

Over the AI ― AIの向こう側に(6):

時をかける人工知能 ～たった1つの数値で結果から原因に遡る

<http://eetimes.jp/ee/articles/1612/28/news014.html>

「ベイジアンネットワーク」は、私が最も使い倒している人工知能技術の1つです。ある事柄について、たった1つしか信頼に値する数字がなくても、「現在の結果」から「過去の原因」を、遡(さかのぼ)って推測できる。そんな技術なのです。

2016年12月28日 09時30分 更新

[江端智一, EE Times Japan]



今、ちまたをにぎわせているAI(人工知能)。しかしAIは、特に新しい話題ではなく、何十年も前から隆盛と衰退を繰り返してきたテーマなのです。にもかかわらず、その実態は曖昧なまま……。本連載では、AIの栄枯盛衰を見てきた著者が、AIについてたっぴりと検証していきます。果たして”AIの彼方(かなた)”には、中堅主任研究員が夢見るような”知能”があるのでしょうか――。[⇒連載バックナンバー](#)

14時間の計算が1分以下に、プロからの助言

[前回](#)、合衆国大統領選挙の簡単なモデルを作ってシミュレーションを行い、対立する2人の立候補者の支持率が拮抗している時、獲得選挙人数に大きな差が発生することを示しました*)。

*) 合衆国大統領選挙の「選挙人制度」については、[前回の記事](#)の、図「合衆国大統領選挙の方式」を参照のこと。

具体的には、対立する2人の立候補者に、米国の全州の国民が投票率100%で全員ランダムに投票するという仮想の簡易モデルにおいて、

(1) 合衆国大統領選挙に独特の「投票人制度」を使うと、実際の投票数に対して、その投票数に比例しない選挙人数を弾き出す可能性が極めて高いこと ― 具体的には、分散(または標準偏差)が驚くほど大きい(全選挙人数 538人中、平均 $M=269$ 人に対して、標準偏差 $\sigma=51$ 人)

(2) 投票比率のわずかな変化に対して、選挙人数が劇的に変化する(投票数の0.01%の変動に対して、平均選挙人獲得数が121人変動)

というシミュレーション結果を示しました。

私は、このシミュレーションを、乱数を用いて1人1人投票させる方法(モンテカルロ方式)を1万回繰り返したので、約5万秒、14時間近くの時間を費してしまったのですが、この記事を読んだくださった、数値分析のチームのデータエンジニアの[gauraさん](#)から、高速計算方式(大統領選のシミュレーション(in R)の[詳細解説](#))をご教示いただきました。

ご開示頂いた[C++のプログラム](#)を、私の自宅のPCで実行^{*)}してみたら、実行時間はたったの59秒にまで短縮され、私のやり方の約0.2%の時間で、シミュレーション結果を出力できることが分かりました。

*) パフォーマンスを正確に評価するため、標準入出力(`printf()`など)は全てコメントアウトしました。

帰宅途中の電車で、gauraさんからメールで送って頂いた解法の解説をガラケーで読んでいたのですが、「あっ! そうか!!」と思わず声を上げてしまい、車内で恥ずかしい思いをしました。そして、『さすがは、プロのデータエンジニア』と感嘆したのです。

計算時間が0.2%まで劇的に短縮できるということは、それだけ検討範囲を広げることができることになり、さらにバラエティに飛んだ結果を得ることができることになります。この世界でも、「高速化は無条件に正義」が成り立っていることは明白です([参考:「カップ1つ作るのに、ご主人様とメイドは4000回会話する」](#))。

それはさておき、Twitterなどのコメントを読んでいると、(私が開示したプログラムを試すこともせずに)前回の私のシミュレーション結果を、間違いであると決めつけている人もいたようです(少数ですが)。

しかし、それぞれ異なるアプローチで導き出したgauraさんと私のシミュレーションは、同じ結果となりましたので、

—— まだ、文句があるなら、かかってきやがれ

という気分になっています*)。

*)ただ、「乱数を全く使わない定式化」の検討が残っていると思っています。私が検討すると、今回のシミュレーションの時間(約14時間)以上の時間を持っていかれた揚げ句、失敗に終わるような気がしますので、ここは数学や統計学の研究者の方からの数理モデルのご教示をお待ち致します。

□

米国大統領選挙の「選挙人制度」では、州ごとに選挙人の全てが加算される操作で、選挙人

数の離散化を発生させるので、選挙人の数にある程度の分散や感度を生み出すことは、最初から分かっていました。ただ、私には、その数値の規模感が分からなかったのです。

ですから、私は前回のコラムの執筆前に、この選挙の分散と感度に言及している文献をネットで探しまくりました。

最初は、『どうせ、統計学の入門書の教科書の例題として出てくるだろう』と軽い気持ちで探し始めたのですが、どうしても見つけ出すことができませんでした。やむなく簡単なシミュレーションプログラム(各州の人口と選挙人の数を調べる時間を除けば、作成時間は30分弱)を作り、力ずくのシミュレーションを実施し、そして、その選出される選挙人の数の分散と感度の数値にビックリしたのです。

Twitterのコメントを読んでいると、シミュレーションも定式化も行うことなく、「そんなことは、全然ビックリするようなことではない」と豪語し、私のコラムをけなしている人もいます*)。

*) 前述のgauraさんや、後述の持論を展開して頂いた12人の皆さんほど理路整然と説明して頂く必要はないので、[これ](#)と、[これ](#)をご一読いただき、あなたが言いたいことを全て、このメールアドレス("over_the_ai@kobore.net")に送ってください。どんな内容であれ、そのメールの内容の完全な全文公開を約束します。

しかし、私は、私と同様にビックリしている人は、かなり多いと思っています。例えば、今回、米国在住の、シリコンバレーでITアナリストをされている方から、以下の文面のメールを頂きました(さらに、講演のご依頼まで頂きました)。

先日発表された「沈黙する人工知能 ～なぜAIは米大統領選の予測に使われなかったのか」を拝見しました。こちらで10回大統領選挙を見ておりますが、日本の解説者も解説記事も的を射たものがなく、選挙人選挙であることを全く理解していないものがほとんどでした。江端様の記事は選挙制度を解説したうえ、データ分析でこの結果を解説なさっており、正にこれだともおっしゃいました。

(ご許諾を頂きご本人のメールから引用。太字の強調は江端)

正直なところ、私も前回のコラムを執筆するまで(というか、シミュレーションを強行するまで)は、この大統領選の「選挙人制度」について、十分な理解ができていなかったと思います。

大統領選の「優れた点」は？

私は、前回のコラムで、米国大統領の「選挙人制度」を、批判的な観点から論じていました。しかし、gauraさんからは、前述の超高速シミュレーションの方式解説に加えて、合衆国大統領の仕組みの優れた点についての、かなり長文のご意見を頂きました。

このメールを読んだ時、正直、頭をぶん殴られたようなショックを受けました。

「合衆国大統領選挙制度の優れた点」――その観点、イける。

米国大統領選挙の欠点については、既に前回の連載で、[私が言及した内容](#)で足りていると思いますが、これだけの欠点がありながら、(歴史的経緯があるとはいえ)、米国民がこの選挙制度の修正を試みようとならないのは、もっと深い合理的な理由があるはず ―― そういう観点で、この選挙制度を見直してみることにしました。

そこで今回、私は、もう1つの連載「[世界を「数字」で回してみよう](#)」でアンケートに参加を申し込んでくださっている方に、無理を承知の上で、以下のメールの文面を送付し、ご協力を要請しました。

江端は、"Over the AIの第5回"にて、米国大統領選挙の選挙人制度が、民意を反映しにくいシステムであると指摘しました。これに対して、この制度が「優れている」という観点から、皆さんの論をご展開して頂ければ幸いと存じます。ご協力の程、なにとぞよろしくお願い致します。

