



How much better can we predict the future: predictability and the limit.

Kenshi Miyabe

October 2, 2010



規則を見つける どうしたら儲かるか

規則を見つける

テーマ

規則を見つける どうしたら儲かるか

今日の話のテーマ

予測

- 未来を予測するにはどうしたらよいか？
- そもそも未来を予測するとはどういうことなのか？
- 予測はどれくらい可能なのか？

予測ができると嬉しい

規則を見つける どうしたら儲かるか

未来どうなるか、誰でも知りたい。

- 天気予報
- 試験問題
- ゲームで敵がどう動くか
- PK でボールが右に来るか、左に来るか
- プレゼントの反応

今の時代、最も重要なのは情報だ。その中で究極の情報は「未来の情報」だ。(by ノストラード)

過去の傾向を調べよう

規則を見つける どうしたら儲かるか

ポイント 1

未来を知るためには過去を調べよう。

- 過去の試験問題から次回の試験問題を予測する。
- 歴史を学ぶのはなぜか？
- 過去の反応は？

しかし、過去を「覚えても」意味がない。。。

規則があると予測ができる

規則を見つける どうしたら儲かるか

「規則」とは？私を縛るもの??
いやいや、そうではありません。
「規則」とは法則みたいなもの。

- モノは下に落ちる。
- 戦争をすれば両国に死者を出す。
- 肉は焼くとおいしくなる。

ポイント 2

規則を見つけると予測ができる。

偶有性

規則を見つける どうしたら儲かるか

今日、覚えて欲しい言葉 (その1)

偶有性

意味・・・完全に規則的でも完全にランダムでもないこと
例えば、

- 明日の天気
- 期末試験問題

規則と予測可能性

規則を見つける どうしたら儲かるか

ポイント 3

- 完全に規則的ならば、完全に予測可能
- 完全にランダムならば、全く予測不可能
- 偶有的ならば、ある程度は予測可能

突然 Q&A

規則を見つける どうしたら儲かるか

Q 科学や学校の科目で習うことには意味があるの??

