汎用人工知能 RUSAGI

内容

はじめに

RUSAGI は、どんな問題でも、計算に時間を掛けるほど、どこまでも予測精度が上がり続ける汎用 AI である。以下、汎用人工知能とはどういうものかの説明から始め、まず、少量のﾃﾞｰﾀでうまくいかない原因である過剰適合について、k 近傍法を例にして解明していく。帰納のための新しい確率の理論を使って、最適な予測とは何かを導き、ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀを消去する。さらに、他のあらゆる誤差の原因や、汎用性のない原因についても、最小化する方法を示していく。さらに、再帰的な選択によるフレーム問題の解決法を示す。また、汎用 AI に必要な要件を示す。最後に、どのような段階を経て汎用 AI を実現するのか、RUSAGI のロードマップについて述べる。

汎用人工知能とは？

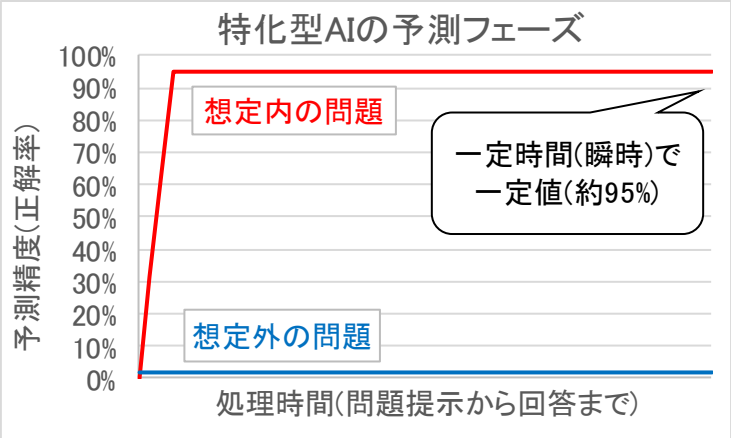
ほぼ同義の呼称	一般的な分類法
汎用 AI, 汎用人工知能, AGI, Artificial General Intelligence, General AI, 広い AI, 強い AI, ASI, Artificial Super Intelligence, Super AI, 超 AI, 人工超知能	**【実力による分類】** チューリングテストや、様々な試験で、人間と同レベルなら汎用 AI。脳と同じアルゴリズムである必要はなく、特化型 AI の寄せ集めでも良い。人間を超える成績なら超 AI。 **開発方針**：特化型 AI の改良による延長線上
	【脳を基準とする分類】 脳と同様のアルゴリズムであるものが汎用 AI。特に、意識があるものは強い AI。脳より優れたアルゴリズムであれば超 AI。 **開発方針**：脳を模倣する
	【理論的な一般性・普遍性による分類】 理論的に、どんな問題でも普遍的に解くことができる一般化されたアルゴリズムが汎用 AI。ハイパーパラメータが存在しない。この分類では脳も厳密には汎用 AI ではない。 **開発方針**：知能の本質を理論的に解明する **（汎用人工知能 RUSAGI はここ）**
(単に)AI, 人工知能, 特化型 AI, 狭い AI, 弱い AI	上記以外の AI。 特定の問題向けにハイパーパラメータ※が最適化。 ※ニューラルネットワークのモデル・活性化関数・学習率。ディープラーニング(深層学習)の層数。サポートベクタマシン等の機械学習のカーネル関数。k 近傍法の k。ランダムフォレストの数等。

特化型 AI と汎用 AI(超 AI)の比較

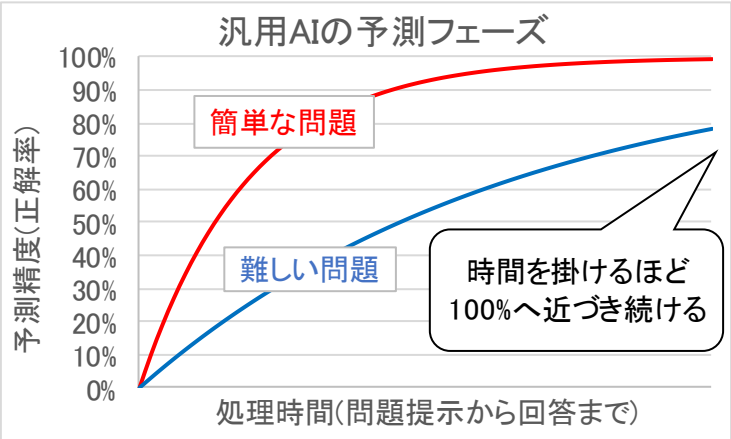
学習フェーズ

特化 AI	学習用データから、説明変数(例えば画像)と目的変数(例えば名称ラベル)間の規則性を導出する。ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀによって、学習結果は変わる。ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀは、特定の問題で結果が良くなるように、あらかじめ調節する(最適化)。 **処理時間：長**（調節次第である程度、精度 UP するが処理時間も UP）
汎用 AI (超 AI)	基本的には学習データを記憶するのみ(怠惰学習)。ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀが存在せず調整不要。特定の問題に適した方法で規則性の導出はせず、問題が与えられてから予測フェーズにて規則性の導出をする。 **処理時間：不要**（予備計算で、予測フェーズの処理の軽減は可能）

予測フェーズ



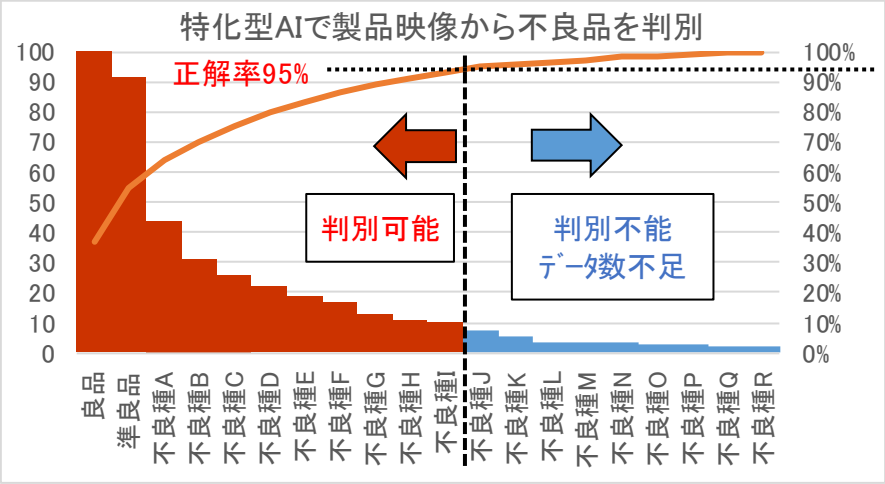
特化型 AI は、問題(例えば名称ラベルの無い画像)が与えられると、瞬時(一定時間)に予測値が得られる。精度(正解率)は約 95%程度。さらに予測精度を上げたい場合は、ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀを調整した上で、再度学習フェーズからやり直さなければならない。



汎用 AI は、学習フェーズがなく、予測フェーズ内で、説明変数-目的変数の規則性の探索を行う。そのため、一定時間で十分な精度が得られなくても、そこから更により良い規則性を追

加で探索し続けられる。難しい問題ほど時間が掛かるが、どんな問題でも、時間を掛ければ掛けるほど、より良い回答を探し続け、精度は 100%へ近づき続ける。

正解率 95%で十分、不十分？



特化型 AI で製品映像から不良品を判別する例を考える。各不良種の件数は、パレートの法則(80:20 の法則)のように、ロングテール型となる。合計件数は多くても、テール部分の不良種の件数は少ない。正解率 95%だとしてもテール部分はほぼ判別不能となる。テールの不良も判別するには、データを増やすしかなく、十分なデータ(数十件以上)が集まるまでは使いものにならない。一方、熟練した人間なら、判別に自信がなければ、納得のいくまで僅かしかない不良のサンプルとじっくりと見比べて判別できるだろう。汎用 AI なら、その熟練者と同様に、決定的な判断材料となる特徴(規則)が見つかるまで、データを解析し続けることができる。