頂いた、12人の皆さんの持論(仮説)の全部が[こちら](#)になります。長文ですが、いずれも素晴らしい説なので、ぜひご一読ください。

さて、皆さんの「合衆国大統領選挙がなぜ優れているか」のご持論(仮説)は、大きく以下の3つのカテゴリーに分けられるようです。

(その1) 多民族、多宗教、かつ国家の合衆体(合衆国)であるという面倒な問題に対応しうる選挙制度であるから

大統領選が優れている理由(仮説) その1

江端のハートにヒットした仮説を要約

お名前	大統領選の効果	通常選挙なら・・・
gauraさん Aさん	政治的に少数派意見をくみ取ることができる	母集団の大きい集団の意見だけが重要になる
	地政学的に安定し、地域的に小人口地域の意見を反映することができる	激戦になるほど、有権者数の多い州に「媚びる」政策が打ち出される→結果、大統領選の議論が行われない地域が生じる→大統領の正当性に疑問が生じる

多民族、多宗教、州間の人口比率の問題を、できるだけ回避する選挙制度であるから

確かに、私にしても、宗教間や民族間の対立というのは字面では理解できても、肚の底から

理解できているかと問われれば ―― 正直、さっぱり訳が分かりません。

私は ―― 正月には神社または寺社(この2つは、全く性質の異なるモノですが)に行き、お盆には先祖の霊をお迎え/お見送りし、10月には「トリック オア トリート」と叫びながら仮装に興じ、クリスマスイブにはミサに行き*)、死んだ時はお坊さんに遺体の処理手を委託する ―― という典型的な日本人です。

*)「[英会話スクール出逢い機関論](#)」「[クリスマス脅迫症候群](#)」

もし私が、イスラム原理主義者のテロリストに銃を突き付けられて、「イスラムに改宗しろ!」と言われたら、1秒もたたずに「いいよ」と言いますし、1分後にはコーランの暗記を始めることもできます(というか、銃を突きつけられてまで「改宗に同意しない」って感じが理解できない。ぶっちゃけ、言うだけならタダじゃん、と思っています)。

このように、宗教的観念が空気のように希薄なわれわれ日本人の在り方は、宗教を根拠とした戦争や虐殺や殺りくの無意味さを、世界に対して雄弁に語るはずなのですが、多分、多民族、多宗教国家である米国では、そうも言っていられないのでしょう。

それに、合衆国の各州間の格差は、もっとシビアです。なにせカリフォルニア州の選挙人55人なのに対して、アラスカ州ではたったの3人です。アラスカ州の意見を大統領選にダイレクトに反映させるのは絶望的と思えます。

それなら、嫌がらせくらいはできる仕組みは必要でしょう。アラスカ州の人々は「俺たちを蔑(ないがし)ろにしていると、当落のボーダーラインで痛い目に遭うぜ」くらいは言いたいはずですよ。

“変なヤツ”を事前に排除できる？

(その2) 長期的に変なヤツを排除しつつ、自分の意に沿わぬヤツが当選しても、そこにリーダーシップを発揮できるようにできているから

大統領選が優れている理由(仮説) その2

江端のハートにヒットした仮説を要約

お名前	大統領選の効果	通常の選挙なら・・・
Dさん Yattanさん	カルト集団から 大統領は生まれ にくい	党内選挙を経ないので、変なヤツ(ヒトラーみたいなヤツ)を事前に排除できない
Eさん	選挙での「圧勝」という幻想を作れる	50：50の投票率では、民衆の納得が得られず、リーダシップを発揮できない(が、選挙人制度では「圧勝」が作れる)
Fさん	無関心な投票を排除できる	いい加減な気持ちで投票しても、そこそこの結果になる(が、選挙人制度では、シビアな結果となりやすい)

変なヤツを事前に排除しつつ、嫌なヤツが当選しても「仕方がない」と思える制度だから

合衆国大統領の予備選挙では、約1年間という長い時間をかけて、党内選挙を繰り返すことで、思想的なエクストリームな奴(極左とか極右とかカルトとか)を排除できます。

また、もし最終投票数が拮抗している場合であっても(まさに、今回の大統領選挙がこのケース)、選挙人の数に圧倒的な差を作り出すことで、「圧勝」を演出することが可能となり、国民に「仕方ねーな」という気持ち作り出すことができることの効果は、結構大きいと思います(例えば、カリフォルニア州なんて、その気になれば、合衆国から離脱して、独立国家になれそうな規模がありますしね)。

それと、日本でいう「1票の格差」という意味とは全く違う意味で、「1票の重み」が違うことも大きそうです。前回のシミュレーションで示した通り、たった1万人に1人の心変わりで、選挙の結果がガラッと変わってしまうことは、有権者に対して、強い投票へのモチベーションを発揮させることができるはずで(と言いつつ、今回の選挙の投票率は半分(50%)を割っていたようですが)。

米国民の気質にマッチしている？

(3) 米国民の気質(変化を求める気質、その場のノリを重視する気質)を反映させることができるから

大統領選が優れている理由(仮説) その3

江端のハートにヒットした仮説を要約

お名前	大統領選の効果	通常の選挙なら・・・
Gさん 江端	変化を起こしやすいシステムである	利権が絡んで、政治が硬直化しやすい(が、選挙人制度では、コロっと変わること、変化をうながしやすい)
Yattanさん	本当の意味で「未来を選ぶ」一票になり得る	支持のない候補者が当選する確率は絶無(が、選挙人制度では、その可能性が相当に高い)
Iさん	「勢い」を選挙に反映させやすい	「その場のノリ」で当選する可能性が排除される(が、選挙人制度では、1年間の党内選挙によってノリ(お祭り)が、システムとして組み込まれている)

「未来を選択する一票」が本当にリアルで、支持率以外の要素も組み込まれるから

私たち日本人は基本的に「変化」が嫌いな民族です。ティーン代は、クラス替えや席替えですら憂鬱(ゆううつ)でしたし([参考ブログ](#))、社会人になっても、自分が所属する部署の安寧のために奔走しています。

一方、米国は――米国在住の読者の方(Cさん)から頂いたご意見によれば――『一発当たりたい人たちが多数を占める国民』なのだそうで(実は私もひそかにそう思っている)、「変化」を望ましいものとする人々が多い国として、把握できそうです。



このような「変化を求める気質」は、遺伝子の突然変異のように、その遺伝子を有する個体に劇的な改善をもたらす場合もある*)のです(多くの場合は失敗しますが)。

*)このような遺伝子の仕組みを利用した高速最適解探索手法に、「遺伝的アルゴリズム(Genetic Algorithm:GA)」があります。次回以降にご紹介致します

ところが、米国は、大統領の交代に伴う経済政策によって、劇的に国内経済が改善された実績を持つ国家でもあります(ニューディール政策とかレーガノミックスなど)。つまり大統領選挙は、米国全土を巨大な競馬場と見立てた、4年に一度だけ開催される、「天皇賞」や「菊花賞」みたいなものです。

しかも、米国大統領選挙という「天皇賞」や「菊花賞」は、1年前から地方競馬場で選抜が始められているので(党内の予備選挙のこと)、いやがおうにも気分は盛り上がってきます。そして、4年に一度のギャンブルであれば、「万馬券」を狙っていくのが人情というものでしょう。

そして、単なるその人の過去の実績だけではなく、自分の未来の夢の国までも立候補者に託すあたりなど ―― 私は『心の底から立候補者に未来を託す』などという発想を一度も持ったことはありませんが ―― さすがは「夢と魔法の王国」などという、(日本人であれば恥しくてとても口にできそうもない)コンセプトを伴った近代アミューズメントビジネス発祥の地の、面目躍如とも言えるべきでしょう。