A 完全に規則的な部分を習うことが多いのでうまく利用すれば役に立つ。

Q 学校で習うことはきっちりしすぎてイヤ。

A 現実の問題はもっと複雑だからどうぞご安心を。きっちりしていた方がどんなに楽か。。

Q 科学は完全に規則的な部分しか表現できないから役に立たない。

A 完全に規則的でないものの表現の仕方の研究も進んでいますから是非いっしょに研究しましょう。

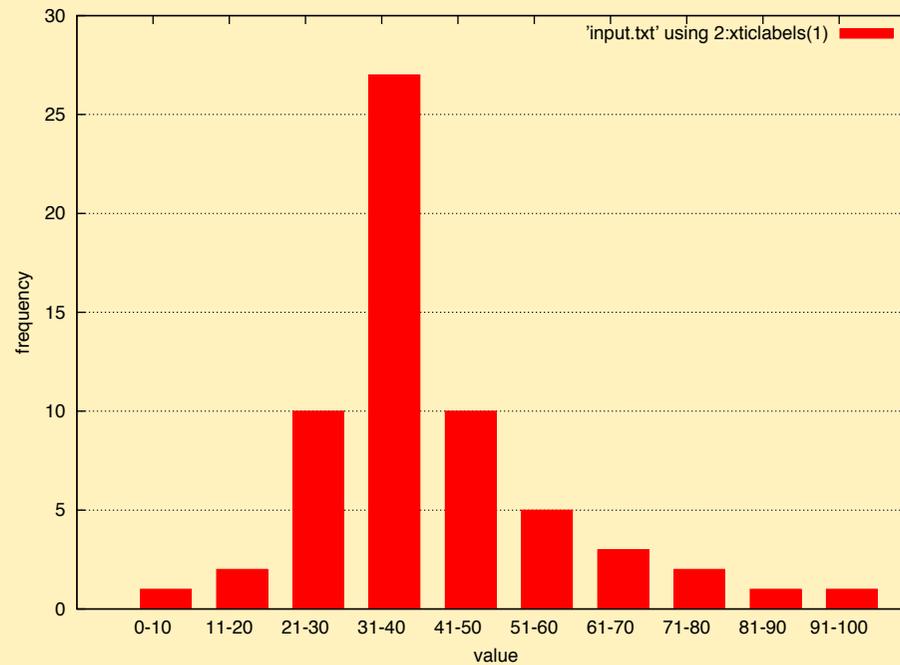
規則を見つけよう（1）

規則を見つける どうしたら儲かるか

最も一般的な規則の見つけ方

平均 だいたいどのくらいの値なのか

分散 どのくらい散らばっているか



規則を見つけよう（2）

規則を見つける どうしたら儲かるか

今、さいころをふったら次のように出たとします。

											平均	分散
A	1	1	1	1	1	1	1	1	1	1	1	0
B	1	6	1	6	1	6	1	6	1	6	3.5	6.25
C	3	4	3	4	3	4	3	4	3	4	3.5	0.25
D	1	2	3	4	5	6	5	4	3	2	3.5	2.24
E	2	5	1	2	4	3	5	6	4	3	3.5	2.24

規則がある列と規則のない列があるよう。

どの列も出る確率は $\frac{1}{6^{10}}$ という意味では同じ。

Aはランダムではないと言うには平均を見れば良い。

BやCがランダムではないというには分散を見れば良い。

Dがランダムではないというのはどうすれば良いだろうか。

平均や分散だけでは表すことができない規則があるようです。

逆にEはランダムだと言えるだろうか。

規則を見つけよう (3)

規則を見つける どうしたら儲かるか

「スパムフィルター」

- (i) 人間がスパムかどうかを判定する。
- (ii) それを元にスパムに良く出てくる単語を検索する。(例えば、free、casino など)
- (iii) 次に受け取ったメールにそれらの単語が出てくるか調べる。
- (iv) よく出てきたらそのメールはきっとスパムだろうと予測する。
- (v) もし間違っていたら訂正をする。

現在のスパムフィルターはいろいろな手法を組み合わせでスパムを判定している。メールの場合にはどんな単語が出てくるかというのが非常に大きな意味を占めている。

規則にはいろいろな規則があって、いろいろな表現方法がありうる。

規則を見つけよう（4）

規則を見つける どうしたら儲かるか

いっそのことすべての表現を考えよう。

例 1. 「aaaaaaaaabbbbbbbbbbb」 という文字列は「 a を 10 個、 b を 10 個並べたもの」

規則があるとは簡潔な表現があるということ。

「圧縮」という操作を考える。

（例えば、zip, lzh, cab, rar など。ここでは可逆圧縮のみを考える。）

ポイント 4

規則があるとは、圧縮できるということ

どれくらい規則があるか、複雑かは、短い表現があるかどうかで判定できる！

- 「aaaaaaaaaaaaaaaaaaaaa」は「 a^{20} 」と表現できて、とても短い。
- 「aceaueirjlij;loigae」は、文字列と同じくらい長い表現しか（きっと）存在しない。

