本サービスにおける著作権および一切の権利はアイティメディア株式会社またはその情報提供者に帰属します。また、本サービスの出力結果を無断で複写・複製・転載・転用・頒布等をすることは、法律で認められた場合を除き禁じます。

Over the AI ——AIの向こう側に(5):

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

http://eetimes.jp/ee/articles/1611/29/news026.html

世界中が固唾をのんで、その行方を見守った、2016年11月8日の米国大統領選挙。私は、大統領選の予測こそ、人工知能(AI)を使い倒し、その性能をアピールする絶好の機会だとみていたのですが、果たしてAIを手掛けるメーカーや研究所は沈黙を決め込んだままでした。なぜかーー。クリントンvsトランプの大統領選の投票を1兆回、シミュレーションしてみた結果、その答えが見えてきました。

2016年11月29日 11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についてたっぷ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

でしょうかーー。<u>⇒連載バックナンバー</u>

世界が注目した米国大統領選挙

今回は、<u>前回</u>の「ベイズ推定」を発展させて「ベイジアンネットワーク」の解説をさせていただく 予定でしたが、2016年11月8日(火曜日、現地時間)に行われた、合衆国大統領選挙 —— 正直、私はショックを受けた —— の結果を受けて、急きょ、内容を変更することにしました。

今回は、さまざまな文献をあさり、検索やシミュレーションを行い、私が理解した範囲で、この 選挙の仕組みについてお話したいと思います。

そして、なぜ、"人工知能技術"は、合衆国大統領選挙に太刀打ちできないのか(当選者を予測できないのか)について、私なりの結論を得ましたので、番外編としてご報告致します。

前回のコラムで、私は、ネイト・シルバーさんの著書「『シグナル・アンド・ノイズ」に出てくる、パンティと浮気の相関関係の事例で、ベイズ推論の説明をしました。

ネイト・シルバーさんは、8年前の2008年合衆国大統領選挙で合衆国50州のうち49州(95%)の、4年前の2012年には、 全州(100%)の勝者を正確に予測しました。

週末エンジニア兼データアナリストである私にとっても、これが、どれほどものすごいことであるかは理解できます。



私がこれまで読んできた本の著者、例えば、イアン・エアーズさん(「その数学が戦略を決める」)、マイケル・ルイスさん(「マネーボール」)などは、数字や統計が本当に役に立つことを、これ以上ないくらいに分かりやすい具体的な成功事例で示し、ビッグデータや数理解析という考え方の、「光」の部分を担当したと思っています。

一方、私は、といえば「『ビッグデータ』からは、『ビッグデータに入っているモノ』以上のモノは 絶対に出てこない」という明々白々の事実を、全く理解しようとしない人たち*)に苦しめられ続 けた「闇」の部分を担当した(させられた)という自負があります。

*)「ビッグデータ」という言葉を使う人は、せめて「<u>標準偏差</u>」くらいは理解しておいてください。 それはさておき。

神のごとき的中率を誇ってきた、ネイト・シルバーさんは、2016年の合衆国大統領選挙*)(以下、大統領選という)において、投票日の10日前の時点でヒラリー・クリントンさんが勝利する確率を80~85%と予想していました。

*) 民主党指名のヒラリー・クリントンさん、共和党指名のドナルド・トランプさん、その他、多くの政党から立候補があった大統領選挙

しかし、今回の、合衆国大統領選挙の結果については、ネイト・シルバーさんだけではなく、世界中の多くの人の予想を覆す結果になりました*)。

*) 事実上のヒラリー・クリントンさんとドナルド・トランプさんの対決となり、多くの世論調査を覆しドナルド・トランプさんが勝利した。



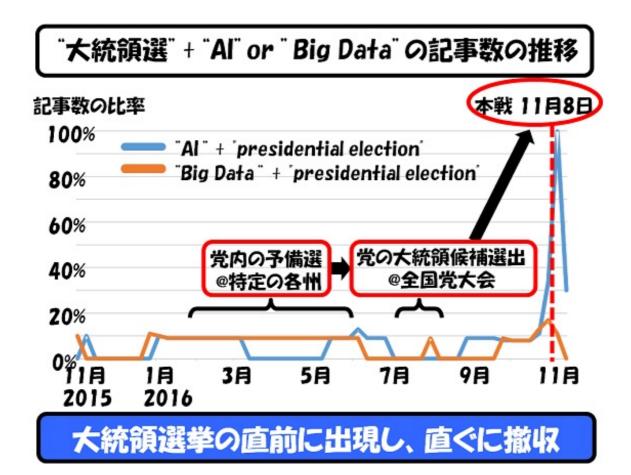
まさかの結末に世界が驚いた

そして、私はといえば、この大統領選の話題に言及してこなかったことを、今、心の底からホッとしています —— 『ああ、いらんこと言わんといて、本当によかった』と。

大統領選に絡むAIの記事がない!?