以上、総括しますと、

—— 合衆国大統領選挙(の選挙人制度)は、あの国の国民と、いい感じでマッチングしている

と言えそうです。

いずれにしろ、この合衆国選挙と、この連載のテーマである人工知能との関係においては、

(1) 対立する2大候補の支持率が拮抗している状態では、大統領選挙の当落予測は、次に出るサイコロの目を予測するようなもので、人工知能"技術"はもちろん、いかなる技術を用いても困難である(これは前回の結論)、

という事実に加えて、

(2) その未来不確定で、私には不合理にも見える選挙システムを、数値化困難なファクター(変化とかノリとか、薔薇色の未来予想図とか)を属性として有するユーザー(米国国民)自身が、今なお、この選挙制度を支持し運用を続けている(これは今回の結論)

という事実が加わって、相乗的に「予測不能ポテンシャル」が極大化し、まあ、今後1000年先も、合衆国大統領選挙は、"人工知能"の予測の対象外としてもいいのだろうなー、と私は考えています。

ベイジアンネットワークとは何か

さて、ここからは、前回の合衆国大統領選挙のシミュレーションで1回分スキップした、[前々回](#)の「ベイズ推定」の後半、「ベイジアンネットワーク」*)の解説をさせて頂きたいと思います。

*)「ベイジアンネットワーク」「ベイズネット」の2つの呼び方があるようですが、Google裁定(検索件数1万1900件 v.s. 641件)によって、今回から「ベイジアンネットワーク」で統一することにします。

なお、本当に毎回しつこいとは思いますが、「『ベイジアンネットワーク』が“人工知能”なのかどうか」については、今回も『[江端AIドクトリン](#)』に基づいて私が勝手に判定しました。

まず、「ベイズ推定」についてのザックリとしたおさらいをしてみたいと思います。

ベイズ推定とは何か(その1)

■江端は「難しい話」はできません

→というか、よく分かりません。興味もありません。
「道具」として使えればいいんです

■ある物事の原因とその結果が分かっている時、 結果からその原因に遡ることができる推定

例: 「合コンすれば、結婚できる」というデータに対して、
「結婚した人が、合コンを月1回以上でしていた確率は？」を求めることができる

「金を払ってでも知りたい」情報に変わる

というものです。詳しくは、[こちら](#)をご覧ください。

ベイジアンネットワークとは、条件付き確率の計算を行う仕組みなのですが、普通の確率手法と異なる点は、この因果関係を図で記述(設計)できる点に特徴があります。以下の図はその一例です。

ベジアンネットワークとは何か(設計)

例1:「浮気をすると、パンティが発見される」



例2:「風が吹けば、桶屋が儲かる」

- (Step.1) 大風で土ほこりが立つ
- (Step.2) 土ほこりが目に入って、盲人が増える
- (Step.3) 盲人は三味線を買う(当時の盲人が就ける職に由来)
- (Step.4) 三味線に使う猫皮が必要になり、ネコが殺される
- (Step.5) ネコが減ればネズミが増える
- (Step.6) ネズミは桶をかじる
- (Step.7) 桶の需要が増え桶屋が儲かる

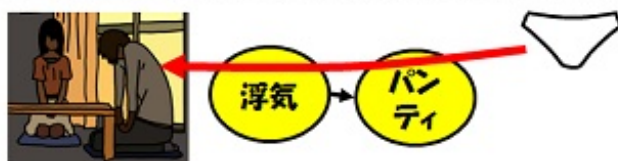


「因果の連鎖」をネットワークで表現

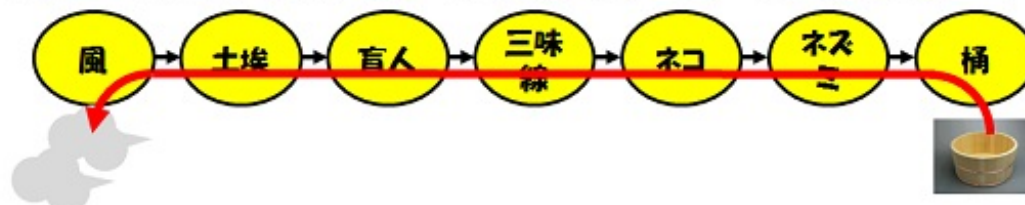
そして、ベジアンネットワークは、この因果の関係を、結果から原因に遡れる点に特徴があります。

ベジアンネットワークとは何か(推定)

例1:「パンティ発見時に、浮気をしてる(確率は)?」



例2:「桶屋が儲かった時、風が吹いている(確率は)?」



「因果の連鎖」ネットワークを使って、結果から原因へ遡る

ベイズアンネットワークの設計方法

さて、次に、実際のベイズアンネットワークの設計方法を御説明致します。

今回は、これまでも何度が登場したNTT数理データ数理システムの「[BAYONET](#)」を使って説明しますが、ベイズアンネットワークのツールの使い方は、おおむねどれも同じ様です。

説明の題材は、[第4回](#)の時にも使ったネイト・シルバーさんの著書「シグナル&ノイズ」の「パンティ問題」の話をそのまま使います。

ベイズ推定を「パンティ」で理解する

■出展:「シグナル&ノイズ」ネイト・シルバー著

■事案

出張から帰宅したあなた(女性)は、
タンスの中に自分のものでない
パンティを発見した



これが「条件」

**パンティ発見時に、パートナー(男性)が
浮気をしている確率は？**

これが「条件付き確率」

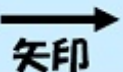
「単なる確率」ではない

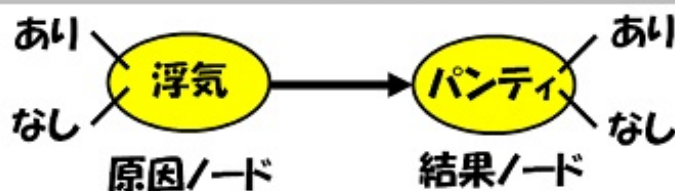
ベイズアンネットワークの設計は、ひと言で言えば、原因(浮気)のノードと、結果(パンティ)のノードを作って、原因から結果に向けて矢印を引くだけです。

ベイジアンネットワークの作り方(その1)

 と  だけを使います

■Step.1:  に、名前と状態(例:あり/なし)を指定して下さい

■Step.2:  の、始点を「原因ノード(例:浮気)」
終点を「結果ノード(例:パンティ)」
として、繋いで下さい



構成要素は「ノード」と「矢印」の2つだけ

ベイジアンネットワークの設計で、最も混乱しやすいことが、何が原因で、何が結果なのかが、設計している自分自身でも分からなくなる点にあります。

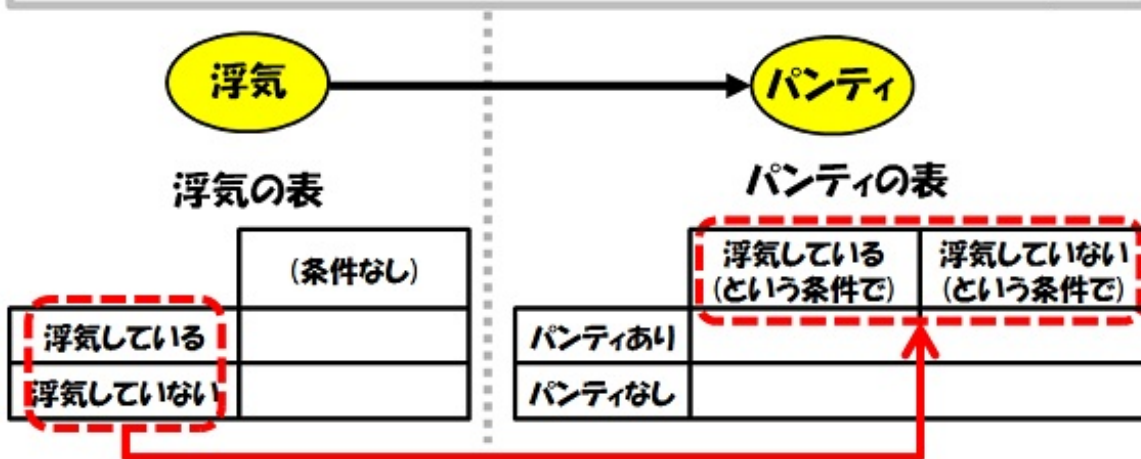
では、どう考えれば良いかというと、最初にノードだけ作っておいて、矢印については「後から考えればいい」のです。