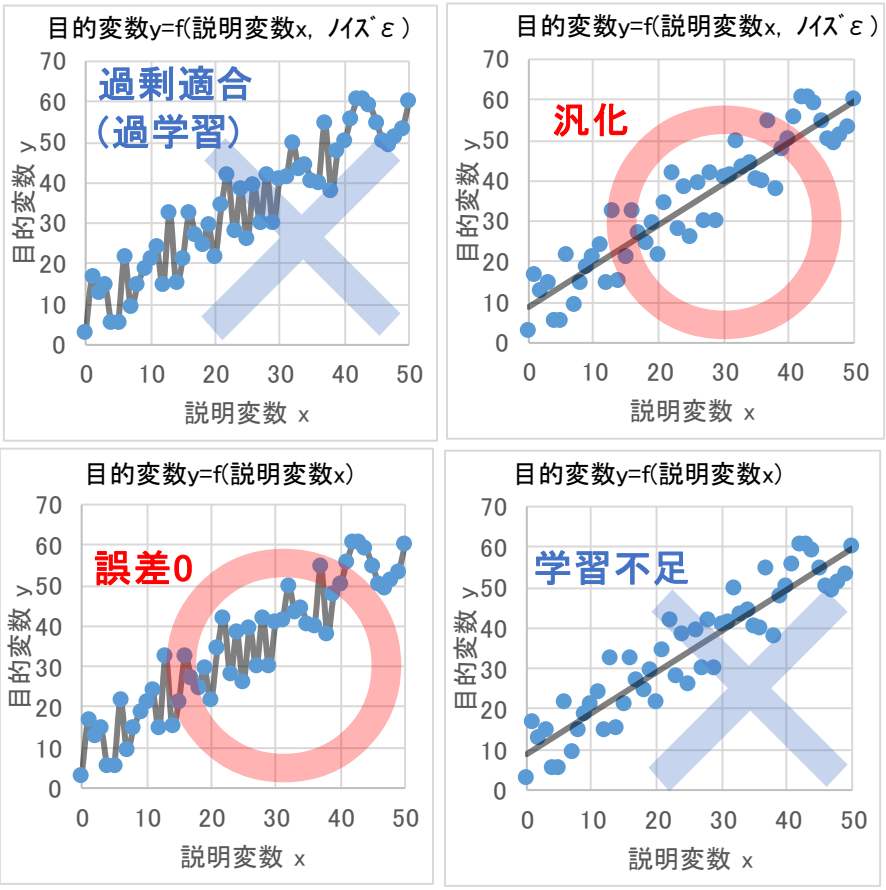
向いている用途

特化 AI	・多少間違えても良いので、瞬時に予測結果が欲しい。 …商品のレコメンド。人間によるﾀﾞﾌﾞﾙﾁｪｯｸのある不良の判別。 ・予測したい分類が限られており、十分なデータが用意できる。 …手書き文字認識。特定の病変のみの画像判別。
汎用 AI (超 AI)	・多少時間が掛かっても良いので、できるだけ予測精度を上げたい。 …株価予測。翻訳。 ・十分な学習データがない場合でも、人間並みの予測をしたい。 …不良の判別。自動運転。

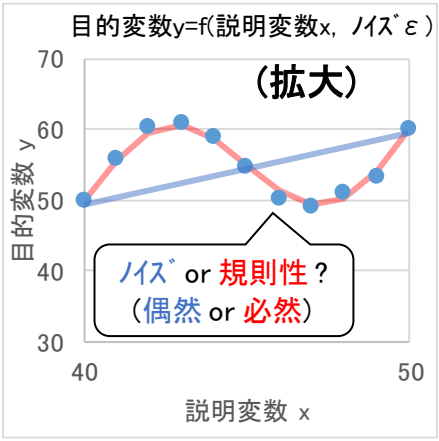
過学習、過剰適合(ｵｰｱﾊﾞｰﾌｨｯﾃｨﾝｸﾞ)

汎用 AI はどうして少ないデータでもうまくいくのか説明する前に、特化型 AI は少ないデー

タでは何故うまくいかないか説明する。その主要因は、過学習・過剰適合(ｵｰﾊﾞｰﾌｨｯﾃｨﾝｸﾞ)と呼ばれる現象にある。



左上のように、ノイズも含めて学習しまうのが過学習である。右上はノイズを除いた本来の姿で、学習に使用していないデータにも適合する汎化した状態である。しかし、下二つは、上二つと全く同じグラフの形をしているが、逆に左下の方が優れる。上二つ「目的変数 y=f(説明変数 x, ﾉｲｽﾞ ε)」なのに対して、下二つは「目的変数 y=f(説明変数 x)」でありノイズはないと仮定している。ノイズが無いということは、x が決まれば、y が決まる。既にデータがある x のところに、新しいデータが来たとき、その y は、既にある y と一致する。したがって、左下は全ての点を通るため、誤差 0 なのに対し、右下は通らないため学習不足といえる。



ノイズ以外だけを学習するのが理想だが、このグラフを拡大すると、部分的に正弦波のような形になっている。これはノイズのためで、偶然だとすれば直線を引いて良いのだが、このデータだけから偶然なのか必然なのかは区別できない。

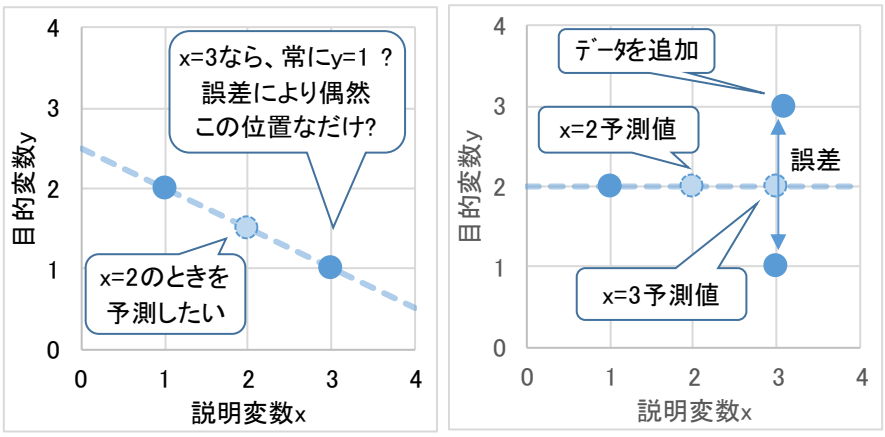
過剰適合の対策法

複雑さの制限(正則化)	例えば、n 次曲線で近似する n 次数を減らす。正則化という手法で、適切な複雑さにすることができる。しかし、前述のｸﾞﾗﾌのように、部分的に正弦波だけど、ほぼ直線的な場合、直線近似してしまうと、曲線部分は推測できない。単純化すれば、複雑なものは推測できなくなる。あらかじめ、直線になることが分かっているような場合には適する。回帰分析での説明変数の数、ﾆｭｰﾗﾙﾈｯﾄﾜｰｸの層数、決定木の深さ等の制限が、この手法に該当。
平均化	平均する数を増やすほど、ﾉｲｽﾞの影響は 0 に近づく。例えば、k 近傍法の k を増やす。しかし、説明変数が遠いところにあるものまで平均に含まれるようになるため、予測値がボヤけてしまう。少ない事例のﾃﾞｰﾀは、多い事例のﾃﾞｰﾀで薄められてしまう。平均数を増やすなら、ﾃﾞｰﾀ数も増やさなければならない。ﾆｭｰﾗﾙﾈｯﾄﾜｰｸの学習率（1 個の学習ﾃﾞｰﾀで、どれだけ結合強度を更新するか）を下げるのも、この手法に該当。
交差検証	学習に使っていないﾃﾞｰﾀを使用して、正解率の実力値を調べる。結果が良くなるようにﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀを調節する。