今回、私は初めて大統領選について調べてみたのですが、米国のマスコミがここ1年間、ずっと出し続けてきた、その「予測の数字」について算出根拠(データ、計算方法、推定アルゴリズムなど)が、ほとんど「ない」ことに気が付きました。

そこで、Googleトレンド使って、"AI" + "presidential election" と "Big Data" + "presidential election"で、記事数の推移を調べてみました。



大統領選の選挙戦は、約1年間続きます(後述)が、"AI"に絡んだ記事は、11月8日の本戦の 直前と直後に突然大量に登場する以外、全体的に単調です。予備選挙や党大会でも選挙が行 われていますが、これらのイベントの影響は見られません。

--- 変だな

ビッグデータや人工知能を華々しく登場させるなら、大統領選ほど利用価値のある事例があるとは思えません。もっと前から大統領選挙の予測を大々的に宣伝し、自社の「人工知能」や「ビッグデータ」の売り込みをかけるのがビジネスというものです。

しかし、私が調べた限り、日本はもちろん、地元の米国ですらビッグデータや人工知能を取り 扱うメーカーやベンチャーは、大統領選の予測に対して完全な沈黙を決め込んでいたのです。

こんにちは、江端智一です。

共和党のトランプさんが、オハイオ州を制したという情報を入手した時に、私は、今回 の2016年の合衆国大統領選挙は、

―― ビッグデータ解析と予測技術("人工知能技術"を含む)の完敗

と、判断しました。

今回の大統領選においても、各種のシンクタンク、調査機関が、最新のデータと予測技術("人工知能技術"も含んでいるだろう)を使って、具体的な数値を上げて、予測を展開していました(ただ、前述の通り、その数値の根拠は提示していません)。

特に、米国の3大テレビ局(ABC、CBS、NBC)は、最初に「当選確実」を出すことに血道を上げているといってもよく、お抱えの統計学者やデータアナリストを数多く雇っています*)。

*) 「異端の統計学 ベイズ」 (シャロン・バーチュー・マグレイン著 草思社 2013年)

有効投票数が億の単位(今回は1億1943万人)にもなる、この世界最大の選挙の予測は、 投票が行われる前における当選者予測もさることながら、世界中がじっと見守る中、刻々と変化 する膨大な数の、不完全で不確かな情報の意味を、トップスピードで理解しなければなりません 。このリアルタイム予測*)こそが、このテレビ局の最大の「売り」になるからです。

*)ご存じの通り、日本でも、次の日になれば確実な当選結果が明らかになるのにもかかわらず、投票日の夜には、開票速報番組で各局がしのぎを削っています(そして、過去には、間違った「当選確実」も相当数ありました)。

今回の大統領選においては、さまざまな調査会社、米国のテレビ局、その他のメディア、そして日本政府や日本国民、もっと言うなら恐らくは世界中の人々も、民主党のクリントンさんが、米国初の女性大統領となることを確信していました。

一方、共和党のトランプさんは、政治経験なし、米軍所属経験なし、過激で差別的な発言の数々を連発し続け、米国はもちろん、わが国でも多くの人も眉をひそめ、間違っても、トランプさんが大統領選本戦で勝つなど、ありえないと思っていました。

二人の候補者の政策や言動、資質うんぬんよりは、むしろ「トランプさんに勝てる要因がない」という恣意的な直感の方が先行し、多くのメディアで「『クリントンさんが勝つ』という数字が、あちらこちらにばらまかれ」て、さらにその恣意的な直感を強化していった、という見方もできるかと思います(まあ、終わった後なら何とでも言えますが)。

米国の選挙システムに仕込まれたパラドックス

今回の大統領選の結果を受け、私は初めて選挙の仕組みについて勉強をしてみました*)。

*) 例えば、「選挙のパラドクスーなぜあの人が選ばれるのか?」 (ウィリアム・パウンドストーン著 青土社 2008年) などを読みました。

そして、「自分が『いいな』と思った人に投票する」という単純明解な選挙というシステム自体に、民意を反映できないパラドックスが自動的に仕込まれる*)という事実を知って、びっくりしています。

*) 例えば、「アローの不可能性定理」などです。

これは、どんな選挙の仕組みをプログラミングしても、そこに、取り除くことができないバグが 運命的に仕込まれ、そのことを皆分かった上で、それでも、なお、私たちは選挙というプログラム を実行しなければならない、ということです。