具体的には、「IF(浮気), THEN(パンティ)」の関係と、「(浮気)と(パンティ)の確率」が分かっているならば、(浮気)から(パンティ)に矢印を引くことで、ベイジアンネットワークの設計は完成です。

つまり、ベイジアンネットワークは、この矢印の逆方向、つまり、"THEN"から"IF"に向けて解くものとだけ覚えておいていただければ十分です。

ベイジアンネットワークの作り方(その2)

■Step.3: ノードを矢印で繋ぐと「表(CPT)」が自動生成されます CPT:Condition Probability Table



原因ノードの「状態」が結果ノードの「条件」

ノードを作ると、上記のような表が自動生成されます。この表のことをCPT (Condition Probability Table (条件付き表)と呼びます (名前なんかどうでもいいですが)。

ただ、ここでは、浮気 (条件) のCPTが"2×1"の表になるのに対して、パンティ (結果) のCPTが"2×2"の表になることだけ覚えておいてください。

めんどくさくても大事な数値

次に、その表に分かれている数値をたたき込みます (この数値については、[このページ](#)の「パンティ問題に必要な情報」を参照してください)

ベジアンネットワークの作り方(その3)

■Step.4: 「表(CPT)」に、確率の数値を入力します



CPTだけは真面目に入力する必要があります

面倒ですが、この数値だけは真面目に入力する必要があります。

既にこの段階で「面倒くせーなー」と思っている人がほとんどだと思います(ベジアンネットワークの挫折は、おおむねCPTの作成から始まる)が、実際にBAYONETを使う場合は、CPTの入力を手動で行うケースなどほとんどありませんので安心してください(私も、CPTを作成したのは、この「パンティ問題」の時だけです)。

ともあれ、これでベジアンネットワークの設計は完了です。

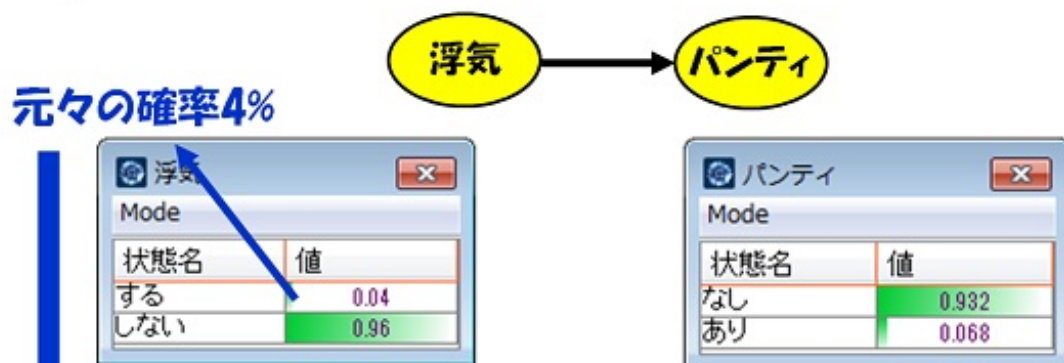
「浮気」の可能性が7倍に!?

では、ここから実際にベイズ推論を行ってみたいと思います。求めるものは「パンティが発見された時に、パートナー(夫)が浮気をしている確率」です。

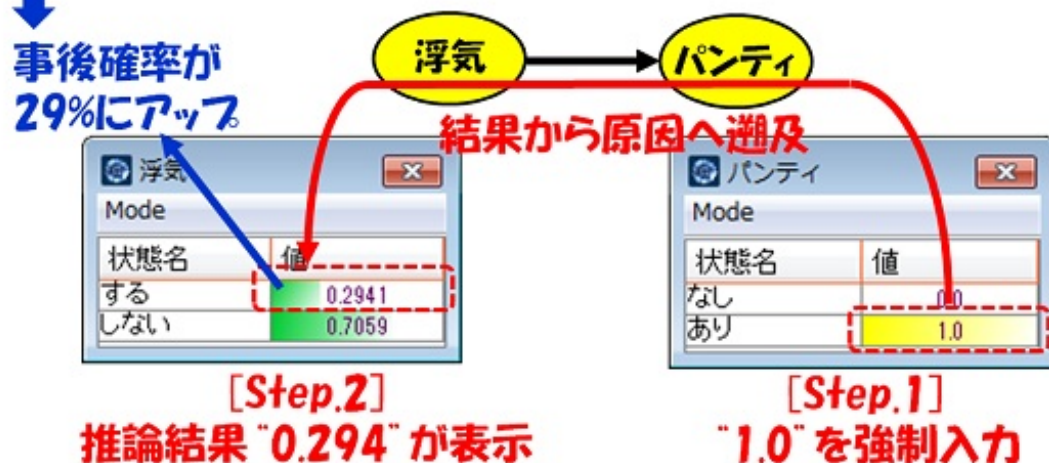
推論モニター(CPTではないので混乱しないように)を開いて、「結果のノード」に強制的に値を入力してください。あなたは、ここで初めて「結果から原因に遡る(さかのぼる)計算結果」を目にすることになります。

ベイズネットワークの使い方

■Step.1: 「推論モニタ」を表示します



■Step.2: 「推論モニタ」に、結果の値を(強制的に)入力します



同じカラムを入力にも出力にも使う

パンティを発見されてしまったという事実が確定した(確率100%)ことによって、浮気の実確率が、4%から29%以上にアップしていることが分かります(この4%→29%の内容については、[前回のこのページ](#)の「パンティ問題を数字で実感する」を参照してください)。

このようにベイズネットワークによって、物証(パンティ)を部屋に置いておくという「うかつさ」が、浮気の可能性を7倍以上も引き上げることが、定量的にも確認できることをご理解いただけた(?)と思います。

時をかける人工知能

さて、私、このベイズネットワークとBAYONETを、とことん使い倒してきたユーザーの1人です。なぜかという、世の中には分からない数字が多すぎるからです。

以前、別媒体で、[性同一性障害についての連載](#)を担当していた時、心底困っていました。

「性同一性障害(GID^(*))を計算する」シリーズ

(*)GID:Gender Identify Disorder

■2014年6月～11月
全5回の集中連載

■試してみた計算

(1)自死者数から見た、GIDの
「痛み、苦しみ」の数値化

(2)性同一性障害(GID)の
国内人口の推定



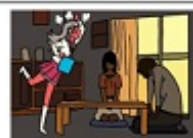
■痛み、苦しみ、
■自己意識



■施術の内容
■戸籍性別変更の方法
■日常生活での困難さ
■GID人口数の推定



(番外編)
■民法772条の考え方



■結婚と出産の問題
■出産の技術的可能性

BAYONETの「変な使い方」をご紹介します

それは、「性同一障害」と認定される前の段階の「自分の性に違和感を覚える人」の国内人口数が全然分からなかったことです。論文、ネット、その他いろいろと調べてみたのですが、データの値がバラバラな上に、その根拠の記述が絶無だったのです。