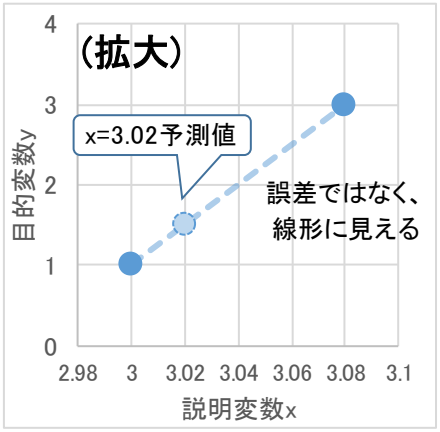
・複雑なモデル：精密に予測できるが、過学習しやすい
　↑↓トレードオフ
・単純なモデル：過学習しにくいが、複雑なものは予測できない(学習不足)

・平均化を減らす：少量のデータで学習するが、過学習しやすい
　↑↓トレードオフ
・平均化を増す：過学習しにくいが、大量のデータが必要

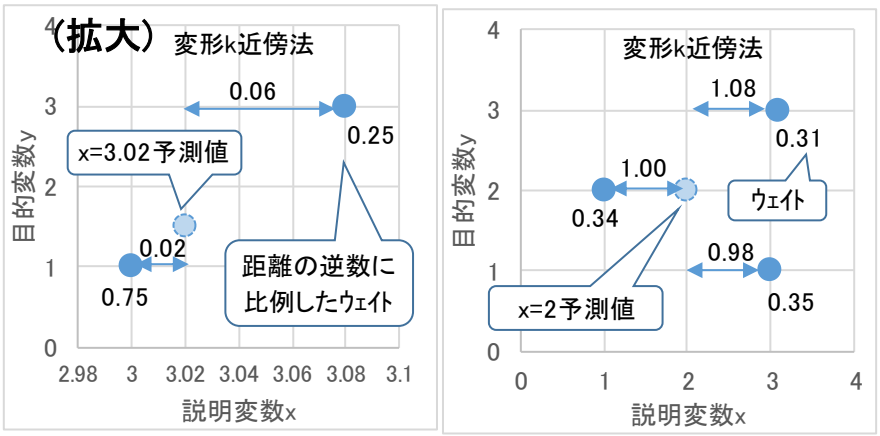
ノイズ(誤差)か否か



もし、ﾊﾞﾗついてるように見えるﾃﾞｰﾀが、ﾉｲｽﾞか否かさえ分かれば、最適な予測ができる。上の例では、x=3 のときの、常に y=1 なのか、ﾉｲｽﾞ(誤差)によりたまたま y=1 なのか知りたい。最も簡単な検証方法は、x=3 の場合のﾃﾞｰﾀを追加で採取することである。x が同じなのに、y が異なる値なら、明らかに誤差である。



x が連続値なら厳密に同じ値はありえず、ｸﾞﾗﾌを拡大し続ければ、差が見えてくる。このｸﾞﾗﾌのように、x=3.02 のときの y を求める際は、2 点の差は誤差とみなすより、線形補完した方が良さそうに見える。



変形 k 近傍法で、距離に応じてうまくｳｪｲﾄを掛けて平均すれば、線形補完で、x=3.02 のときの

予測ができる(左上図)。また、同じ計算で、x=2 のときも予測できる(右上図)。この方法なら誤差(ﾉｲｽﾞ)か否かを区別することなく予測できてしまう。yの差が一定なら、xの差が小さいほど、誤差のように感じるだけで、同じ方法で計算できる。

・バラツキか誤差(ノイズ)か否かは、予測対象との距離で変わるため、予測対象次第で最適な回帰曲線も変わる。
・予測対象が与えられてから計算する k 近傍法のような怠惰学習なら、誤差か否か区別する必要がない（汎用 AI もこちら)。

変形 k 近傍法におけるﾃﾞｰﾀの質と量

k 近傍法は、予測対象と距離が近い順に k 個のﾃﾞｰﾀを選び平均値を予測値とする。原理的には、y=f(x)の x が近いほど、y も近いだろうという予想に基づく。関数の形が未知でも、他に情報がなければ、真逆の予想より同等以上と推測できる。距離が近いﾃﾞｰﾀのほど、目的変数も近そうな、高品質のﾃﾞｰﾀといえる。最近接ものが最高品質だが、誤差の影響を減らすため、より低品質なものも含めて k 個平均する。予測に使うﾃﾞｰﾀの品質と量はトレードオフの関係にある。k は大きすぎても小さすぎてもダメで、経験的な値か、実力値を調べて決められる。

	予測に使うﾃﾞｰﾀの量	予測に使うﾃﾞｰﾀの質
k が小さい	少量	高品質
k が大きい	多量	低品質
	多量の方が良い	高品質の方が良い

変形 k 近傍法は、距離の逆数をｳｪｲﾄとして平均する。より近いものが重視されるため、k が同じなら、単純平均より距離の平均値が近くなる。この方法でも、k の値は指定する必要がある。計算量の問題ではなく、ｳｪｲﾄの小さい遠いﾃﾞｰﾀでも、増えるほど、ｳｪｲﾄの合計値は増え続け、距離の平均値も増え続けてしまう。また、距離が 0 のﾃﾞｰﾀがあった場合、ｳｪｲﾄは逆数の∞となり、そのﾃﾞｰﾀ 1 つをそのまま予測値とすれば良いことになる。しかし、ﾉｲｽﾞがあれば、説明変数が同じでも目的変数が同じとは限らないため、この重みの付け方は適切とはいえない。

・予測に使うデータの品質(近さ)と量はトレードオフ
・距離の逆数をｳｪｲﾄとして平均しても、質と量の関係は最適値ではない

帰納

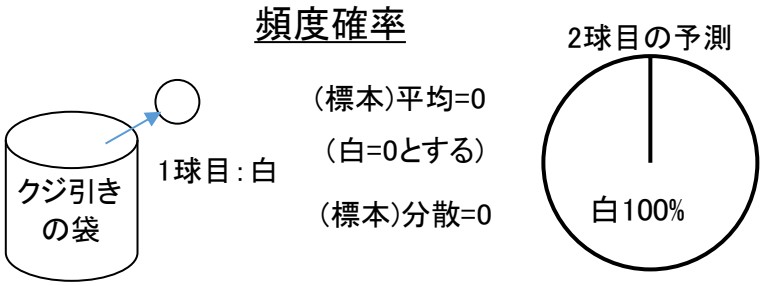
予測に使うﾃﾞｰﾀの質と量の最適な関係を導くため、帰納についての理論を 1 から構築する。まず帰納というのは、類推とほぼ同義である。例えば、画像認識では、全く同じ画像が学習ﾃﾞｰﾀになくても、画像が似ていれば、名称ﾗﾍﾞﾙも同じだろうと推測する。画像は似ている方が良く、似た画像の数が多い方が良い。原理的には、確証性の原理と呼ばれ、「関連する観測が多いほど、確証が増す」とされる。しかしこれは、どんな些細な情報でも無いよりはマシと言って

いるのに過ぎない。「良く似ているﾃﾞｰﾀが少量」と「やや似ているﾃﾞｰﾀが多量」のどちらが優れているか示されていない。

・「高品質、少量」←どちらが良いか分からない→「低品質、多量」

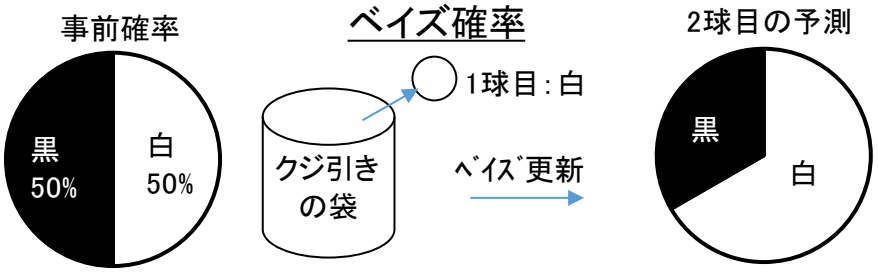
頻度確率とベイズ確率と新しい確率

ﾃﾞｰﾀが少ないと、うまく予測できないﾒｶﾆｽﾞﾑを、簡単な例で考えていく。ｸｼﾞ引きで、袋から取り出した球の色を予測する場合を考える。ﾃﾞｰﾀの品質は一定だと仮定する。すなわち、球を取り出すﾀｲﾐﾝｸﾞ等は、色に影響しないとする。1 球目が白球で、2 球目の色を予測する場合を考える。