不可思議な投票方式

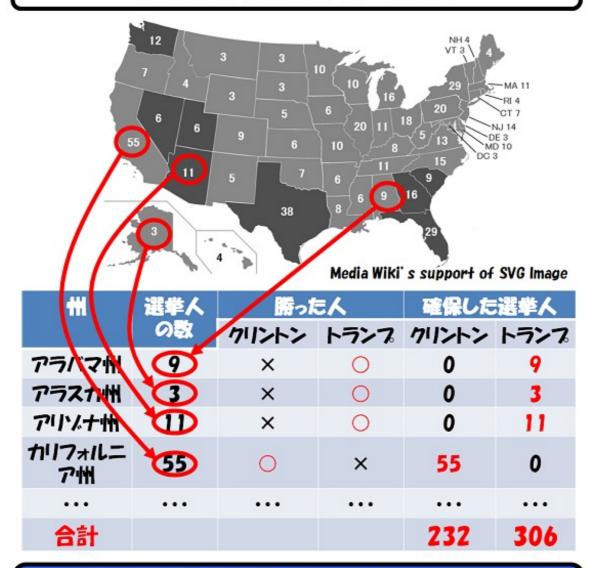
合衆国大統領選挙も、私にとっては理解不能で不思議な投票方式が運用されています。

ひと言で言えば、「投票数で大統領が決まるわけではない」という投票方式です。

大統領選挙は、州ごとに勝ち負けを決める選挙です。しかし「勝った州の数が多い方が勝ちというわけでもない」のです。

各州には、(おおむねその州の人口に応じた) "持ち点" があり、その持ち点の総計の多い方が大統領として選ばれます。この持ち点のことを選挙人といいます。その州で勝利した党が、その選挙人の投票数を全て総取りできるのです。

合衆国大統領選挙の方式



なんで、こんな面倒な方式を 取りやがるのかなぁ

例えば、アラスカ州で勝っても3点しかもらえないのに、カリフォルニア州で勝つと55点もらえるのです。そして、この3点とか55点は、2人の候補者で分けることができません。例えば、カリフォルニア州で勝てば55点は自分のもので、負ければ0点となります。

で、この点数(選挙人)を合計すると全米で538点になります(なんで、わざわざ偶数のままにしておくのかも気持ち悪いですが)。その半数の269点に1点を加えた270点となった段階で、大統領選の勝利が確定する、ということです。

一一どうです、このすさまじいまでの分かりにくさ。

この訳の分からない投票の仕組みの理由については後述致しますが、今は、投票数の多かった人や、多くの州で勝利した人が、必ずしも大統領選で勝者となれるわけではない、ということだけ理解していただければ十分です(実際、今回の選挙でも、有権者投票数では50.08%対49.92%でクリントンさんが過半数を取得していますし、2000年の大統領選(ブッシュ対ゴア)においては、票の数え直しを求めて裁判になったこともあります)。

大統領選、連続1万回シミュレーションを決行

あまりに訳が分からなかったので、私は、乱数を使ったとても単純なシミュレーションで、合衆 国大統領選挙の性質を理解してみようと考えました。私には、大統領選の数式モデルを考える 頭脳も時間もないので、いつでも、このような力づくのシミュレーションをやるのが常です。

今回は、<u>このようなプログラム</u>を作って、「米国の各州の住人(未成年や公民権が停止中の人も含めて)、約3億人の人間を、投票率100%で、クリントンさん、トランプさんのいずれかに、ランダムに投票する」というシミュレーションを実施しました(2016年の実際の有権者登録数は、1億1943万人です)。

また、本来のシミュレーションであれば、地盤州(ほっといても勝てる州)の選挙人を定数として、激戦州(スイング・ステート)の選挙人だけを変動対象にしてシミュレーションすべきなのでしょうが、今回の目的は「選挙の性質を理解する」ことなので、あえて、米国の全州(コロンビア特別区を含む)を、スイング・ステートとして取り扱うことにしました。

とはいえ、3億人分の投票は私のPCでも大変だったようで、1回の選挙に約5秒間を必要としました。

また、この「デタラメに投票する」を作るために、当初C/C++言語が標準実装している乱数の 関数(rand())を使ったのですが、3億人を相手にするシミュレーションでは精度が悪すぎて(2 の31乗 = 21億回くらいで乱数が循環してしまう)、変な結果(例えば、「クリントンさんが不利 な状況のシミュレーションで、クリントンさんが全勝する」など)が出てきてしまったので、乱数に 関しては精度の高いものを組み込むことにしました(参考)。

さて、このシミュレーションでは、投票者は「ランダムな(デタラメな)投票」をするので、普通に考えたら、選挙の結果は、クリントンさん、トランプさんが、勝ったり負けたりして、勝率50%となり、選挙人の総数も、270人±10人程度に納まるだろう、と予想していました。ところが、何回シミュレーションをし直しても、私の予想を超える結果(選挙人数)になってしまうのです。