「自分の性に違和感を覚える人」の人口は？

■絶望的なほどデータがない

例3:「自分の性に違和感を覚える人」の人口数

→ バックデータなしの「**根拠なき数値**」が一人歩き

■流布されている数値の例

(1)GIDに関する最も古い文献→MTF **3万人に1人**
FTM **10万人に1人**

(2) 文部科学省のレポートから→ **16万人に1人**
(コメント:「実数を反映しているものとは言えない」)

(3) その他、新聞に「**2800人に1人**」等の記載

(4)娘(小学校6年生)へのインタビュー
→「そうかな」と思う子は、**学校に2人ほど**いる

はっきりとは分からない

その連載の時点において、はっきりしていたデータはたった1つ、戸籍上の性別変更が認められた人の人数だけでした。

このたった1つの数値から、「自分の性に違和感を覚える人」の人数に遡れるか、が目下の解決すべき課題となっていたのです。

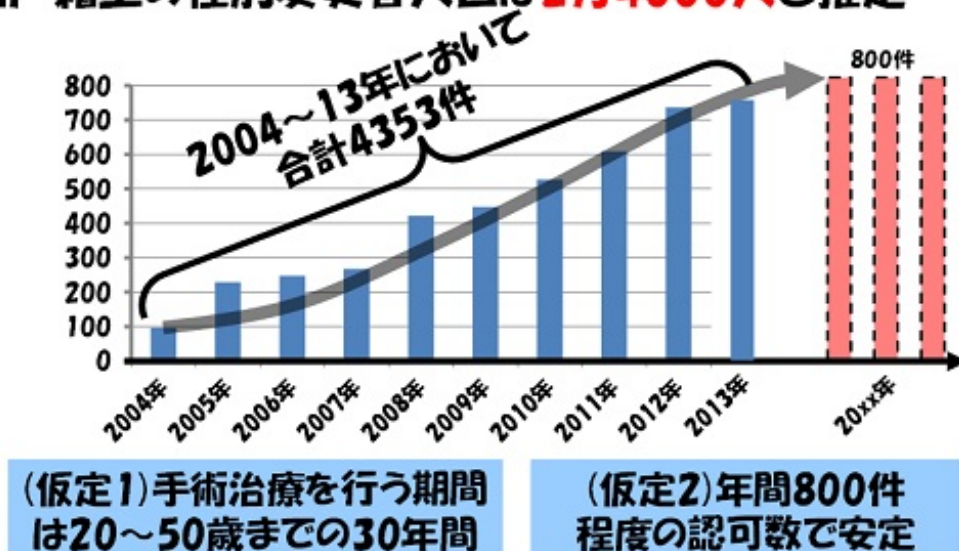
では、ここから、現在の結果から過去の原因に遡る、「時をかける"人工知能技術"」をご覧くださいませ。

ベイジアンネットワークで推測できるか？

■絶対的に確実な数字が「1つだけ」あった

戸籍上の性別変更が認められた人→ 4353件(～2013年)

■戸籍上の性別変更者人口は2万4000人と推定

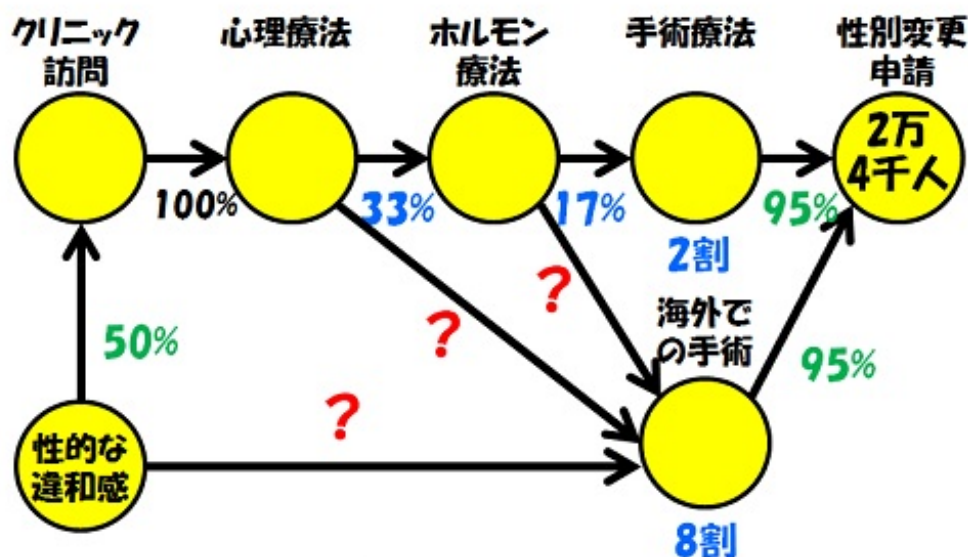


私は、ここで上記の“超ざっくりな仮説”を2つ立てて、まずは戸籍上の性別変更人口を推定しました。

次に、さまざまな文献を読んで、戸籍上の性別変更の手续に至るまでのプロセスをベイジアンネットワークに見立てて、そこに、ネットや書籍から拾ってきた数値を、片っぱしから組み込みました。

さらに、分からないところは、拾ってきた数値に矛盾が生じない範囲で、いろいろな数値を手作業で入れ直しながら、ベイジアンネットワークのチューニングを繰り返しました。

こういうベイジアンネットワークを考えてみた



青色: ネットの文献から拾った数値

緑色: ネットのブログ等や直感から推定した数値

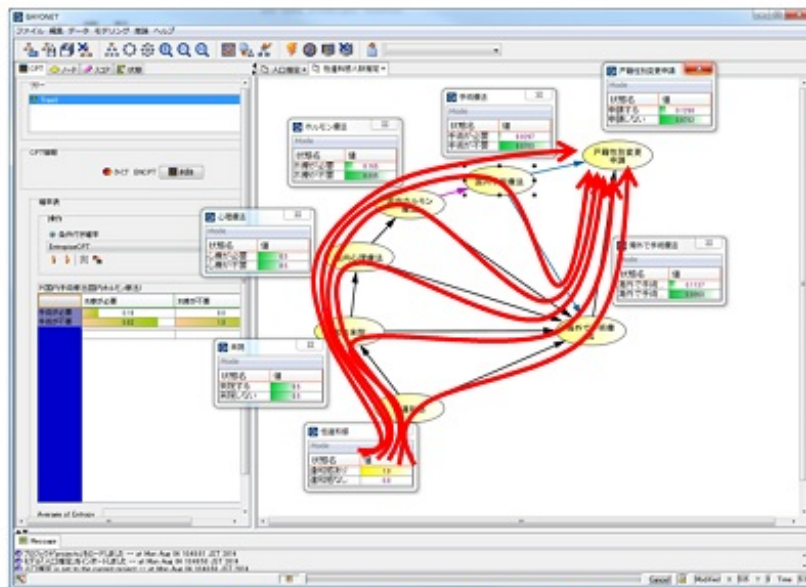
“?”: どうやっても見つけれなかった不明な数値

「不明な数値は、主観で作っちゃえ」を強行

「680人に1人」、推論で根拠のある数値に

これをBAYONETにたたき込み、力づくで計算させた結果、わが国において「自分の性に違和感を覚える人」は、680人に1人という推論結果をたたき出しました。

ベイジアンネットワークを使った推論結果



http://biz-journal.jp/2014/08/post_5706_6.html

性に違和感を感じる人の約**13%**
が、戸籍の性別
変更に至る

↓
日本全国で
18.5万人

↓
680人に1人

単に「条件付き確率」をやっただけ

図版の一部出典:[Business Journal](http://biz-journal.jp)

