クラスとして見た混合モデル
- ハイパーパラメータとして見た階層ベイズ
- 主観確率としてみた信念モデル、意思決定



ベイジアンネットの学習(グラフ構造の探索)

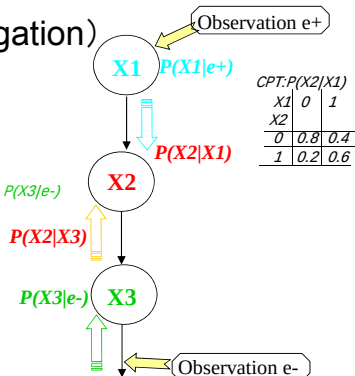


確率推論

(belief propagation)

もっとも単純な場合 (基本系)

$$\begin{aligned} \text{Belief}(X_2) &= P(X_2|e^+, e^-) \\ &= P(X_2|X_1) P(X_1|e^+) \cdot P(X_2|X_3) P(X_3|e^-) \end{aligned}$$

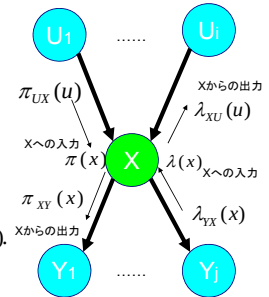


CPT: $P(X_2|X_1)$

X_1	0	1
X_2	0.8	0.4
	0.2	0.6

Belief propagationアルゴリズム

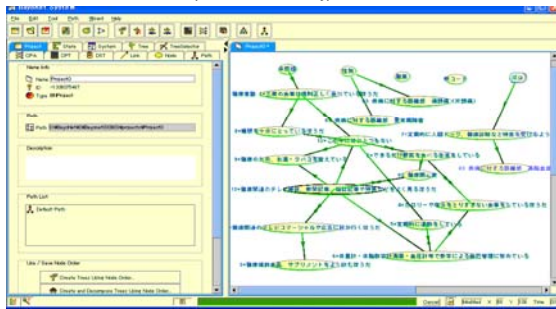
$$\begin{aligned} \Pr(X=x) &= \alpha \lambda(x) \pi(x), \\ \pi(x) &= \sum_u P(X|U=u) \prod_{U_i} \pi_{U_i X}(U_i), \\ \lambda(x) &= \prod_{Y_j} \lambda_{Y_j X}(x), \\ \pi_{XY}(x) &= \pi(x) \prod_{k \neq j} \lambda_{Y_k X}(x), \\ \lambda_{XU_i}(u) &= \sum_x \lambda(x) \sum_{k \neq i} P(x|U) \prod_{k \neq i} \pi_{U_k X}(u_k). \end{aligned}$$



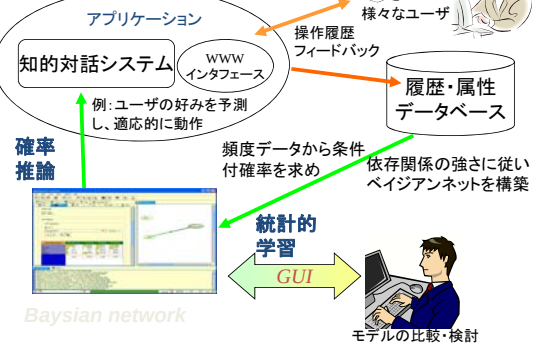
ベイジアンネットワーク構築・推論ソフトウェア BayoNet

(産総研で開発され、数理システムから販売)