例えば、ある乱数初期値では、クリントンさんが400人を取り、トランプさんが138人になることもあれば、その逆のパターンも何度も出てくるのです。

バグ? とも思ったのですが、このシミュレーションプログラムが、アホみたいに単純(main()の行数20行くらい)なものなので、バグとも思えませんでした。

そこで、私は、自宅のPCで、

――「合衆国大統領選挙、連続1万回シミュレーション」

を強行しました。

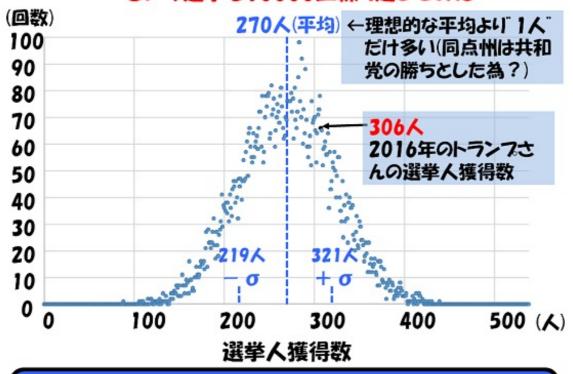
シミュレーション結果の平均値が一定であれば、バグではないという確信が持てると思ったからです。ただ、今回のシミュレーションは1回につき5秒かかるので、1万回繰り返せば5万秒、つまり約14時間かかります。今月の江端家の電気代は、結構な金額になるだとうと腹をくくっております。

それでは、江端家の電気代と引き換えに算出した計算結果を以下に示します。確かに、バグではなかったようですが、江端家のPCは、バグよりも驚くべき結果を私に見せつけたのです。

クリントン:トランプ=50.00%:50.00%

米国民全員が、一人も棄権しないで、なーんにも考えずに、クリントン、トランプのいずれかに、デタラメに投票する。

・・・という選挙を10000回繰り返してみた



完全な乱数を使っても、選挙ごとに こんなにも結果が変わってしまう

信じられないかもしれませんが、1万回の選挙の、クリントンさんと、トランプさんへの投票数は、ほぼ正確に50.00%(最大値50.0102%、最小値49.9987%、標準偏差0.0028%)といえるものです。

にもかかわらず、乱数を使った同じシミュレーションで、これだけのバラバラな結果が出てく

ることは、私のこれまでの統計計算の常識を超えていました。

1万人に1人の心変わりで、勝率がほぼ100%に!?

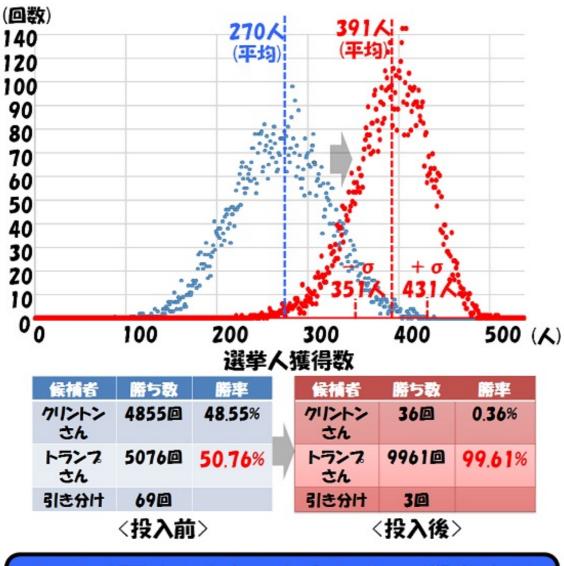
今回の選挙で、トランプさんは、選挙人306人を獲得して「大勝」したことになっていますが、 このシミュレーション結果を見た後では、全く信じられなくなりました。

それでは、もし、この50.00%の均衡を破ってみたらどうなるだろう、と思って、クリントンさんとトランプさんへの投票率を、49%:51%にして再度シミュレーションしてみたら、1万回のシミュレーションの全てで、トランプさんが選挙人の全員538人を獲得するという、仰天の結果が出てきてしまいました。

これでは、比較ができないので、今度は、クリントンさんとトランプさんの投票率の差を、さらに僅差の49.99%:51.01%として、再度、連続1万回シミュレーションを再実行してみました。その結果を以下に示します。

クリントン:トランプ=49.99%:50.01%

「10000人に1人だけ」、トランプさんに投票したいという 意思のある人を投入してみたら、こうなった



たった1万人に1人の心変わりで、選挙人が121人増え、勝率は99%以上に跳ね上がる

―― 1万人にたった1人の心変わりで、勝率が50.76%→99.61%に跳ね上がるって、マジ?