黒球が出た場合：予測は間違っていた

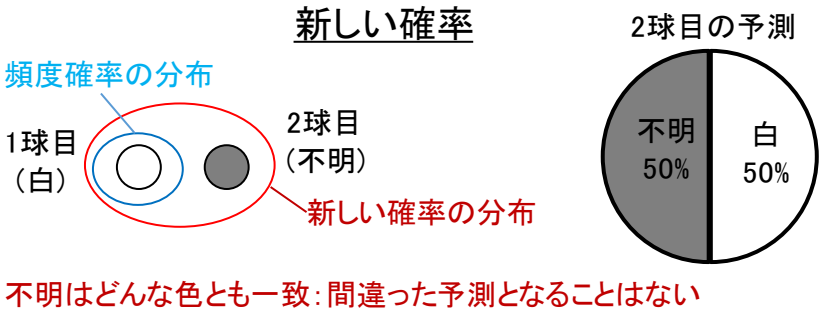
古典的な頻度確率の考え方では、白球を 0、黒球を 1 とした場合、平均値=0、分散=0 と予想できる。さらに球を取りだし続けても、同じ色が出続ける限り、分散=0 となる。次も 100%同じ色が出ると言い切ってしまうのは明らかにおかしい。次に別色が出たら、その予測は誤りだったことになる。



白、黒以外が出た場合：予測(事前確率)は間違っていた

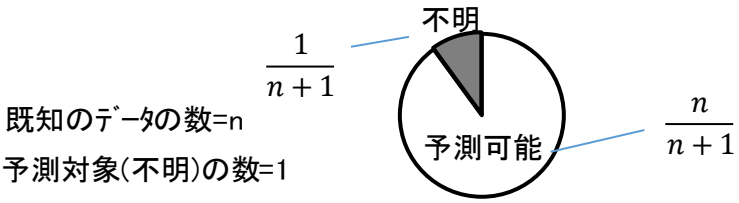
一方、ベイズ確率では、1 球も取り出していない状態で、白 50%、黒 50%という事前確率を仮定する。白を 1 球取りだした後は、白の確率が高めに更新される。仮に 2 球目が黒だったとしても、低確率のことが起こっただけで、予測が間違っていたとはいえない。しかし、白か黒のみと分かっていれば良いが、それ以外の色が出たら予測は外れたことになる。どんな色が入っているか分からければ､あらゆる色が均等に入っていると仮定しなければならない｡しかし、均等といっても、クジで 1 等・2 等・3 等賞があるとして、さらに 1 等 A 賞と 1 等 B 賞に分かれている場合、AB を分けて考えるかどうかというサジ加減で結果が変わってしまう。事前確

率が無いと計算できないため、ｸｼﾞを一つも引いていない状態でも、しぶしぶ「全ての当たりが等確率だ」と予測するしかない。しかし、予測に使える材料が一切ないのなら、「不明」という方が正しいのではないだろうか。同じ「全て等確率」という予測結果でも、全色の球が同じ数だけ取り出されてそう予測したのか、それとも、ﾃﾞｰﾀがないから適当に答えているだけか区別できない。サジ加減で決めた不純な情報なら、予測結果に混入しない方が良い。

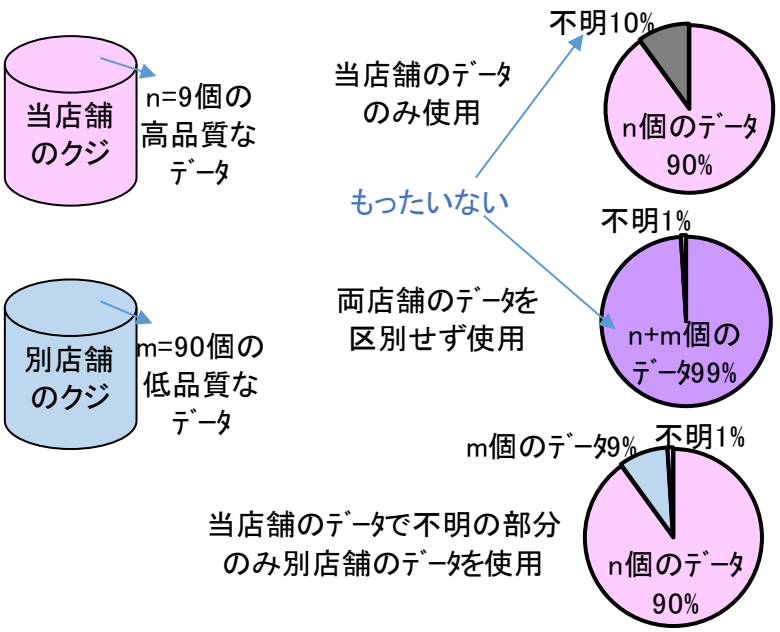


「不明」なものを「不明」なまま扱う方法を考える。頻度確率では、既に取り出した球のみで統計するが、これから取り出そうとしている球も統計に含めてしまおう。その球の色は「不明」という値とする。白玉 1 球取りだされた後に、次の 2 球目を予測する場合を考える。「白」：50%、「不明」：50%と予測できる。「不明」はどんな色とも一致するため、取り出した球がどんな色であっても予測は誤っていない。どんな色が入っているか事前確率を知っている必要もない。頻度確率やベイズ確率は球の色を観測すると前後で確率分布が変化するが、この新しい方法では観測によって「不明」が明らかになるだけで、確率分布は変化しない。

予測能力の限界



球の色を予測する例で、色が未知(不明)の球(予測対象)の数を 1、色が既知の球の数を n とする。「不明」の割合は、1/(n+1)となる。「不明」ではない部分は、n/(n+1)となる。すなわち、ｻﾝﾌﾟﾙ数が n 個なら、どう頑張っても割合として n/(n+1)しか予測する能力がない。ﾃﾞｰﾀの品質が一定なら、ﾃﾞｰﾀの数が減るほど、予測能力の限界が減ることが示された。