<http://www.msi.co.jp/BAYONET/>

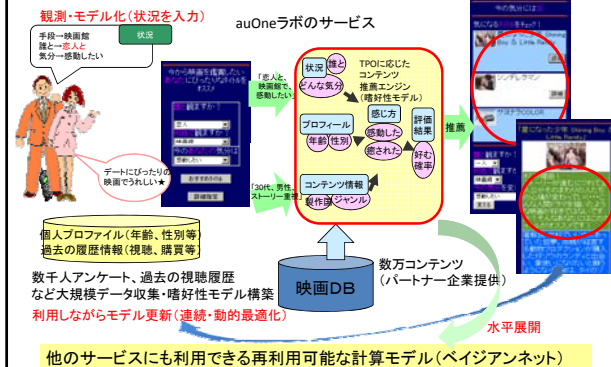


ベイジアンネットワークを用いた大規模データからのモデリング

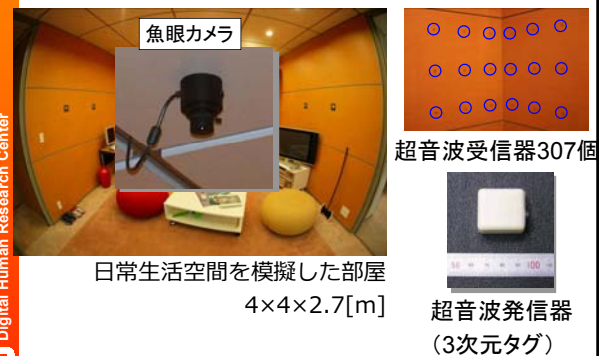


生活者の嗜好性のモデル化

映画推薦サービスにより集積する大規模データをマーケティングに活用



生活行動データの収集と予測モデルの構築



行動履歴データ

子供の行動観測

1. カメラで撮影した動画に対して、子供の行動に行動ラベルを振る
2. 同時に子供の位置も超音波センサで計測して対応づける



観測画像

- ▶ 1秒ごとに行動に対してラベル付けを行う
- ▶ 例
 - ▶ 歩いている
 - ▶ 座っている
 - ▶ 手に〇〇を持っている
 - ▶ など

ベイズ推定

ベイズ事後確率

$$P(C = c_i | x) = \frac{P(x | C = c_i)P(C = c_i)}{\sum_j P(x | C = c_j)P(C = c_j)}$$

C : クラスラベル
 $P(C|x)$: 事後分布
 $P(x|C)$: 特徴量 x に対する尤度

X : 特徴量 (35次元の高次自己相関特徴)
 $P(C)$: 事前分布

Naïve Bayes識別器

クラスラベルの尤度が分解できると仮定する

$$P(x | C = c_i) \propto \prod_j P(x_j | C = c_j)$$

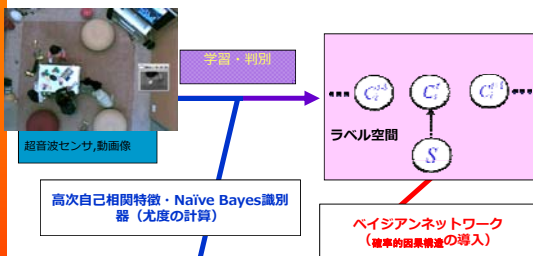
事後分布を最大化するクラスラベルを推定する

事前分布には一様分布を与える

最尤推定に置き換えられる

ベイジアンネットを用いた状況依存ベイズ推定モデル

(本村・西田: 情報処理学会CVIM誌, 2007 他)



$$P(C_i | x, S) \propto P(x | C_i) \times P(C_i | S)$$

再利用可能な確率的因果構造へ

行動推定を行うための確率変数

- x, y 方向の速度: $V_{x,y}(t)$
- 子供の高さ: $Z(t)$
- 1秒前の行動: $C(t-1)$
- 子供の位置情報

子供の位置情報



室内の座標系の利用
絶対座標での分割



モノとの相対距離
相対距離

日常生活の確率的因果構造のモデル化

長期生活型情報収集システム



掃除行動の例

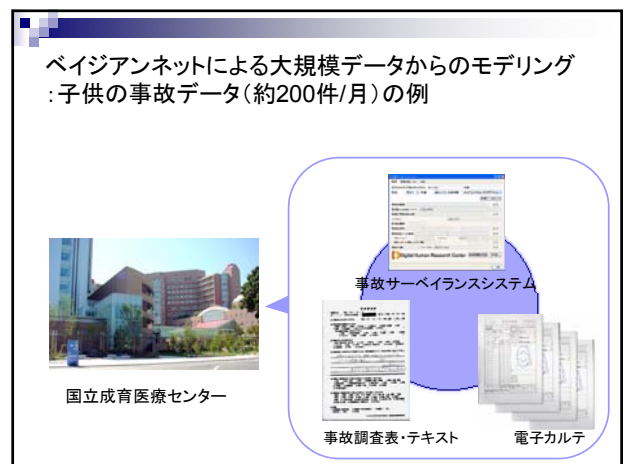
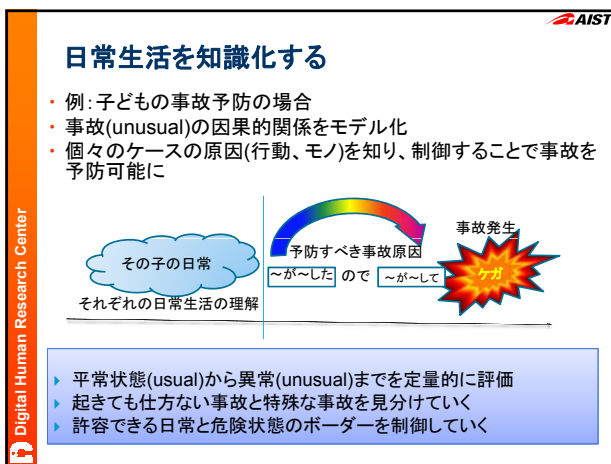
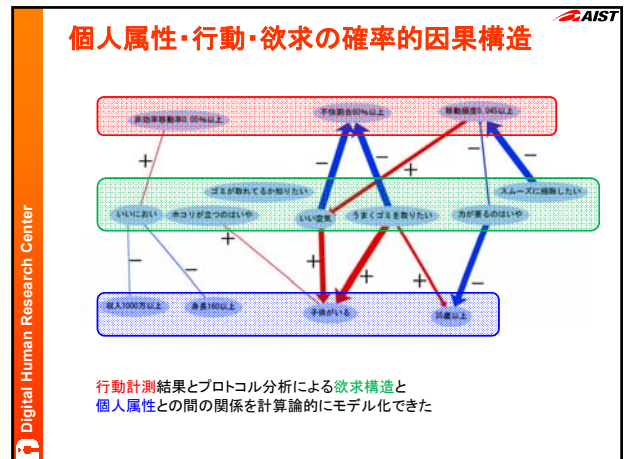
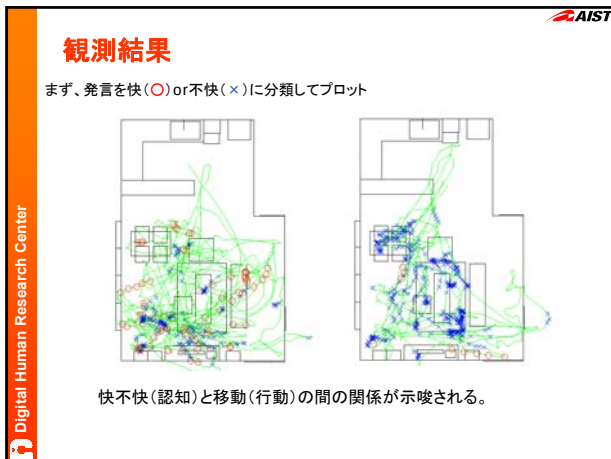
掃除行動実験