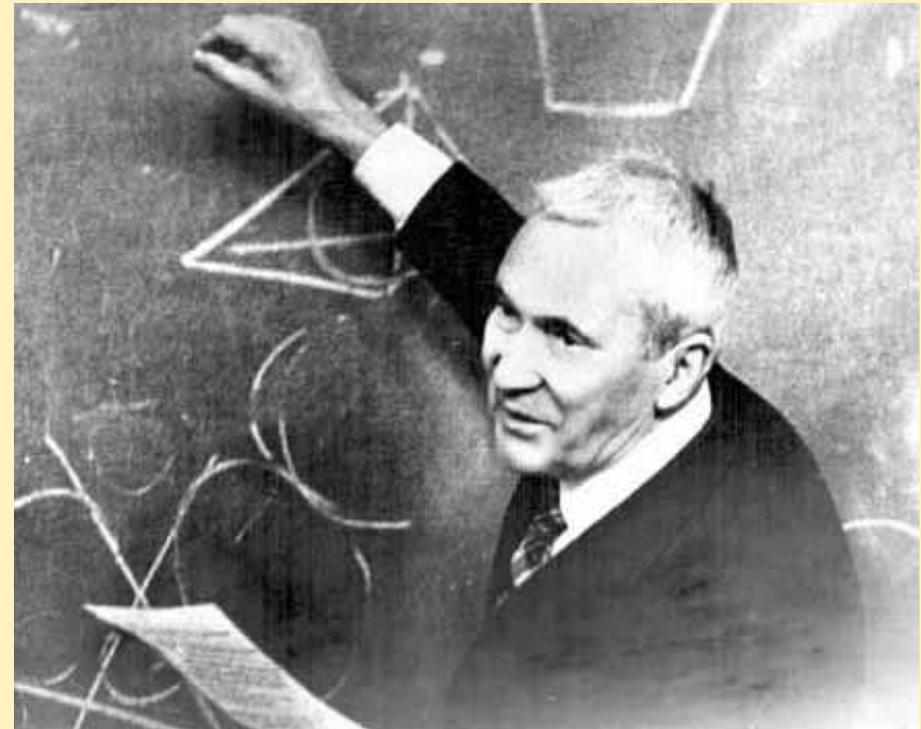
規則を見つけよう (5)

規則を見つける どうしたら儲かるか

今日、覚えて欲しい言葉 (その2)

コルモゴロフ複雑性... 最も短い表現の長さで、複雑さの指標

- アンドレイ・コルモゴロフ
- 1903年4月25日-1987年10月20日
- ロシアの数学者
- 確率論の発展に貢献した。



規則がないランダムな列

規則を見つける どうしたら儲かるか

実は

ほとんどの文字列が圧縮不可能 !!

すなわち、全く規則が見つけられない。

よって、全く予想ができない。

それに対し、日本語の文章は圧縮可能である程度予測可能である。

例えば、「ある」の後には「。」が来ることが多いなど。

世界は規則的

規則を見つける どうしたら儲かるか

なぜ？

どうして世界はこんなにも規則的なのか？

この世界は私たちが語るることができるくらい規則的である。

すなわち驚くほどに予測可能である。

予測不可能であることの方が自然。

予測不可能なことがあっても驚くことではない。

前半まとめ

規則を見つける どうしたら儲かるか

規則がある	\iff	圧縮可能	\iff	予測可能
規則がない	\iff	圧縮不可能	\iff	予測不可能
規則が少しある	\iff	圧縮が少し可能	\iff	予測が少し可能

予測可能性と圧縮可能性は同じ !!

後半は圧縮可能性と予測可能性の関係についてもう少し詳しく。

規則を見つける どうしたら儲かるか

どうしたら儲かるか

予測ができるとは

規則を見つける どうしたら儲かるか

「予測ができる」とはどういうことかを見るために次のようなゲームを考えよう。
「丁半」

- 2つのサイコロをつぼの中に入れて降り、合計が偶数か奇数かを当てる
- 偶数か奇数かの確率はそれぞれ 50%ずつだとする
- ただしコンピュータでシュミレーションする !!
- 所持金を賭けて当たれば2倍もらえ、はずれれば没収される
- 借金はできない

予測ができる \iff 所持金がどれだけでも増やせる

予測ができない \iff 所持金は増やせない

できれば公平なゲームを戦おう

規則を見つける どうしたら儲かるか

このゲームは「公平な」ゲームである。
しかし、現実には公平なゲームはめったにない!!
例えば、

- 宝くじ...47%
- 競馬...75%
- ルーレット...95%
- パチンコ...97%

ポイント5

ルールに従って勝つより、ルールを作る人になる方が簡単である。

規則があると信じよう

規則を見つける どうしたら儲かるか

規則と予測の関係は。。

- 規則があって、圧縮できれば予測ができる。
- 規則がなければ所持金を増やすことはできない。
- 規則があっても見つけられなければ所持金を増やすことはできない。