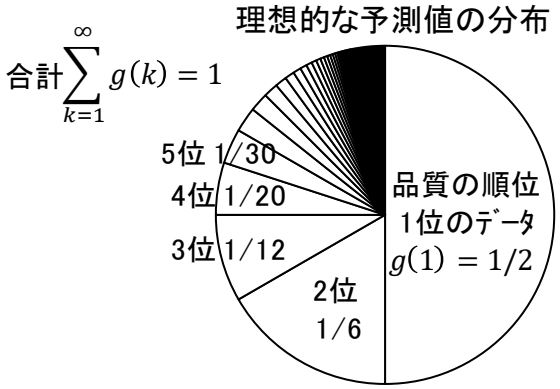
次に、品質の異なるﾃﾞｰﾀが混在する場合を考える。ある店舗で行われているｸｼﾞの結果を予測する。球の色が既知のﾃﾞｰﾀが、当店舗について n 個、別店舗について m 個あるとする。店舗によって当たり確率が同じかもしれないし違うかもしれない。当店舗のﾃﾞｰﾀは高品質なﾃﾞｰﾀ、別店舗のものは低品質なﾃﾞｰﾀといえる。感覚的には、当店舗のﾃﾞｰﾀが十分ならそれだけで統計するが、当店舗のﾃﾞｰﾀが不十分なら、別店舗のﾃﾞｰﾀも使うべきだろう。両方のﾃﾞｰﾀを使うとしても、当店舗のﾃﾞｰﾀが重視されるべきと感じるだろう。n=9、m=90 の場合を考えてみよう。当店舗のﾃﾞｰﾀでは、n/(n+1)=90%まで予測できる。当店舗と別店舗のﾃﾞｰﾀを両方使う場合は、(n+m)/(n+m+1)=99%まで予測能力がある。品質の悪いﾃﾞｰﾀも使うことで予測能力の限界が上がるﾄﾚｰﾄﾞｵﾌの関係にある。ここで、ﾃﾞｰﾀの品質に差があるので、予測能力の限界値が高いほど、予測精度が高いわけではないことに注意。次に、当店舗のみのﾃﾞｰﾀと、両店舗のﾃﾞｰﾀで予測した結果を組み合わせて、最適な予測を考えよう。当店舗のみで 90%まで予測できるが、残り 10%を「不明」としてしまうのはもったいない。別店舗のﾃﾞｰﾀも使えば 99%まで予測できるのだから、99%-90%=9%分は、別店舗のﾃﾞｰﾀも使った予測値を付け加えてしまおう。ただし、この 9%分には別店舗のﾃﾞｰﾀのみを使う。当店舗のﾃﾞｰﾀを 90%部分と 9%部分の両方で重複して使えば、ﾃﾞｰﾀ 1 個当たりの予測能力の限界を超えてしまうからである。最終的に、当店舗のﾃﾞｰﾀは 1%/個×9 個、別店舗のﾃﾞｰﾀは、0.1%/個×90 個、「不明」1%で、合計 100%の予想になる。

理想的帰納

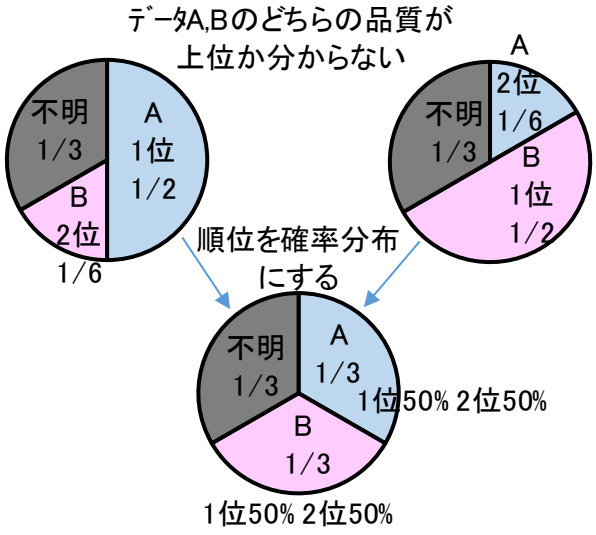
ﾃﾞｰﾀ数がxのときの最大予測能力　$f(x)=\frac{x}{x+1}$

品質の順位がxのﾃﾞｰﾀが、予測値を占める割合

$$g(x)=f(x)-f(x-1)=\frac{x}{x+1}-\frac{(x-1)}{(x-1)+1}=\frac{1}{x(x+1)}$$

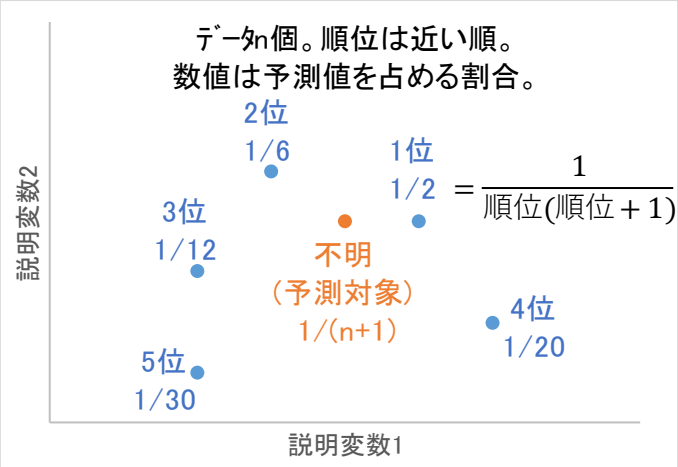


前述のｸｼﾞの例では、ﾃﾞｰﾀの品質が 2 水準のみであったが、一般化した場合を考える。ﾃﾞｰﾀが n 個あり、品質が良い順にｿｰﾄされていると考える。品質上位 i 個のﾃﾞｰﾀのみ使う場合、i/(i+1)の予測能力がある。関数 f(x)=x/(x+1)と置けば、f(i)と表記できる。ここで、i 番目ﾃﾞｰﾀが予測結果を占める割合は、より高品質の i-1 番目のﾃﾞｰﾀで予測できなかった分に限られる。従って、f(i)-f(i-1)が、i 番目のﾃﾞｰﾀが結果に占める割合となる。関数 g(x)=f(x)-f(x-1)と置けば、g(x)=1/n(n+1)となる。すなわち、品質上位 i 番目のﾃﾞｰﾀが、予測結果の 1/i(i+1)割を占める。1 つ目が 1/(1・2)=1/2、2 つ目は、1/(2・3)=1/6、3 つ目は、1/(3・4)=1/12 と続き、無限個足すと 1 になる。この方法では、ﾃﾞｰﾀの数を増やすほど、必ず精度が上がり、決して下がらない。なぜなら、ﾃﾞｰﾀが増えるほど不明部分が明らかになっていくからである。「不明」はどんな予測結果よりも精度が悪い。また、どんな精度の悪いﾃﾞｰﾀでも、無いよりは僅かにマシになるため、確証性の原理と合致している。



ここまでは、全ｻﾝﾌﾟﾙを品質順にｿｰﾄしていたが、同品質のﾃﾞｰﾀが複数ある場合が考えられる。しかし、同品質のように見えて、実際には優劣があるか、どちらが優れるか見分けられていないだけと解釈できる。全てのｻﾝﾌﾟﾙは、品質順位の確率分布を持つとすればよい。ｻﾝﾌﾟﾙ A が 1 位、B が 2 位の場合、予測結果は A：1/2、B：1/6、不明：2/3、順位が逆なら、A：1/6、B：1/2 となる。AB ともに、1 位に確率が 50%、2 位の確率が 50%なら、先の 2 つを平均して、A：1/3、B：1/3、不明：1/3 となる。AB を同品質として考えた場合と一致する。実際にどうやって品質を決めるかというと、ｸｼﾞの例では、当店舗のｸｼﾞ結果の確率分布と、別店舗のｸｼﾞ結果の確率分布を比較することになる。全く同じ分布なら品質は等しい。全く別の分布なら、明確に優劣を付けられる。

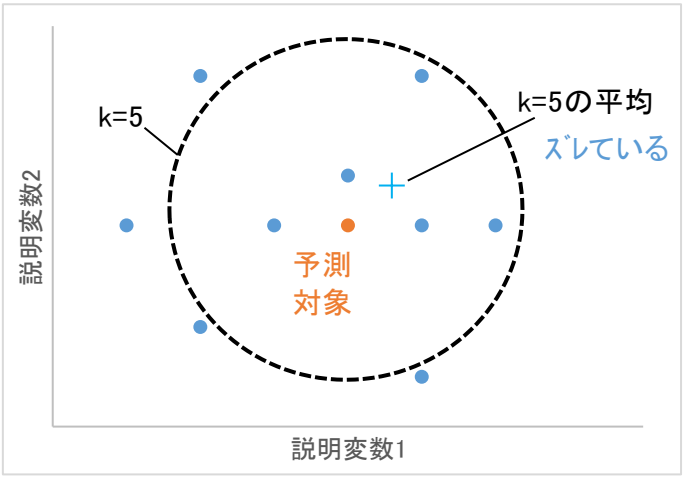
ﾊﾟﾗﾒｰﾀのない k 近傍法



理想的帰納を応用して k 近傍法のﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀ k を消去する。まず n 個のﾃﾞｰﾀを、予測対象との距離が近い順にﾃﾞｰﾀをｿｰﾄする。1/i(i+1)の確率で、i 番目のﾃﾞｰﾀと同じ値になると予測する。k 番目までの確率の合計は、k/(k+1)で、残りの 1/(k+1)は「不明」と推測する。k が増すほど精度が増し、k=n とするのが最適だが、途中で打ち切ることもできる。「不明」があるため、厳密には期待値は求められない。近似的に期待値を求めるには、「不明」の確率が一定のところで、「不明」に適当な値を割り当てるか、「不明」を除いて 100%になるよう正規化する。