- ・5人の被験者対象
- ・ほぼ同じ状態の部屋を自由に掃除してもらう
- 5人の行動の違いを見る

掃除中、思ったこと感じたことを発言してもらう

- 5人の発言の違いを見る





事故事例データの入力項目

表1 主な調査項目

番号	ISS調査項目	番号	ISS調査項目
1	受診年月日	9	怪我の種類
2	生年月日	10	怪我の部位
3	発達段階	11	事故に関係したモノ
4	体重	12	事故が起きた場所
5	身長	13	事故直前の当事者の行動
6	事故発生日	14	一緒にいた人の事故直前の行動
7	事故発生時間	15	事故の経過
8	事故の種類	16	事故の詳細

事故状況を記述したテキストデータ

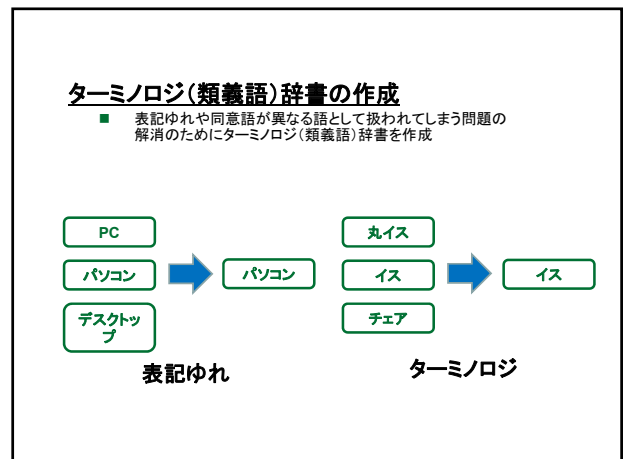
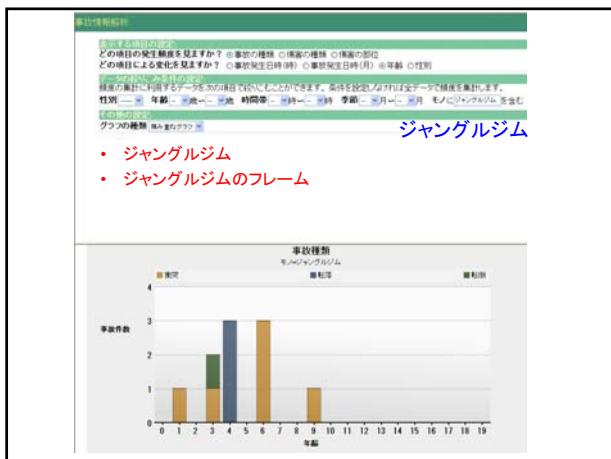
表2 収集されたテキスト形式の事故データ(一部抜粋)

No.	事故の詳細
1	つかまり立ちして机の角に下腹をぶつけた。
2	車に乗ろうとして転んでドアの角に顔をぶつける。
...	...