と思いました。これ、多数決で民意を反映することを目的としている選挙としては、世界最低の 選挙方式なんじゃないかなぁ、とまで思いました。

大統領選においては、えげつないスキャンダル合戦が繰り返されているといわれていますが、 そりゃ当然です。 全有権者の、1万人にたった1人の気持ちを変えるだけで、選挙人総数538人中、獲得選挙人数が100人以上がコロっと入れ替わるような選挙であれば、敵方に撃ち込むスキャンダルの弾は、多ければ多いほどいいはずです ―― かっこつけて、スキャンダル合戦に参戦しない候補者がいたら、そいつはアホだといっても良いでしょう。

このように、現状の米国大統領選の仕組みは、そもそもが、投票のパターンに対して結果(選挙人獲得数)が恐しく敏感であり、加えて、ほんのちょっとした有権者の気まぐれで、選挙結果が簡単に大逆転してしまうものであることも分かってきました。

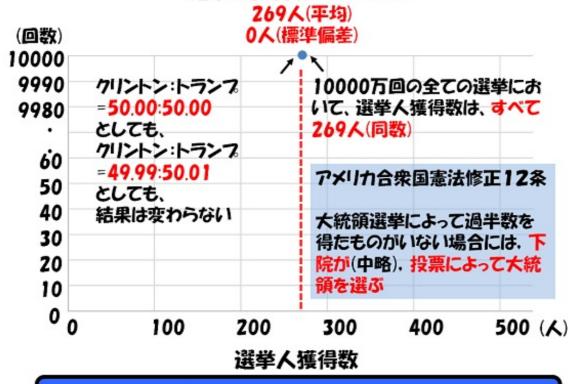
獲得投票数で勝負が決まるとしたら

そこで、もし、これが、一般の選挙のように、獲得投票数で勝負が決まるものであったと仮定した時、今回のシミュレーション結果はどのようなものになるかを、調べてみたくなりました。

以下に、これまでのシミュレーションと全く同じ乱数、有権者数を使って投票を行い、それを、 選挙人獲得数に置き替えてみたものを示します。

もし投票数で大統領が決まるとしたら(1)

前回と全く同じ投票結果を使って、投票数に応じた 選挙人獲得数としてみた



普通に考えれば、この結果が正しい

先ほどの2つのケース、"50.00%:50.00%"と、"49.99%:50.01%"のいずれにおいても、1万回のシミュレーションの全てで、クリントンさんとトランプさんの獲得選挙人数は、269人

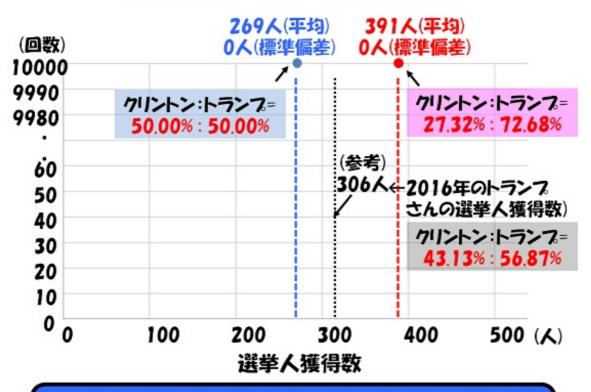
対269人の同点となり、勝負がつきません(標準偏差が0という、バラツキ絶無の結果となります)。

この場合は、合衆国憲法修正12条により、下院の投票によって大統領が決せられることになりますが、普通に考えれば、この結果が自然だと思います。

では、逆に、先ほどの投票比率が"49.99%:50.01%"のケース(勝率が50.76%→99.61%に跳ね上がるケース)を、この投票者数で決する普通の選挙に置き換えた場合、どれくらいの投票数の差が必要になるかを逆算してみました。

もし投票数で大統領が決まるとしたら(2)

投票数に応じた選挙人獲得数としてみて、同じ選挙人の 差を作るには、どれくらいの差が必要か



これが普通の結果になるはず→今の大統領 選は「サプライズ」を作る装置に見えてくる

普通の選挙であれば、トランプさんが選挙人を391人獲得するには、有権者の73%近くの票数を獲得しなければなりません。つまり「圧勝」することが必要です。

仮に、今回のトランプさんの「306人」という選挙人獲得数であっても、そこそこ「目に見える勝利(56%程度の票の獲得)」が必要となります。

今回、最終投票数は、クリントンさんが5917万票を獲得したのに対して、トランプさん

は5904万票となっており、普通の直接選挙であれば、クリントンさんの勝ちです。

しかし、もし投票数に応じた選挙人を選ぶのであれば、269人対269人の引き分けになり、 下院で決戦投票となっていたはずですので、下院で過半数を占める共和党員の投票により、最 後にトランプさんの勝ちになると思います。