距離によって順位を決めているが、ﾉｲｽﾞの影響を受けないよう、僅かな距離差で、順位が入れ替わるべきではない。距離がほぼ同等なら、ｳｪｲﾄもほぼ同等でなければならない。順位を確率分布とすれば良い。確率分布の計算の仕方としては、距離差が 0 に近づくほど、順位の確率分布の差も 0 に近づくという性質を満たす必要がある。

説明変数の平均値の差による誤差



k 近傍法は、予測対象と距離が近いほど、ｳｪｲﾄを重くするのは、必ずしも最適ではない。例えば、近いものから順に k 個選んだ結果、説明変数が＋,-側どちらかに偏っているかもしれない。近い順の k 個の説明変数の平均値は、予測対象の説明変数の値とは必ずしも一致しない。説明変数と目的変数に相関性があるなら、説明変数の平均値は、予測対象の平均値と近い方が良い。ｳｪｲﾄを調節して加重平均にすることで、説明変数の平均値を一致させた方が良いと考えられる。しかし、説明変数の+側にﾃﾞｰﾀが多く、-側は僅かな場合、-側のﾃﾞｰﾀは距離が離れていても、大きなｳｪｲﾄを占めさせねば、平均値を一致させられない。そのため、必ずしも平均値を一致させるのが最適ではない。また、例えば説明変数が時刻で、未来を予測する場合、過去のﾃﾞｰﾀしかないため、どうがんばっても時刻の平均値を一致させることができない。平均値を近づけようとするほど、近いものほどｳｪｲﾄが大きいべきという関係が崩れるため、どこかでﾊﾞﾗﾝｽを取る必要がある。

ここで、ﾊﾟﾗﾒｰﾀのない k 近傍法の場合を考える。この方法は、距離順の各順位のﾃﾞｰﾀが、理論的に持つ予測能力の最大になるようにｳｪｲﾄが決められる。つまり、順位から決められたｳｪｲﾄを上方修正するのは許されない。そこで各ﾃﾞｰﾀのｳｪｲﾄを下方修正のみで合わせられる範囲で、できるだけ説明変数の平均値を近づければよい。具体的には、各ﾃﾞｰﾀの個数を 1 個とするのではなく、0～1 個の範囲で調節する。一部分のﾃﾞｰﾀを偏った形で作為的に取り出しているように見えるが、全く問題ない。なぜなら、手元のﾃﾞｰﾀは、母集団から抽出した時点で既に偏っているのだから、偏ったままにせず、間引いても良い。

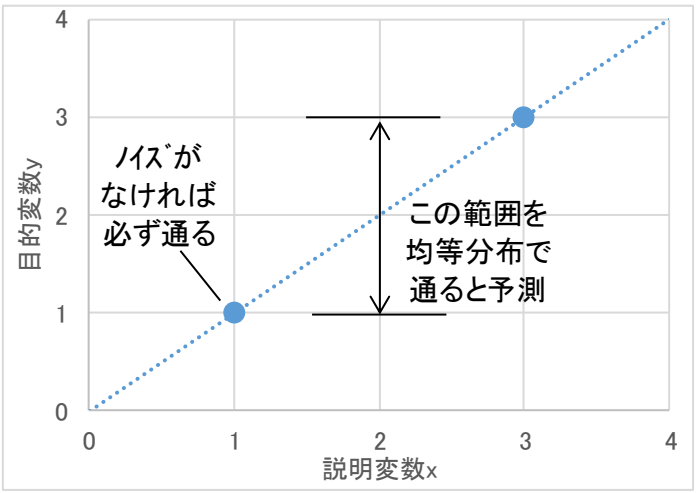
非線形性による誤差

予測値と実測値の差の成分

・説明変数で決まらないﾉｲｽﾞ成分
・予測に使用したﾃﾞｰﾀの説明変数(平均値)と予測対象の説明変数の差
・線形予測に対する実際の非線形との差

k 近傍法で、説明変数の平均値を合わせるだけでは、まだ最適ではない。k 近傍法における、予測値と実測値のズレは、3 つに分解して考えることができる。1 つ目は、ﾉｲｽﾞ成分。説明変

数が全く同じでも、目的変数が変動する分が、ﾉｲｽﾞ成分である。2 つ目は、予測に使ったﾃﾞｰﾀの説明変数の平均値と、予測対象の説明変数の値の差である。ここで、ﾉｲｽﾞが無く、説明変数の平均値も一致している場合を考える。予想対象の説明変数 x=0 とする。このとき、x=+1 と x=-1 の 2 点を平均した場合と、x=+2 と x=-2 の 2 点を平均した場合のどちらも x の平均値は 0 である。もし、y=x のような線形であるとしたら、両者とも正解が得られるが、非線形かもしれない。



(x,y)=(1,1),(3,3)の 2 点から、x=2 のときの y を予測する場合を考える。ﾉｲｽﾞがないのだから、解が非線形だとしても、必ず(1,1),(3,3)の 2 点を通る。x=1～3 の間では、どのような曲線になっているか分からないが、この区間では y=1～3 と予測できる。実際にはこの範囲外の可能性もあるが、y＞1、y＜3 の根拠となるﾃﾞｰﾀがない。名義変数で考えると分かり易く、白色・黒色のﾃﾞｰﾀしかない場合に、その中間の色を予測することはあっても、金色と予測することはない。今までに存在しない値が現れる確率は、「不明」の方に含まれるため、無視することができる。よって、x=2 のとき、解が y=1～3 の均等分布と仮定すると、y=2 と予測すれば、平均で y=±0.5 ズレることになる。ここで、予測対象と 2 点との平均距離は 1 であり、また傾き 1 として予測している。0.5*平均距離*傾きが、非線形性による誤差と予測できる。

x_0 ：予測対象の説明変数

x_k ：ﾃﾞｰﾀkの説明変数　　w_k ：ﾃﾞｰﾀkのｳｪｲﾄ

$$\beta = \sqrt{\underbrace{\left(\left|\frac{\sum_{k=1}^{n} w_k x_k}{\sum_{k=1}^{n} w_k} - x_o\right| \times \alpha\right)^2}_{\text{説明変数の差による誤差}} + \underbrace{\left(\frac{1}{2} \times \frac{\sum_{k=1}^{n} |w_k x_k - x_0|}{\sum_{k=1}^{n} w_k} \times \alpha\right)^2}_{\text{非線形性による誤差}}}$$

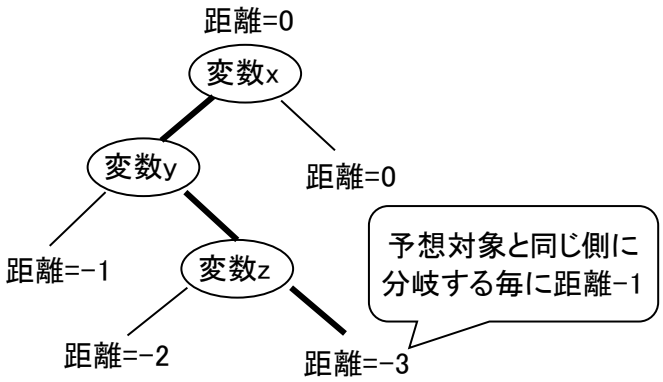
上式 β を最小化するようにｳｪｲﾄを設定すれば良い
傾き α は知らなくてよい