[illegible][illegible][illegible][illegible]

行ID	文章ID	単語ID	見出し語	原形	品詞	係り先
185	479	2779	風呂に	風呂	名詞	2780
185	479	2780	入って、	入る	動詞	2783
185	479	2781	洗いで	洗う	動詞	2782
185	479	2782	で	着る	名詞	2783
185	479	2783	靴履	靴履	名詞	-1
185	480	2784	浴槽の	浴槽	名詞	2785
185	480	2785	角に	角	名詞	2787
185	480	2786	顔面を	顔面	名詞	2787
185	480	2787	ぶつけた	ぶつける	動詞	-1
185	481	2788	自分の	自分	名詞	2789
185	481	2789	面で	面	名詞	2792
185	481	2790	尊の	尊	名詞	2791
185	481	2791	下	下	名詞	2792
185	481	2792	切れた、	切れる	動詞	-1





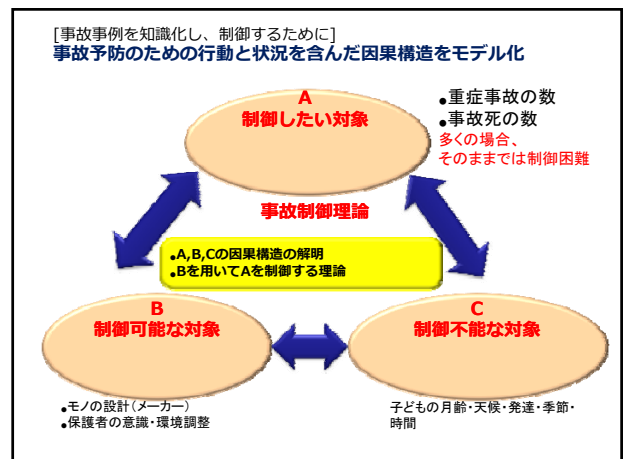
モノに関する辞書の作成

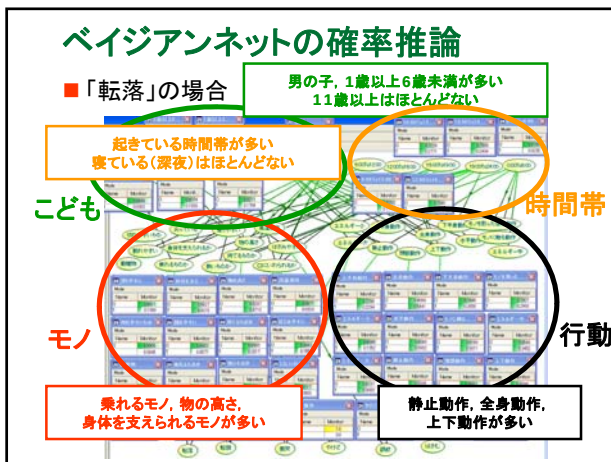
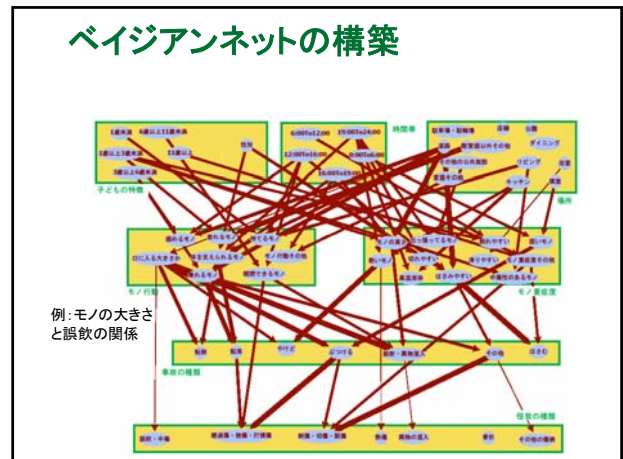
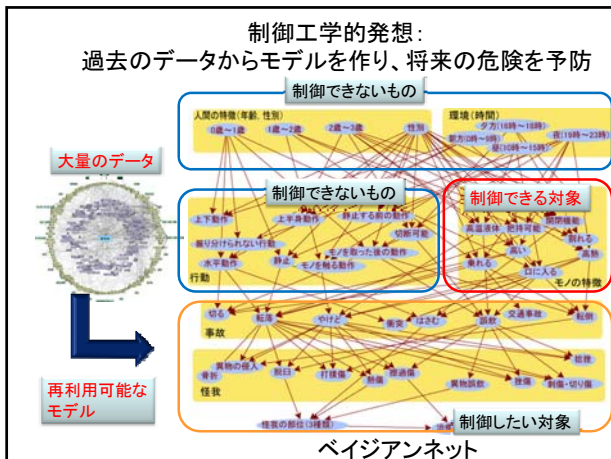
- モノに関して、表記ゆれ辞書、ターミノジ辞書を作成
- モノの特徴を示す、高さ、材質、使用頻度も導入
- モノの表記ゆれ辞書
 - 表記ゆれを解消するために作成した辞書
 - 単語数: 1421
 - 項目数: 590に集約

見出し語	品詞	単語群
イス	名詞	椅子
チェア	名詞	チェア
...	...	...