上記の例では、ベイズ推論をしたというより、ベイズ推論の条件付き確率の連鎖の結果を導き出した、というものになっています。

この結果は、連載当時に私がかき集めた数値や推定や完全な主観をベースとしているもので、当時も今も、信頼性があるものではありません。

しかし、もし、今、さらに新しい事実(データや因果関係)や加わったのであれば、それを後からバカスカ追加して、ベイジアンネットワークを組み直していけば良いのです。

これが、ベイジアンネットワークの「新しいデータが入ったら、どんどん(確率を)変化させていこうぜ」のノリ、"ベイズ改訂"の実体です。

いずれにしても、現在においても「自分の性に違和感を覚える」の統計的なデータを収集することは絶望的に困難だと思います。まだまだ、「性に対する違和感」に対する世間の理解は十分ではなく、偏見を持つ人が存在する以上、その当人にとって、その違和感を告白する恐怖は、計り知れないものがあると思うからです。

それでも、

「よく分からないけど、違和感を覚え、どうして良いのか分からずに苦しんでいる子どもがいる」

というフレーズと

「あくまで推定の域を出ないが、少なくとも日本のほぼ全ての学校(全校生徒数が800人程度)で、1校当たり少なくとも1人は、その違和感に現在進行形で苦しんでいる子どもがいる」

というフレーズのどちらに価値があるかと問われれば、私は後者の方だと信じています。

しかし、結局のところ、「お前は、一体誰のために『数字を回して』いるのか」と尋ねられれば、私は「自分のためだ」と答えると思います。私は、私自身をだますために、これらの数字を回しているのです。

私にとってのベイジアンネットワークとは

■この推論結果は、間違いなく「正解ではない」

- そもそもベイジアンネットワークの設計が怪しい
- ネット上のデータ自体も怪しい
- 江端の主観値が入った段階でもう駄目だろう

■キャッチコピー

それが、本当に正しい値かどうかは分からないけど、根拠は「ない」より、「ある」方がいいよね

■効果

- 自信を持って、コラムを書き進めることができる

自分で自分を騙す為の道具(ツール)でO.K.

アンケート結果を分かりやすい表示に変える

では、次の事例に入ります。

次は、「結婚を計算する」シリーズにおいて、ベイジアンネットワークを使って、「人工的な出産技術」に対する読者の皆さんの心の中を観察した事例です。

「結婚を計算する」シリーズ

■2013年10月～2014年5月、全12回の集中連載

<http://www.kobore.net/kekkon.html>

	【第1、2回】 人口予測 非婚者予測		【第5、6回】 不妊		【第9、10回】 同性出産
	【第3回】 出産させないシステム		【第7回】 人工出産 技術		【第11回】 事実婚
	【第4回】 結婚=幸福 製造装置		【第8回】 AID、クロー ン技術		【第12回】 子どもの値段

■「人工的な出産技術」を、読者はどう考えるだろう？

読者から、直接意見を聞きたい

読者の皆さんには、以下のような5つだけの、単純で、直接的で、そして遠慮のない質問に対する回答をお願いしました。

江端個人によるアンケート

■ブログで報知して、アンケートの回答を貰った

http://biz-journal.jp/2014/01/post_3940_5.html

はい / いいえ / 分からないで 回答してください

- ・質問1:「冷凍卵子」「冷凍精子」を活用すべきですか？
- ・質問2:「人工子宮装置」を開発すべきですか？
- ・質問3:「人工子宮装置」で生まれた子どもと、そうでない子どもは違いますか？
- ・質問4:「人工子宮装置」で生まれた子どもを愛せますか？
- ・質問5:「技術」による少子化対策は必要ですか？

回答数 わずか16人

頂いたアンケート結果の表は以下の通りですが、正直、この表をどんなに眺めていても、何も見えてはきません。

そして、世間一般のアンケート結果とは、大抵の場合こういう形になっており、アンケート結果の分析とは、どこからどう手を付けたら良いのか、さっぱり分からないものなのです。

アンケート結果一覧

コード名	属性			質問1: 「冷凍卵子」 「冷凍精子」の 積極活用	質問2: 「人口子宮装 置」の研究開発 と実用化	人工子宮装置で産まれた子 ども、に/を		質問5: 技術による少 子化対策は必 要か
				質問3: 「違い」はある		質問4: 愛せるか		
A	女性	15歳	未婚	賛成	反対	ある	愛せない	不要
B	女性	20歳	未婚	反対	反対	ある	分からない	不要
C	女性	20歳	未婚	分からない	反対	ある	愛せない	分からない
D	女性	20歳	未婚	賛成	分からない	分からない	愛せる	分からない
E	女性	25歳	未婚	分からない	分からない	ない	愛せる	不要
F	女性	35歳	既婚	賛成	賛成	ある	愛せる	必要
G	女性	35歳	既婚	賛成	賛成	ない	愛せる	不要
H	女性	35歳	未婚	賛成	分からない	分からない	分からない	必要
I	男性	35歳	既婚	賛成	反対	ある	愛せない	必要
J	男性	35歳	既婚	賛成	賛成	ない	分からない	必要
K	男性	35歳	既婚	反対	賛成	ない	愛せる	分からない
L	男性	40歳	未婚	賛成	反対	ある	愛せる	不要
M	男性	45歳	既婚	反対	反対	ある	愛せる	必要
N	男性	45歳	既婚	賛成	反対	ある	分からない	必要
O	女性	50歳	既婚	反対	反対	ある	分からない	分からない
P	女性	55歳	既婚	賛成	反対	分からない	愛せる	必要

あなたは、ここから何が読み取れますか？

ところが、これのアンケート結果の[csvファイル](#)をBAYONETに突っ込むだけで、解析作業のほとんどは終わったも同然です(本当に、これ以外のファイルはいりません)。

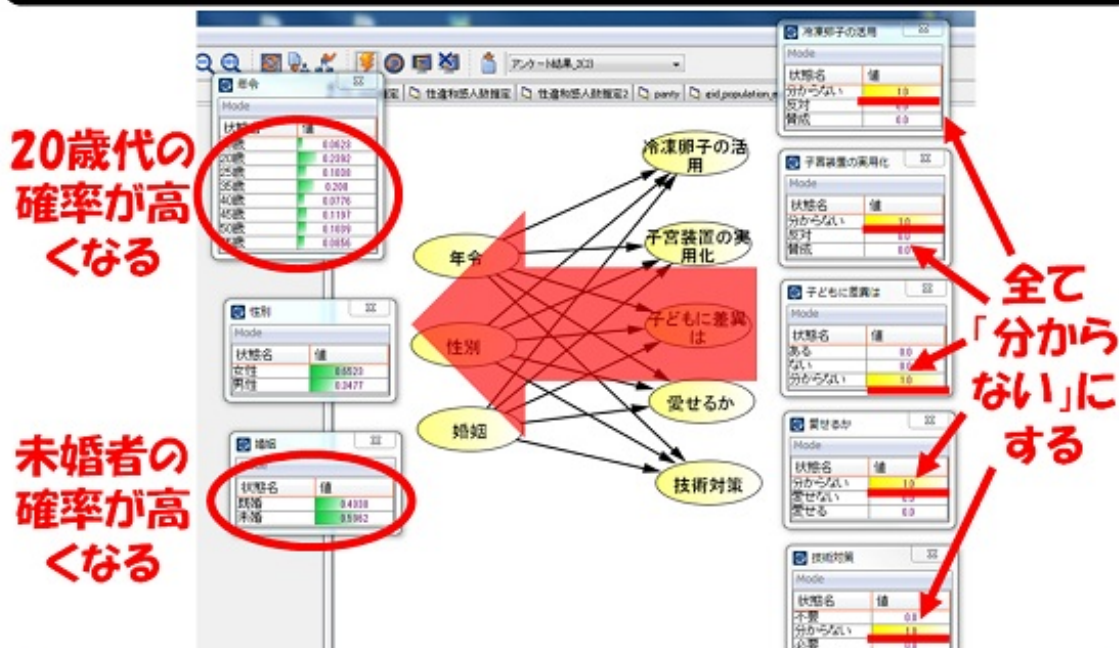
BAYONETは、アンケート回答者の属性(性別、年齢、既婚/未婚)および質問事項のそれぞれを、ノードとしてポコポコと自動生成します。私は、そのノードの間を自分の望む結果が得られるように、矢印を引く作業をただけです。