誤差の成分の内、説明変数の平均値のズレによる誤差は、|予測対象の説明変数-利用ﾃﾞｰﾀの説明変数の平均値|*傾きである。また、非線形性による誤差は、|予測対象の説明変数-利用ﾃﾞｰﾀの説明変数|の各利用ﾃﾞｰﾀについての平均値*傾き*0.5 である。これら 2 つの誤差の和(2 乗和平方根)を最小化するように、利用ﾃﾞｰﾀを選ぶのが最適な予測となる。どちらも傾きは同じため、

傾きは未知でも計算できる。目的変数をみる必要が無く、説明変数のみで最適な予測ができる。また、この方法はﾉｲｽﾞ成分を区別する必要が無い手法であるため、全ての誤差が最小化される理想的な方法である。ただし、最適解を得るには、距離が正しく決められている必要がある。

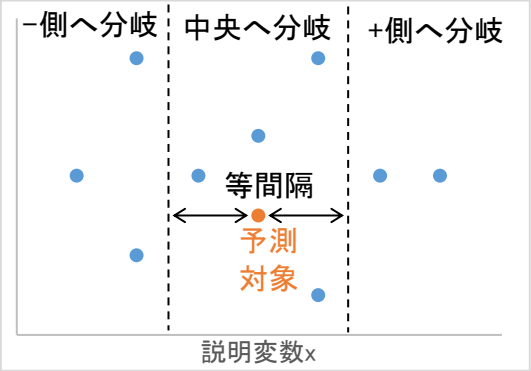
決定木による距離決め

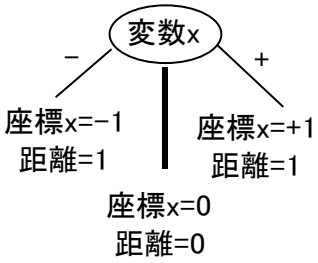
k 近傍法では、説明変数が複数ある場合は、主にﾕｰｸﾘｯﾄﾞ距離が使われるが、変数毎にうまく尺度を調節しなければ、距離＝似ている度合にならない。説明変数毎に平均 0、分散 1 への正規化は必ずしも最適ではない。関係のない変数ならば、分散が 0 になるようにするべきである。k 近傍法は、どの説明変数をどのくらいの尺度で使うかという人間のｻｼﾞ加減で距離が変わってしまう。この説明変数が効くだろうという人間の先入観が入ってしまうが、そのような事前情報の混入は精度を悪化させるだけだ。ｻｼﾞ加減に頼らず、実際にどの説明変数が目的変数に効くか調べて、距離を決めるべきだ。そのような手法の一つが決定木だ。



決定木は k 近傍法の距離を決める手法と解釈できる。初めに全てのﾃﾞｰﾀが距離 0 にあり、予測対象と逆側に枝分かれしたﾃﾞｰﾀは距離を無限遠まで離す。最後まで残った距離 0 で最近傍ﾀｲの k 個のﾃﾞｰﾀの平均値が予測値となる。枝分かれのさせ方が良ければ、k 近傍法より良い結果となる。ﾗﾝﾀﾞﾑﾌｫﾚｽﾄという手法もあるが、悪い木が混じってしまうので最適値には至らない。

ﾊﾟﾗﾒｰﾀのない k 近傍法を決定木に応用する。最初に全ﾃﾞｰﾀは距離 0 とする。分岐する毎に、予測対象と同じ側に分岐したﾃﾞｰﾀの距離を-1 する。逆側に分岐した場合は距離を維持。最後まで残ったﾃﾞｰﾀの距離が最小となる。あとは、この距離を使って、ﾊﾟﾗﾒｰﾀのない k 近傍法を行えばよい。距離の近い順に順位を付けると、同順位に複数個のﾃﾞｰﾀが存在する。順位 i までのﾃﾞｰﾀが予測できる範囲を f(i)=i/(i+1)とする。現在の順位までに合計 n 個のﾃﾞｰﾀがあり、この次に近いﾃﾞｰﾀが m 個あるとする。それらのﾃﾞｰﾀは、予測値の{f(n+m)-f(n)}/m 割を占めるようにする。これを上位から順に繰り返せばよい。

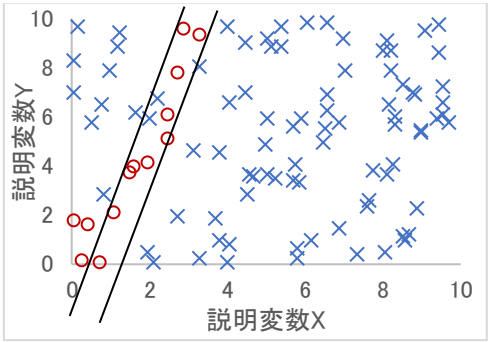




距離の大きさを合わせるだけでは、k 近傍法のときのように利用ﾃﾞｰﾀの説明変数の平均値を予測対象の説明変数に合わせられない。一般的な決定木の手法では、k 近傍法と違って、説明変数のどの値を境界に分岐させるか、事前に決めている。しかし、予測対象の説明変数の値が、境界のすぐそばであるより、十分に離れている方が、うまく分離できるだろう。そこで、予測対象の説明変数の値が分かってから、決定木を行うこと（怠惰学習）を考えよう。2 つではなく、3 つに分けよう。中央の枝に予測対象が入るようにし、+側と-側に同じだけマージンを取って、3 つに分ける。+側と-側に分岐したﾃﾞｰﾀを、同じだけ使うようにするほど、説明変数の平均値は、予測対象の説明変数に近づく。

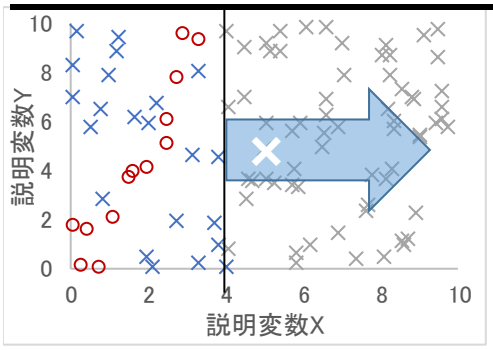
選択肢の絞り込みと再帰的選択

決定木には、ﾃﾞｰﾀの分類を苦手とするﾊﾟﾀｰﾝがある。説明変数が 2 次元であり、平面上にﾌﾟﾛｯﾄされた点を、2 つに分ける場合を考える。縦軸、横軸のうまく分離できる場所を探して境界線を引いていく。しかし、境界線は軸に対して垂直 or 平行にしか引くことができない。斜めにﾃﾞｰﾀが分かれていたら、斜めの線が引けず、ｶﾞﾀｶﾞﾀになってしまう。しかし、縦軸が緯度、横軸が経度であったとして、初めから座標を回転させて斜めにした状態でﾃﾞｰﾀを分離していたらうまくいっただろう。また、決定木は XOR も得意ではないが、初めから 2 変数の差を説明変数としていれば、うまくいく。つまり、決定木がうまく分離できるかは、ﾃﾞｰﾀをどう数値化しているかというｻｼﾞ加減に依存する。しかしながら、座標を回転させても表現の仕方が変わっただけで、本質的な情報(量)は変化していない。そのため、複数の説明変数を組み合わせて新しい説明変数を作り、その変数でうまく分離できないか調べれば、どんなものでもうまく分離できる。しかし、組み合わせ爆発が起こるため、すべて調べ切るのは非現実的だ。