- モノのターミノジ辞書
 - さらに493項目に集約

見出し語	品詞	単語群
イス	名詞	回転イス
自動車	名詞	乗用車
...	...	...





AIST

成功体験の想起を促す電子日記帳

産業技術総合研究所
本村陽一
中島正人
三浦未央

Digital Human Research Center

AIST

使用データ

国リハ: 自己高揚経験データ

記述内容
内容
「あの時、あの場所で、あの行動をとった自分はえらかったな」というような「自分をほめてあげたい」と思える過去の経験
思い出した順番に3つ記述
ラベル
経験を2つに分類
達成経験
自分の力で何かを達成した経験
自分で成し遂げたことを人から評価された経験など
親和経験
他者との共生に関する経験

行数: 154
単語数: 1438

Digital Human Research Center

AIST

TMSによる評判分析

- 構文解析→評判分析
 - 肯定的であると判断された事例
 - 親としての強い責任感のもと、4人の子育てに専念し、仲の良い家庭を作りつづくと確信できる自分
 - 暑期中練習を頑張り、短距離でよいタイムが出せたこと
 - 否定的であると判断された事例
 - 小さい頃、自分のやりたいことを我慢して、兄弟の面倒を見たこと
 - 25〜30歳の頃、ハードな仕事をこなしたこと
 - 母が亡くなった時、心がぼろぼろだったけれど、葬儀のことを色々切り盛りし、父と兄の悲しくてやり場がなく当り散らした感情を冷静に受け止め、前向きになるよう言葉をかけることができたことは、本当に今までの人生ですごいことだったと思う

Digital Human Research Center

TMSによる評判分析のまとめ

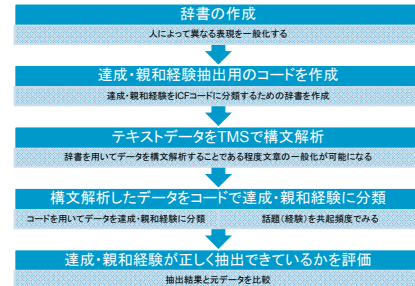
問題点

- 認識率が不十分
 1. 人により同じ意味でも表記が異なる
 2. 文章に応じたアルゴリズムの修正ができない
 3. 親和経験と達成経験を分けて抽出することはできない

アプローチ

1. 表現を一般化するための辞書を作成する
2. 独自の文書分類アルゴリズムの作成
3. 生活機能分類コード(ICF)を用いて達成・親和経験の識別を試みる

実験1：辞書とICFコード化→TMS評判分析



構文解析の出力例

構文解析結果

行ID	文章ID	単語ID	見出し語	原形	置換語	品詞	品詞詳細	係り先	述語属性
1	1	1	34歳の頃、	34歳	34歳	名詞	数	13	なし
1	1	2	家事に	家事	家事	名詞	一般	3	なし
1	1	3	協力的でない	協力的	協力的	名詞	サ変接続	4	否定
1	1	4	夫に	夫	e310	名詞	一般	6	なし
1	1	5	不満を	不満	不満	名詞	一般	6	なし
1	1	6	持つ	持つ	持つ	その他		7	なし
1	1	7	自分を	自分	自分	名詞	一般	8	なし
1	1	8	変えたい	変える	変える	その他		9	要望
1	1	9	一心で、	一心	一心	名詞	一般	13	なし
1	1	10	倫理を	倫理	倫理	名詞	一般	11	なし
1	1	11	学び出し、	学ぶ	学ぶ	動詞	自立	13	なし
1	1	12	実践を	実践	実践	名詞	サ変接続	13	なし
1	1	13	継続できたこと	継続	継続	名詞	サ変接続	-1	可能
2	2	14	家業が	家業	家業	名詞	一般	17	なし
2	2	15	宿屋で	宿屋	宿屋	名詞	一般	17	なし
2	2	16	小さい頃から	小さい	小さい	形容詞	自立	17	なし

生活機能分類(ICF)コードと辞書

達成・親和経験をICFコードに分類するための辞書

達成	d1551 複雑な技能の習得 d1750 単純な問題の解決 d1751 複雑な問題の解決 d177 意思決定 d2201 複数課題の達成 d2302 目標の達成 d2400 責任への対応 d2402 危機への対応	d1551 免許 取得 できる d1750 目標 達成 できる d1751 目標 達成 できる d177 自分 意志 取り込む d2201 複数 出来る d2302 地味 出来る d2400 業務 やり運ぶ d2402 ストレス 克服
親和	d6600 他者へのセルフケアの援助 d6601 他者への移動への援助 d6603 他者への対人関係への援助 d7100 対人関係における敬意と思いやり d7101 対人関係における感謝 d7102 対人関係における寛容さ	d6600 家族 世話 d6601 救急車 手配 d6603 地域密着化 貢献 d7100 相手 立場 理解 d7101 家族 協力 感謝 d7102 家族 受け入れる