ポイント6

規則があると信じよう。

いろいろな規則にどう分配するか

規則を見つける どうしたら儲かるか

規則といってもいろいろある。

- 0 が 100 回続いた。
- きっと次も 0 だと思って全財産を 0 に賭けた。
- 1 が出た。
- 一文無しになってしまった。

敗因は??

一つの規則にこだわって他の規則を見失ったこと。

- ずっと 0 が出る。
- 0 と 1 が 100 回ずつ繰り返して出る。

この二つの規則にはどう分配したらよいか?

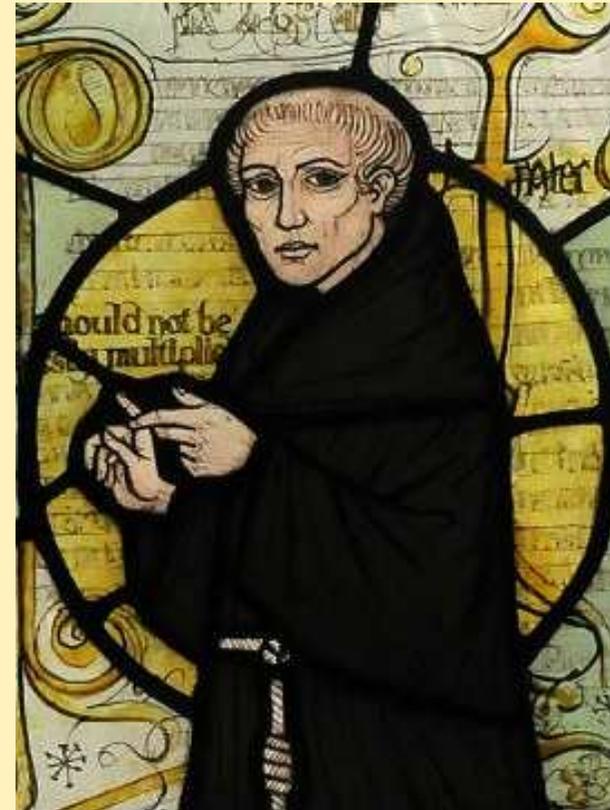
理論は単純な方が良い

規則を見つける どうしたら儲かるか

今日、覚えて欲しい言葉(その3)

オッカムの剃刀... 必要以上に多くの実体を仮定するべきではない。

- 現象を同程度うまく説明する仮説があるなら、より単純な方を選ぶべきである。
- オッカム村のウィリアム
- 1285年-1347年
- 神学者、哲学者



圧縮可能性と予測不可能性

規則を見つける どうしたら儲かるか

そこで「圧縮可能であるほど可能性が高い」として賭けを行う。

定理 2. そのような賭けが定数倍を除いて最も有効な賭け方である。

定理 3. 最初の n 桁が、 sn 桁まで圧縮できるとすると、 n 回の賭けで所持金は約 $2^{(1-s)n}$ 倍にできる。

- 圧縮できないなら、 $s = 1$ で所持金は増やせない。
- 完全に規則的なら、 $s = 0$ で所持金は 2^n 倍にできる。
- その間なら、その間くらい増やせる。

規則を見つけよう（再）

規則を見つける どうしたら儲かるか

まとめ

- 科学は完全に予測可能なもの以外でも表現できる。
- どこまで予測可能かとは、どこまで規則が見つけれられるかということ。
- 規則があると信じよう。
- 予測が難しい世の中だからこそ、予測ができる部分（規則がある部分）を、確実に増やしていきましょう。

終り

規則を見つける どうしたら儲かるか

ありがとうございました。