しかし、このベイジアンネットワーク(のツール^{*)})を使うと、例えば、こんなことが一瞬に分かってしまいます。

*)BAYONETは、申請すれば1カ月間の無料ライセンスをもらいますので、興味のある人はお試しください。

(1)人工生殖に対して、「よく分からない」と回答している人は、どんなタイプの人か？

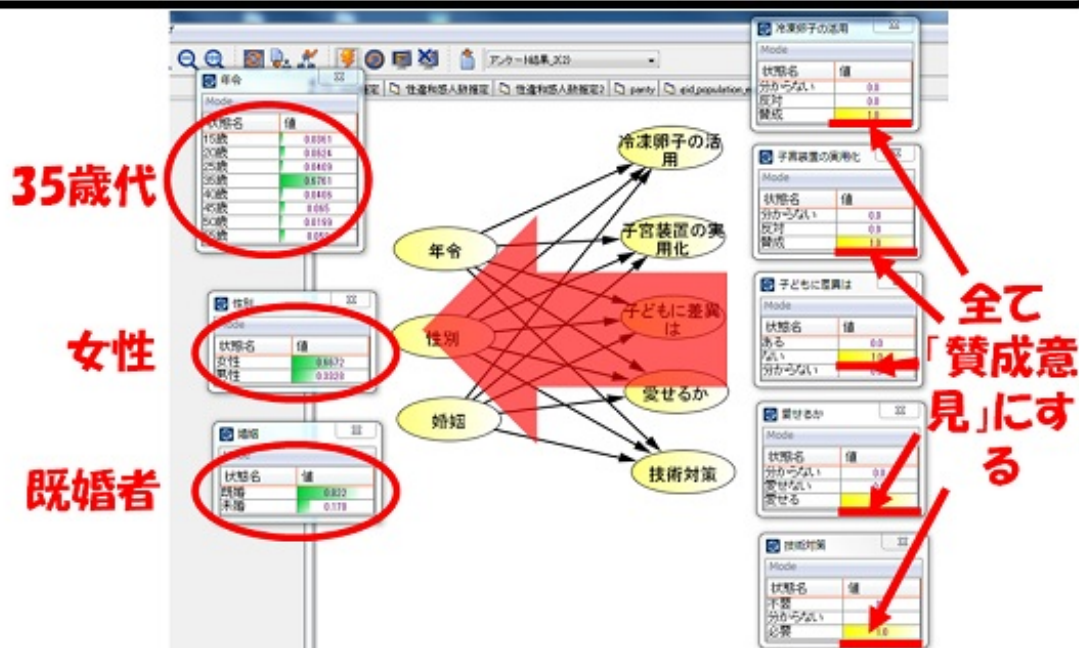
「『分からない』と回答する人」の条件では？



「若い」or「未婚者」は当事者意識が低い

(2) 人工生殖に対して、賛成の意見を表明しているのは、どんなタイプの人か？

「人工的な出産技術」に賛成する人は？



「明示的に賛成」の人は実に良く見える

とまあ、こんな風に、あっという間に、分かりにくいアンケート結果の中身を、丸裸にしてくれる、私にとっては「魔法の"人工知能技術"」なのです。

これ以外にも、これまで、もう1つの連載「[世界を「数字」で回してみよう](#)」では、読者の皆さん(のアンケート結果)と、ベイジアンネットワークと私の連合チームで、以下の結論を定量的(×観念的)に導き出してきました。

題目	結論
部分痩せは可能なのか？ (前編) (第19回)	「 部分痩せ 」に興味があるのは 女性だけ
「ダイエットの目的は…… 結局、ダイエット!？」 (第21回)	「 ダイエット 」と「 健康 」の因果関係は強い 「 ダイエット 」と「 もてたい 」という気持ちの因果関係は 弱い

これからも私は、このベイジアンネットワークという武器を使って、世の中の比較的ど〜でもいい(偉い学者さんや研究員がテーマにしないような、日常の出来事の)数字を、どんどん「見える化」していくつもりです。

「江端による視力偽装問題」

では、ここからは、ベイズ推論における最大の特徴「ベイズ改訂」について、私が独自に編み出した、人間ドックにおいて視力検査装置を欺く「江端偽視力問題」を使って、具体的に説明します……

—— と思ったのですが、今回の私の膨大な初版原稿を手にした担当のMさんの「ひきつった笑顔」が見えたような気がしましたので、ベイズ推論、ベイジアンネットワークについては、これにて「撃ち方やめ」とすることに致します。(興味のある方は、[こちら](#)をご覧ください)

(編集担当のつぶやき) ベイジアンネットワークね。来たな、「浮気とパンティ」(ワード原稿12ページ目)。⇒あ、まだあった。「性同一性障害の計算」(原稿19ページ目)⇒え、まだあるし。「人工的な出産技術」(原稿25ページ目)⇒え!まだあるし!!! (原稿30ページ目)⇒……て、やっとな「撃ち方やめ」だった!! ※江端様には、契約を締結した際、「原稿2〜3ページで連載しましょう」と前担当者が依頼しております。

それでは、今回のコラムの内容をまとめてみたいと思います。

【1】前回、合衆国大統領選挙における「選挙人制度」によって、投票比率と著しく乖離した選挙人の数となる(分散と感度)ことを示し、この制度に対する批判を展開しましたが、今回は、この「選挙人制度」が優れているという観点から合衆国大統領選挙を見直すことにしました。

【2】読者の皆さんからのアンケートをまとめた結果、合衆国大統領選挙の「選挙人制度」は、(1)多民族、多宗教を前提とする合衆国に都合のよい選挙制度である、(2)国民にリーダーシップを認めさせる仕組みが働いている、(3)米国民の気質が反映されやすいものになっている、という可能性が示唆できることが分かり、「あの国の国民といい感じでマッチングしている」と認められることが分かりました。

そして、この制度とこの国民性という2つの解決困難な対象の相乗効果によって、今後も人工知能"技術"は、この選挙の予測には太刀打ちできないだろうという江端見解を示しました。

【3】前々回の「ベイズ推定」を受けて、「ベイジアンネットワーク」の特徴(ネットワークとして図示可能である、など)と、実際のツールの使用例を用いて、ベイジアンネットワークの設計と使用のイメージを説明しました。