ICFコード

達成・親和経験となる単語をICFコード化した辞書

述語属性は「否定」の場合は無視
述語属性は「可能」の場合は出力

実験1：ナイーブベイズ(子ノードが一つのBN)によるテキストデータの分類

結果1(達成経験)

- 元データと抽出結果が一致した事例
 - ・ 毎日神様にお参りすることが出来た(25年間)
 - ・ 35年間、大病もなく、ひとつの会社に勤められ、退職後第二の人生に向かって働いていること
 - ・ 結婚して3人の子供を産んだこと
 - ・ 小学生からスポーツを継続した
- 元データと抽出結果が不一致だった事例
 - ・ 昭和50年、わずかにばかりの貯金を元手に我が家を新築できたこと。
 - ・ 10年前、自分の仕事で、通常ならば三年以上かかるであろうといわれていた仕事を毎晩遅くまで残業し、一年程度で成果が挙げられたこと

結果2(親和経験)

- 元データと抽出結果が一致した事例
 - ・ 近所の人が喘息の発作で苦しんでいるところを発見し、救急車の手配をしてあげたこと。
 - ・ 車で移動中、歩道で自転車に乗っていた人が倒れ、歩けなくなっていたので、車に乗せ、自宅に送っていったこと
 - ・ 姑とあまり仲がよくなかったが、姑が病気になるまで寝たきりになったとき、家で面倒を見て、最後を見取ったこと
- 元データと抽出結果が不一致だった事例
 - ・ 平成6年～8年頃、兄弟の経済的危機に直面した時、自分のことを忘れて協力できたこと
 - ・ 散歩中に、道端に老婆が倒れていたの、家まで連れて行った。家の人も探しており、感謝された。
 - ・ 地区の自治会長と連合会長を引き受けたこと

実験1: 結果識別精度

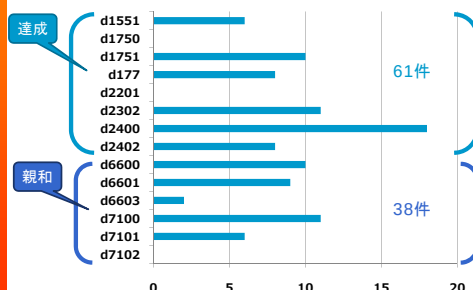
評判分析結果との比較

	正解	間違い
評判分析	32	125
コードを用いた共起頻度による抽出	92	62

ICFコードに変換することで識別精度が向上

付録:ICFヒストグラム

国リハ自己高揚経験リストからの抽出が多かったICFコード



発展：日常場面成功体験日記システム



日常場面成功体験日記システム

- 場面を限定し、その日1日の成功体験を聞き取り、入力
 - ・ 日常で成功した(良かった)場面の体験を語ってもらう。
 - ・ 入力は別の人が行う。
 - ・ 場面は複数あらかじめ設定しておく
 - ex) 家族、家事、仕事、友達、買い物、など
 - ・ 場面は利用者が選択する
 - ←カテゴリー化

《メリット》

- ・想起、記述の容易性 → 辞書作成の負担減(カテゴリー化も容易)
- ・反復性あり
- ・日記の記録のモチベーションアップ(コミュニケーションの促進)
- ・入力後に簡易な評価とフィードバックを行うことも可能

行動から心理へ：
認知・評価構造へのアプローチ

Personal construct theory (G.A.Kelly 1955)

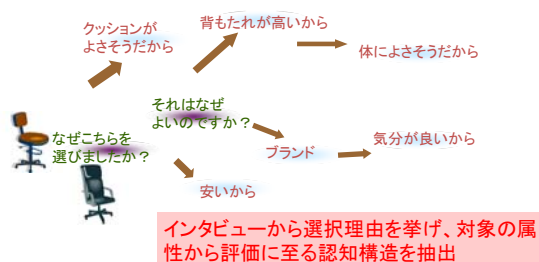
個人構成体:小橋1988「決定を支援する」認知科学選書(東大出版)

- 個人構成体: 事象を二文法的にカテゴリ化する個人の評価構造モデル
- 公理: 人間の諸過程は、事象を予期する様々な仕方によって心理的に経路づけられている。
- これを基本公理として、さらに11の系に従う。

評価グリッド法(讃井 1986)

個人の認知・評価構造を明らかにする調査法

- ・「なぜ」を繰り返すインタビューによる気づき
- ・ 無意識的な行動の背景にある動機が明らかに



運転中の走行シーンの認知構造

(画像提供:トヨタ自動車)

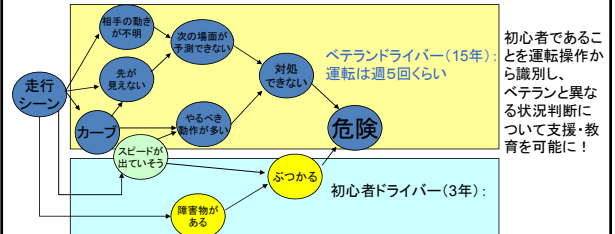


どちらが危険と判断するか？
より危険と認識するのは何が原因か(コンストラクト)?
それはなぜか(理由・原因)?
もしそうであれば、どんなことが起こるか(因果的結果)?