このように、普通の選挙と比較してみると、現行の合衆国大統領選挙の仕組みが、いかに「気持ちが悪い」ものであるかを理解できるかと思います。

大統領選の方式にみる、米国人の本音と建前

以下は、米国がこの大統領選挙の仕組みに固執している理由*^{)、**)}と、私の所感を、ざっくり表にしたものです。

- *)「大統領の作り方」(堀田佳男 プレスプラン 2004年)
- **)「大統領はカネで買えるか?-5000億円米大統領選ビジネスの全貌」(堀田佳男 角 川SSC新書 2008年)

米国民がこのやり方に固執する理由(建前)

理由	問題点	解決策	矛盾点
歷史的	(1)200年前の米 国→テレビもラジ オもない、文盲率 も高く新聞も読め ない (2)投票所までの 距離が100km単 位→自動車などな い	「オラの代りに投票 してくれる人を、選 ぶのがよか」 →選挙人制度の 発生	現在、これらの 問題点は、一つ も残っていない →制度を維持 する正当な理 由なし
経緯		「大統領は、州の総 意として選ばれるも のである」 →「州の選挙人 全員一致による 投票」の考え方 の発生	10年おきに、各 州の人口に応じ て、選挙人の数 が変更されてい る →結局、有権 者数に応じた 選挙になって いる

「建前」が形骸化していることを、 米国民は、みんな分かっている 今回、私は、大統領選挙の仕組みが、民主主義の基本原則「多数決の原理」を働かせているかどうかを調べたかっただけなのですが、いろいろと掘り下げてみると、今や大統領選挙は、現在の社会状況を反映しない形骸化したシステムに成り下がっています。

"効率性や合理主義の権化"ともいえるあの米国において、この選挙の仕組みが、最高意思 決定者を決める方式として採用され続けていることに、ある種の「気持ち悪さ」を感じずにはい られません。

大統領選で(江端が)訳の分からない点

訳が分からない点	(江端の考える) 解決策	その他、 江端の所感
■民主主義の基本原則 「多数決」が正しく機 能しているのか、さっ ぱり分からん(本コラ ムでのシミュレーショ ン結果)	とっとと「選挙人制 度」を止めればいい のに	米国において、過去 何度も法案廃止が議 会に提出されている が、憲法改正には 至っていない
■党内の予備戦や、党の大統領選出全国大会など、本選挙以外に、無駄な選挙が山ほどある	本選挙だけにして、 各党は立候補者を無 制限に受け入れれば いいのに	政権を狙う党として、 党内での票割れを防 ぎたいという気持ち は、理解できる
■選挙期間は1年にも および、1人の候補に 50~500億円(?)も のカネが使われている	なし (思い浮ばない)	「金=権力」を是と する米国の国民性は、 日本人(である私)に は理解しにくい

結局、あの国の国民は「膨大な金を浪費する 「祭」が大好き」ということなのかな?

今回のシミュレーションで、米国の大統領選挙は、その仕組みにおいて「ギャンブル性」が組み込まれていることが分かりました。

また、選挙の期間が1年と無駄に長く(正確には、米国には選挙期間が法定されていない*
))、さらには、選挙で1票を投じる以外の方法で立候補者を支援することが認められている
—— 普通のサラリーマンが政治献金(3000円くらいからスタート)をすることが、大っぴらに認められている —— など、どうも日本の常識とは、かけはなれた世界にある、ということを知りました。

*)日本の場合、長くて17日(参議院選挙、知事選挙)、短いものでは5日(町村議会選挙、町村

米国民がこのやり方に固執する理由(本音)

理由	本音	解決策
政治的理由	選挙人制度を止めると、二大 政党制が壊れる →「余計な政党は入ってくる な」	現在の「選挙 人制度」の絶 対的な維持
国民が求める大統領の姿	大統領たるもの、 (1)1年間の選挙期間を勝ち抜く気力と体力があり、 (2)巨額の選挙資金を集めるカリスマがあり、 (3)口論で相手をやりこめる根性の悪さと、 (4)スキャンダル攻撃に耐えられるメンタルもあり、 (5)運もある	今の大統領選 のままでいい

私たち日本人には理解できないリーダー像

それに、今回調べてみて、米国国民が求めているリーダー像というのが、我が国とはかなり違うということも分かってきました。

そこには、「キレイごとだけで、メシが食えるか」という、究極のプラグマティズム*)に基づく、「スッピンの民主主義」というべきものが、よく現われていると思います。

*) 「知識が真理かどうか」とか「正義は何か」とかはど一でも良くて、「生活上の実践に利益があるかないかだけで物事を決定すべき」とする哲学

「相手」が悪すぎた!?