【4】これまで私が実際にベイジアンネットワークを使った事例(性同一障害の人口数、人工生殖への意識の傾向分析など)を使って、その有効性を示しました。

以上です。

ベイジアンネットワークが"人工知能技術"なのかどうかは、人によって解釈が違ってくるものになると思います。

しかし、今回の事例において、ベイジアンネットワークを設計したのは、「この私」であり、"人工知能"が自動的に作ってくれたものではありません。

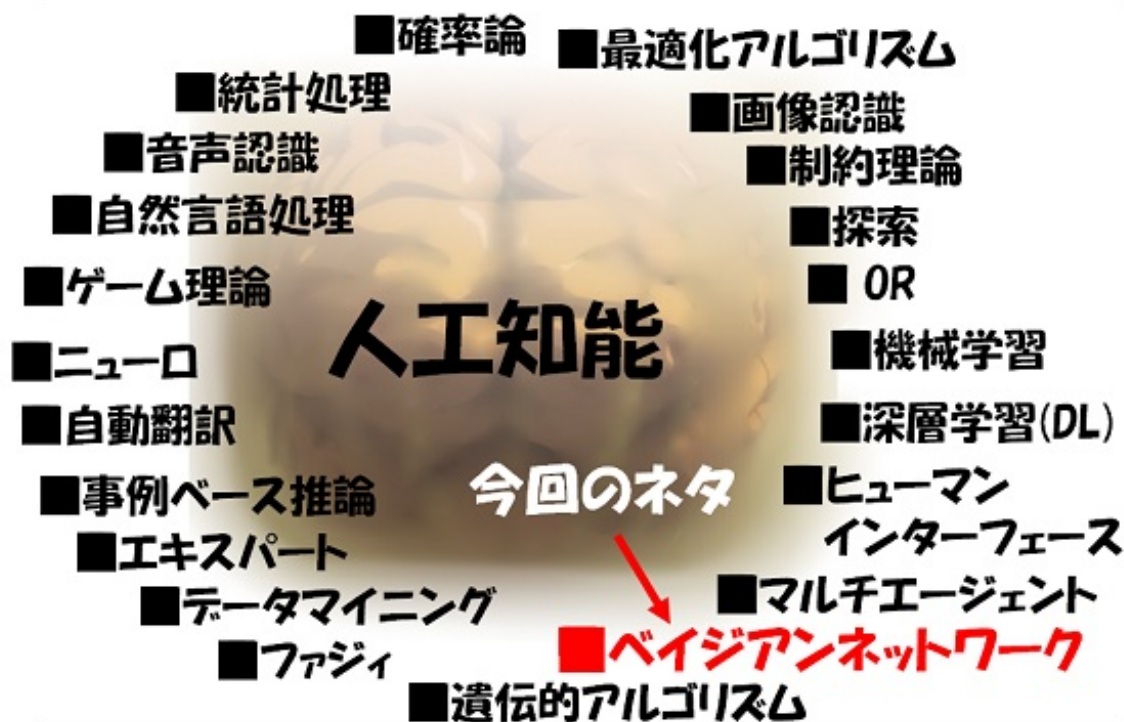
そして、現時点における"人工知能"と呼ばれているものは、全て、人間が、考えて、考えて、考え抜いて設計した設計図に基づいて動いているだけです。

私が知り得る限り、「『性に違和感を覚える人』の人数を推定してよ」と言うだけで、それを算出してくれる人工知能など、存在しません。

2017年、私たちは“人工知能”の夢から覚め始める!?

この連載では、これまで何度も繰り返しており、そしてこれからも何度も繰り返して申し上げていくつもりですが —— "人工知能"という技術は存在しません([「我々が求めるAIとは、碁を打ち、猫の写真を探すものではない」](#))。

“人工知能”という技術は存在しない



(原則として)関連のない個別の技術

私たちの社会には“人工知能”というコンセプトがあり、そのコンセプトを支える“人工知能技術”だけが存在し、その技術を使いこなすためには、“人工知能技術”を利用する技術を習得した人間が、必ず必要となるのです。

来年(2017年)は、私たちが、“人工知能”の夢から覚め始める年になる——私は、そんな漠然とした予感を持っています。

そして、『夢から覚めた時に、私たちが“人工知能技術”を使いこなせる立ち位置に立っているかどうかは、私たち次第である』と私は思っているのです。

(謝辞)

今回の分析では、NTTデータ数理システムのベイジアンネットワーク構築支援システム「BAYONET」(<http://www.msi.co.jp/bayonet/>)を利用させていただきました。

⇒「Over the AI ——AIの向こう側に」⇒[連載バックナンバー](#)

⇒[次回の掲載をメールで受け取る](#)



Profile

江端智一（えばた ともいち）

日本の大手総合電機メーカーの主任研究員。1991年に入社。「サンマとサバ」を2種類のセンサーだけで判別するという電子レンジの食品自動判別アルゴリズムの発明を皮切りに、エンジン制御からネットワーク監視、無線ネットワーク、屋内GPS、鉄道システムまで幅広い分野の研究開発に携わる。

意外な視点から繰り出される特許発明には定評が高く、特許権に関して強いこだわりを持つ。特に熾烈（しれつ）を極めた海外特許庁との戦いにおいて、審査官を交代させるまで戦い抜いて特許査定を奪取した話は、今なお伝説として「本人」が語り継いでいる。共同研究のために赴任した米国での2年間の生活では、会話の1割の単語だけを拾って残りの9割を推測し、相手の言っている内容を理解しないで会話を強行するという希少な能力を獲得し、凱旋帰国。

私生活においては、辛辣（しんらつ）な切り口で語られるエッセイをWebサイト「[こぼれネット](#)」で発表し続け、カルト的なファンから圧倒的な支持を得ている。また週末には、LANを敷設するために自宅の庭に穴を掘り、侵入検知センサーを設置し、24時間体制のホームセキュリティシステムを構築することを趣味としている。このシステムは現在も拡張を続けており、その完成形態は「本人」も知らない。

本連載の内容は、個人の意見および見解であり、所属する組織を代表したものではありません。

関連記事



[困惑する人工知能 ～1秒間の演算の説明に100年かかる!?](#)

コンピュータは、入力値から出力値に至る演算のプロセスを、順番通りに説明できます。ところが、「なぜ、その結果になったのか」と尋ねると、途端に説明できなくなってしまう。困惑し、黙りこくったり、自分が行った演算を“言い訳”のごとく延々と説明したりと、迷走を始めるのです。



[「海外で仕事をしたい」なんて一言も言っていない!](#)

世の中にはいろいろな方がいますが、大きく2種類の人間に大別できます。「英語に愛される人間」と「英語に愛されない人間」です。



[「サンマとサバ」をファジィ推論で見分けよ! 史上最大のミッションに挑む](#)

製品名に「ファジィ」と付けられ、家電が飛ぶように売れた時代。電子レンジの開発に携わる私に突き付けられた課題は、「10種類の食品をファジィ推論で判別する」というものでした。今回は、私の代名詞とも言える、「サンマとサバの研究員」としての原点をお話します。



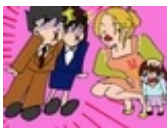
[“ネットワーク研究者”がネット詐欺に遭った日](#)

2015年7月のある暑い日。私は自分がインターネット利用詐欺に遭ったことに気付きました。19歳でPCを譲り受けて以来、ネットワークとともに生き、インターネットの商用発展にできる限り尽くし、自称“ネットワーク研究者”でもある私が、ついにインターネット利用詐欺の餌食となってしまったのです。



“電力大余剰時代”は来るのか(前編)～人口予測を基に考える～

今の日本では、「電力が足りる/足りない」は、常に議論の的になっています。しかし、あと十数年もすれば、こんな議論はまったく意味をなさず、それどころか電力が大量に余る時代が到来するかもしれません。



ラジオ語学番組に出てくる米国人にリアリティはあるのか？

ラジオやテレビの英語講座では、明るくユーモアたっぷりで、決断力にあふれた米国人がしばしば登場します。果たして、米国のオフィスにはこのような人物が本当にいるのでしょうか？ さらに、赴任となると、思いもよらない事態に見舞われることもあります。私の場合、9.11のテロが、まさにそれでした。実践編(米国滞在)となる今回は、赴任中に非日常的な出来事が起こった場合の対処法について、お話しします。

Copyright © 2017 ITmedia, Inc. All Rights Reserved.