危険性の評価・認知構造

走行シーンの認知・評価の因果的階層構造には下図のような個人差がある。

- ・運転者の初心者／熟練者の識別
- ・熟練者:危険 初心者:安全 となるシーンについて
初心者に危険であることを伝える支援システムへ



大規模データモデリング (理解→予測→制御)

従来の統計

- ・各事例の保存と検索(データベース)
- ・単独の頻度、集計、統計
- ・複数の変数の間の関係(相関、独立性)
- ・ある変数を説明するルール、予測モデル
- ・変数群の依存関係の条件付確率(確率的予測)
- ・因果的構造のネットワーク(構造的理解)

ベイジアンネット

集めたデータの説明だけでなく、将来の**予測**と
望ましい状態を実現する**制御**へ！

日常生活支援のために

生活を知る→生活行動を制御

行動分析学

環境変化と行動変化を観察

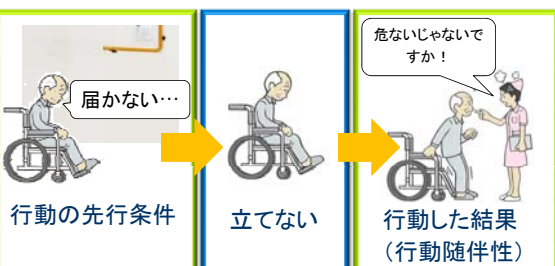
↓
行動と環境の関係を見出す

↓
その関係を使って行動を制御

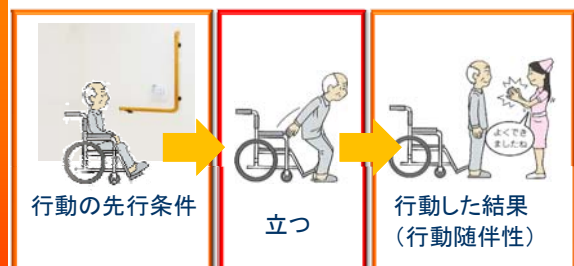


Skinner 1904~1990

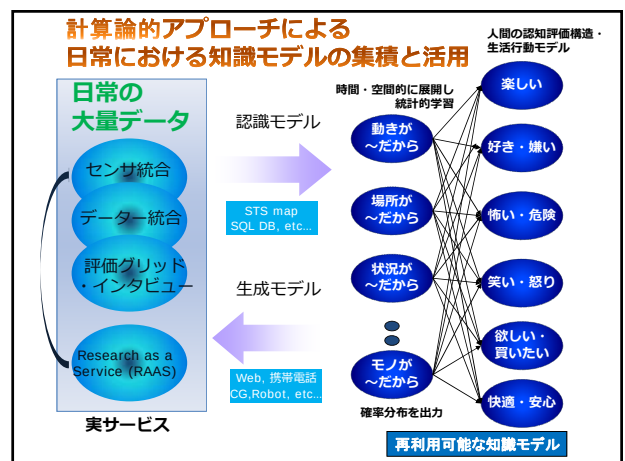