さて、このコラムの冒頭で、私は、「2016年の合衆国大統領選挙は、『ビッグデータ解析と予測技術("人工知能技術"を含む)の完全敗北』であると判断した」と申し上げましたが、この見解を、以下のように変更します。

合衆国大統領選挙は、そもそも、その仕組みの性格上、予測不可能であり、人間はもちろん、 ビッグデータであろうが、"人工知能技術"であろうが、到底太刀打ちのできる対象ではない。

そりゃ、人工知能を取り扱うメーカーやベンチャーが、大統領選の予測に対して完全な沈黙を

決め込んでいたのは当たり前でしょうし、今回、ネイト・シルバーさんが予測を外してしまったのも、至極当然と言えます。

詰まるところ ―― 相手が悪すぎたのです。

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】今回の2016年合衆国大統領選挙では、民主党のクリントンさんと共和党のトランプさんの一騎打ちに。結果、世界中の予想を裏切り、共和党のトランプさんが当選を確実にしました。
- 【2】しかし、このトランプさんが当選を予測した"人工知能"は見あたらないばかりか、そもそも、合衆国大統領選挙の予測を試みたという"人工知能技術"の製品やサービスすら見当たらないことも分かりました。
- 【3】そこで、合衆国大統領選挙の仕組みを調べ、力づくのシミュレーション結果を行った結果、この大統領選は、多数決の民意を反映しにくい「ギャンブル性」が極めて高い方式になっており、現状の"人工知能技術"では、到底太刀打ちできそうなものではないことが分かりました。
- 【4】なぜ、このような「気持ちの悪い」選挙制度を、米国国民が維持している理由を考察した結果、米国人には、日本人の考え方とは根本的に異なるリーダー像がある、という自分なりの結論を得ました。

「数字の敗北」は珍しいことではない

「数字が敗北する」ことは珍しいことではありません。特に「("人工知能技術"などによる)予測の数字」の勝率は、決して高くはありません。

最近では、各国政府もシンクタンクもデータアナリストも、英国のEU離脱についての投票結果の予測も、豪快に外しています。

さらには、(もう多くの人が忘れているかもしれませんが)、2011年の東日本大震災の際、政府が「SPEEDI(緊急時迅速放射能影響予測ネットワークシステム)」のシミュレーション結果の発表を差し控えたことからも、予測(を公表することと、そしてそれが外れた時)の「恐怖」もよく知っているのです(参考)。

ぶっちゃけ、"人工知能技術"などによる数字が威力を発揮するのは「過去の数字」であり、 「過去に起こった出来事(特に、訳の分からないような結果になった時)を、筋の通ったストーリーに(後出しジャンケンのように)組み立てるとき(=仮説構築)」には、威力を発揮します。

しかし、その「後出しジャンケンの仮説」についても、それが正しいかどうかを知る手段はないのです。この「仮説」の意義は、多くの人が、後から「ああ、なるほど、そういうことか」と納得できるストーリーが作れる(かもしれない)、という、その1点にのみ価値があるのだと、私は思ってい

ます。

つまり、ビッグデータ解析や人工知能は、(その対象にもよりますが)

- (1)頻繁に予測を外すが、
- (2)その「外した理由」を説明するためのストーリーメーカーとしては優秀である

ということは言えそうです。

では次回は、今回の予定であった「ベイジアンネットワーク」を解説したいと思います。来月も、 よろしくお願い致します。

「デタラメ」を作ることは結構難しい

これまで私は、C/C++を使って、簡単なシミュレーションプログラムを作ってきました。私には、ある対象の全体像を理解する手段としては、数式や論理で理解するより、直感的で簡単で早いからです。

シミュレーションプログラムでは、なるべく異なる状況を作り出すために、頻繁に乱数を使います。私の場合、C/C++で標準実装されている乱数生成関数(rand())を使ってきました。

rand()は、線形合同法という方式を使っています。これは、簡単な数式で実装でき、計算負荷がかからずお手軽ですが、一定の数の乱数を作り出してしまった後に、同じ乱数を繰り返してしまうという「周期性」の問題と、乱数がほどよくバラけてくれない「均等性」の問題が発生してしまいます。

今回、クリントンさん vs トランプさんの投票を、なんやかんやで、1兆回ほどやってみたのですが、そのシミュレーション結果から、乱数の発生に偏りがあることに気が付きました。これはrand()は、2の31乗回で、同じ乱数を繰り返し発生してしまうからです。