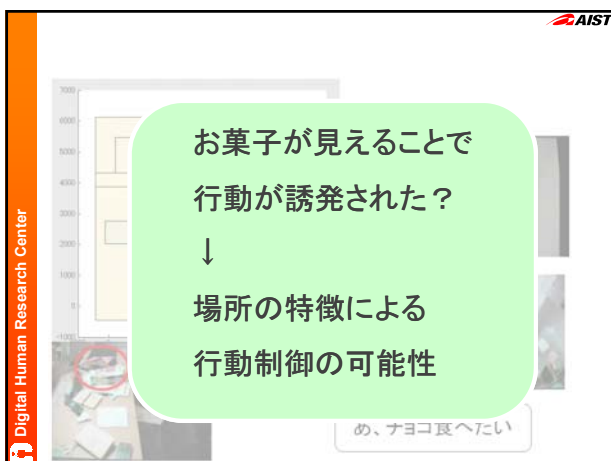
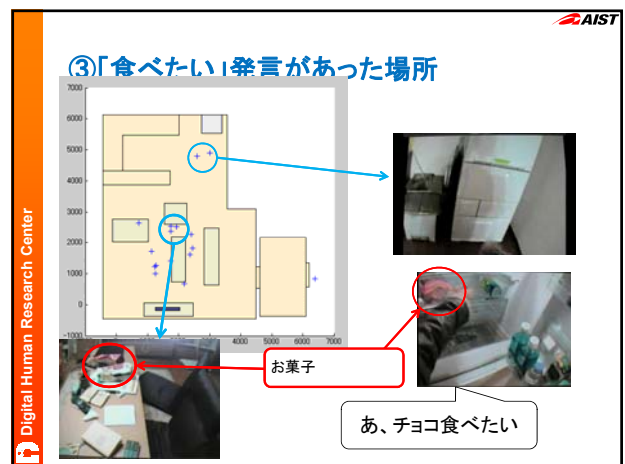
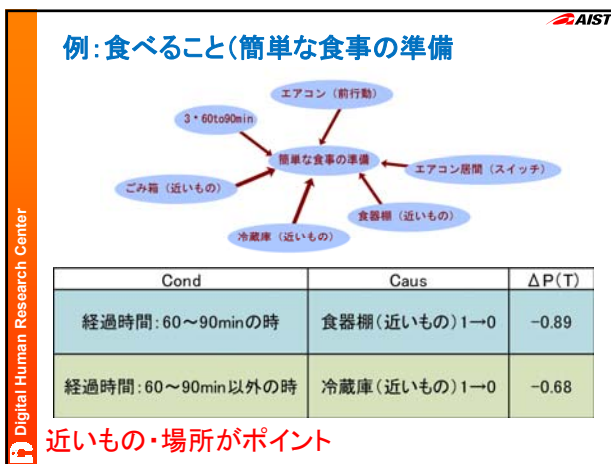
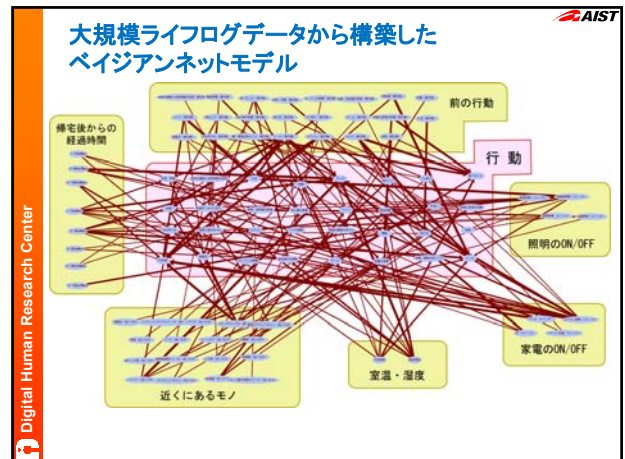
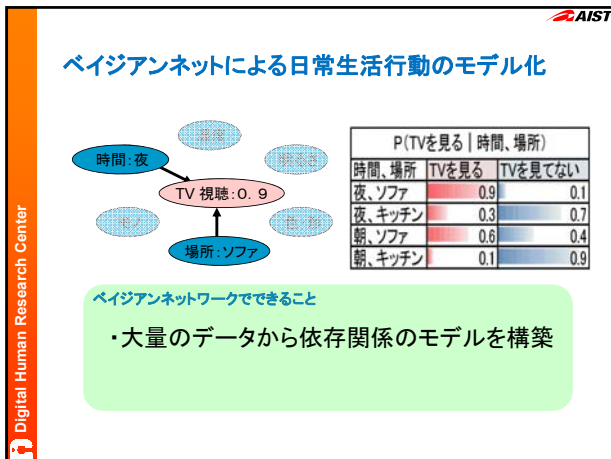
Skinner, The Behavior of Organisms, 1938

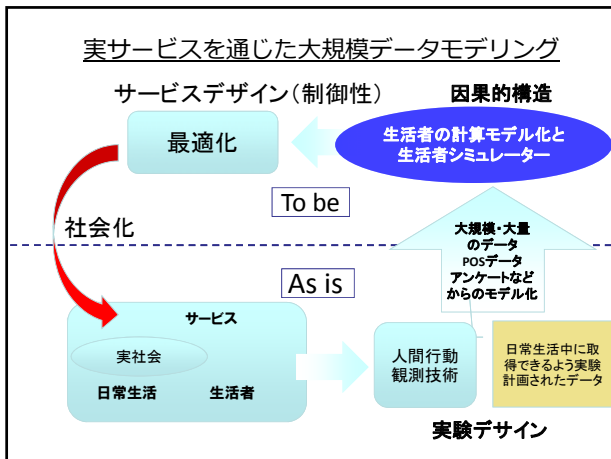


環境により行動が抑制される



環境を修正することで
「立てる」状態を実現する





日常生活の科学の方法論

- ・ マイケルギボンズ: モード論
- ・ 「モード1」: 伝統的なアカデミズムの世界の研究様式
- ・ 「モード2」: 現実社会の具体的な問題を解決することが研究の目的であり、学問領域を越えた連帯が重視される。
この基礎と応用という二元論で対処することが難しい!

そこで、現実問題を解決する「モード2」から出発し、「モード1」の世界で洗練させ、さらに「モード2」へ還元する1と2をつなぐ研究が必要!!

**研究と応用が切り離されることなく、相互作用する。
問題解決型の基礎研究の必要性**