2の31乗とは、約21億です。これは、日常的には、十分に大きな数だと思いますが、1回のシミュレーションで約3億(人分)の乱数を消費するので、(ちょっと乱暴な話ですが)7回目には同じ乱数のパターンを再利用することになってしまうのです。

1兆回の投票に対しては、同じ乱数パターンを1400億回も繰り返し使うことになり、結果、同じパターンの投票が繰り返されるだけ、という、なんとも無駄なことをしてしまっていたのです。

そこで、私は今回、松本眞さんと西村拓士さんが考案されたメルセンヌ・ツイスタという乱数 生成方式を試してみました。これもrand()と同じように「周期性」があるのですが、その周期が ものすごく長いのです。

メルセンヌ・ツイスタの周期は「2の19937乗」=「10の6001乗」です。これは宇宙の誕生から現在までの時間(秒数)を、10の5983乗回繰り返しても、まだ一巡もできないという超長期

の周期性を持ち、事実上、周期性のことは忘れてしまっても良い、というくらい、優れた乱数を生成します。この乱数に切り替えたことで、私のシミュレーションは、妥当な結果を出力するようになりました。

── と、ここまでが前振りです。

話はいきなり変わりますが、1990年代、金融の世界では、複雑化した金融商品やリスクの計算のために、乱数を用いたシミュレーションが必要となってきていました。

そんな時、乱数の生成を、従来の100倍ものスピードで生成するソフトウェアが売り出されました。ところが、その価格が、なんと1億円だったのです*)。

*)「NHKスペシャル マネー革命〈第2巻〉金融工学の旗手たち」(相田洋・茂田善郎 NHK出版 1994年)

この本の中に、どっかの銀行の研究部門の研究員(だったかな?)が、金融商品のシミュレーションをするために、このソフトウェア購入の提案をした、という話が出てきます。

そのソフトウェアの内容を聞かされた重役は「デタラメな数字を出すソフトウェアになんぞに1 億円も出せるか!」と一喝して、その提案を却下したそうです。

しかし「デタラメ」を作ることは、結構難しいのです。なぜなら、一見「デタラメ」に見えている事象であっても、ねちっこく分析し、解析をし続けると、一定の法則が見えてきてしまうものだからです(例えば、このコラムのページなど)。

ですから、「デタラメ」を極めるのであれば、誰が、どこから、どのような解析をしても、―― たとえ、世界最高性能のスーパーコンピュータを持ち込んできたとしても ―― その関連性を絶対に見つけられない ―― というものでなければなりません。

そのような「究極のデタラメ」(あるいは、「限りなく究極に近いデタラメ」)を使わないと、新しい金融商品の正確な評価はできないからなのですが、―― しかし、まあ、その銀行の重役が、 その「デタラメ」の意義を理解できなかったことは仕方がないのかなぁ、とも思えます。

それこそ、合衆国大統領選挙の投票を1兆回も繰り返そうとする物好きでもなければ、この感覚を理解するのは無理だろう、と思えるのです。

⇒「Over the AI ——AIの向こう側に」<u>⇒連載バックナンバー</u>



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「<u>こぼれネット</u>」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



GoogleからAI用プロセッサ「TPU」が登場

Googleが、人工知能(AI)に向けたアクセラレータチップ「Tensor Processing Unit(TPU)」を独自開発したことを明らかにした。同社が2015年にリリースした、オープンソースのアルゴリズム「TensorFlow」に対応するという。



"Al"は死語!? 検索すると、歌手がずらり……

人工知能(AI)の歴史は古い。今は、「AI」という言葉こそ聞かれなくなったが、Appleの「iPhone」に搭載されている音声認識機能「Siri」や、IBMのスーパーコンピュータ「Watson」などの登場により、人工知能自体は再び注目を集めている。では、いったい何をもってして"人工知能"というのだろうか……。



飛び込みにまつわる「6つのなぜ」に新たな仮説が続々?

今回は、「なぜ7月に飛び込みが多いのか」「なぜ火曜日に飛び込みが減るのか」など、私が仮説立案できなかった6つの疑問に対して、読者の皆さまが寄せてくださった仮説の中から、私のハートにヒットしたものを紹介したいと思います。後半では、小田急線の人身事故に巻き込まれた人々の実際のTwitterを分析してみました。



<u>海外出張に行くあなたは、「たった一人の軍隊」である</u>

われわれ英語に愛されないエンジニアが海外に出張する上で、「最大の難所」とも言える場所――。それが入国審査です。今回の実践編(入国・出国)では、海外出張に行くあなたを「たった一人の軍隊」とみなし、敵国(



■■■ =出張先)に首尾よく潜入(=入国)する方法についてお話しましょう。入国審査で使える"レジュメ"も紹介し ます。

Copyright© 2016 ITmedia, Inc. All Rights Reserved.