1度に最適解を求める

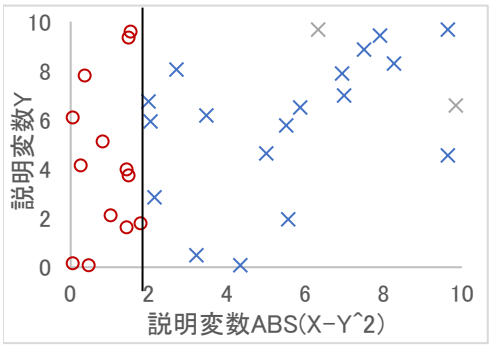
完全に分離できない



徐々に絞り込む

容易に分離できる部分を先に分離

ここで時間切れでもそれなりの精度



時間切れまで、完全に分離できる新たな説明変数を探し続ける

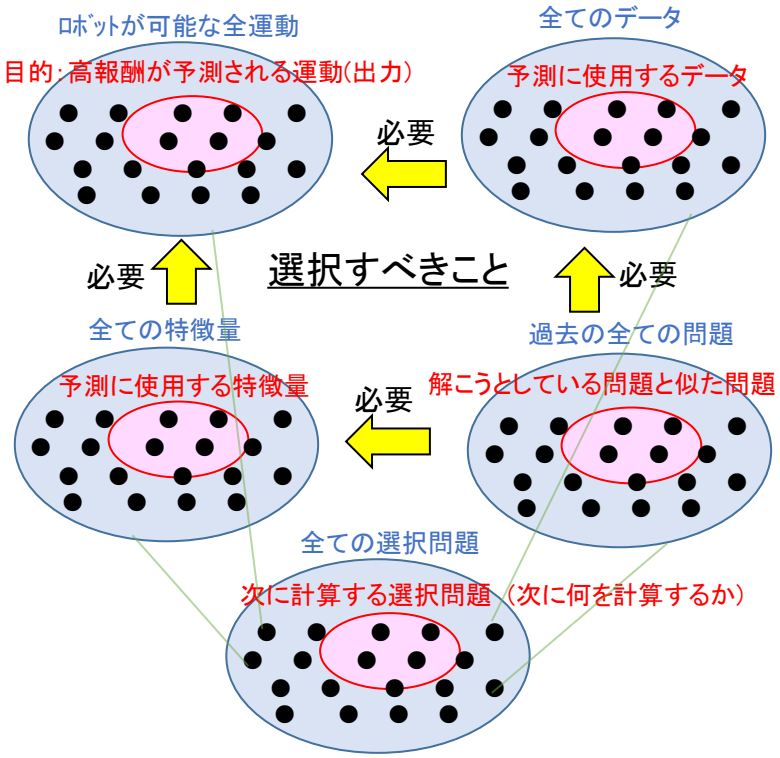
説明変数の組み合わせは無数にあるため、全て計算する時間はない。可能性のありそうなものから優先して調べていく必要がある。ある時点で探索を打ち切り、そこまで調べられた範囲での最適解を答える必要がある。先に、簡単に分離できるﾃﾞｰﾀを分離してしまって、後は時間が許す限り、分離し難いﾃﾞｰﾀの分離を試みるのが良いだろう。こういった段階的な予測は人間も行っている。映像からある人物を認識する例を考えてみよう。まず、映像をﾁﾗっと見た段階で、人か否かを判断する。このとき、脳内のあらゆる映像ﾃﾞｰﾀの中から、人以外のものｳｪｲﾄを 0 にする。さらに女性という特徴が分かれば、男性を 0 とし、個人が特定されれば、その人以外の映像ﾃﾞｰﾀのｳｪｲﾄを 0 とする。最初は全映像のｳｪｲﾄが 1 である。時間とともに、ﾃﾞｰﾀは絞り込まれていき、判断を迫られた時点で、残っているﾃﾞｰﾀが予測値となる。

映像からその名称ﾗﾍﾞﾙを推測する場合、学習ﾃﾞｰﾀの名称ﾗﾍﾞﾙのどれに当たるかを選択することになる。数値を予測する場合は、その数値が取りうる値(例えば実数全体)の内のどれかを選択するといえる。あらゆる推測というものは、取りうる選択肢から一つを選ぶこと解釈できる。最適な 1 つを調べ終えるには、どれだけ時間が掛かるかわからない。あらゆる可能性を考慮して最適解を出そうとすると、計算が終わらず、フレーム問題に陥る。しかし、時間と共に徐々に絞り込んでく形ならば、途中で計算を打ち切っても、そこまで考えた範囲での最適値が得られるため、フレーム問題に陥らない。

ある問題の予測値を選択するためには、どの説明変数を使って分岐をさせるか選択しなけれ

ばならない。過去の事例、似た事例や、数学的なﾃｸﾆｯｸを使用して、うまくﾃﾞｰﾀ分離できそうな方法を選択することになる。それができれば最適な予測ができる。しかし、ある選択をするために、別の新たな選択が必要になってしまっている。例えば、画像の名称ﾗﾍﾞﾙの予測値を選択しようしたら、うまく分離できる特徴量を選択しなければいけない。ただ、どんな問題でも選択問題と解釈できる。新たな生まれた選択問題も、同じ方法で解けば良い。ただし、新しい選択問題を解くには、さらにまた新しい選択問題を解く必要がある。再帰的に、無限に選択問題が発生する。しかしながら、計算不能ではない。ある問題に掛けられる計算時間が 0 であれば、取りうる値の集合全体がそのまま予測値になるだけである。そのとき、特徴量を選択する必要はないため、これ以上、選択問題が増えることはない。再帰的に普遍的な選択をする汎用人工知能であることから、Recursive Universal Selective Artificial General Intelligence の頭文字をとって RUSAGI と呼ぶ。

目的と割引率除去



選択肢が∞にあっても、時間がなければ途中まで絞り込んだ中から無作為に選択するため停止しない…ﾌﾚｰﾑ問題解決

知能とは、「目的に沿って選択肢を絞り込む能力」と定義(解釈)できるが、汎用 AI は、何を選択しているか図示する。強化学習ｴｰｼﾞｪﾝﾄと同様に、初めに人間が指定するのは、入力と出力と目的（報酬の定義）である。AI は、報酬が大きくなるように出力を選択する。例えば、次にﾛﾎﾞｯﾄがどう運動するか、可能なあらゆる運動ﾊﾟﾀｰﾝの中から選択する。どの行動を選択するのが良いかの予測に、入力ﾃﾞｰﾀを参照できる。ただ、現在の入力値に限らず、過去全ての入力値や、過去全ての出力値を参照することができる。しかし、全てのﾃﾞｰﾀを使おうとすると計算量が膨大なため、どのﾃﾞｰﾀを使うべきか選択しなければならない。加えて、ﾃﾞｰﾀのどの特徴量を

使うのか選択しなければならない。これは、どの特徴量が近いものを予測対象と近いと仮定するか、どの特徴量で決定木を分岐させるかといった選択である。ﾃﾞｰﾀや特徴量をどうやって選択するかは、過去の事例を参考にする。現在、解こうとしている問題と似ている問題を過去の事例から探し、そのときに有効だったﾃﾞｰﾀや特徴量を優先的に使う。どれが似た問題化かは、問題が持つ特徴量で評価する。その特徴量の選択も、先ほどの特徴量の選択と同様に行う。厳しく見れば全ての問題は別問題であるが、甘く見れば全て問題は同種の問題である。そのため、参考になる過去の事例が全くないということはない。

ここまでで選択しなければならないことが沢山あるが、さらに、それらのどの選択処理を次に行うべきかを選択しなければならない。どう選択を行うべきかは、初めに与えられた「目的」による。一般的には、今後受け取る報酬の総和の最大化を目的とする。過去の報酬値は演繹的に定まっているが、未来の報酬値は帰納的に推論しなければならない。一般的には「割引率」というﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀによって、遠い未来の報酬の予測値ほど、低く見積もる。これは、未来のことほど予測精度が低いだろうという考え方による。ただし、割引率によって指数的に予測精度が減衰するという根拠はない。しかし、予測精度が正しく求められていれば、割引率のようなﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀは不要である。未来を予測できる限界は、k 近傍法のところで述べたのと同様に、既知のﾃﾞｰﾀ量と、未知のﾃﾞｰﾀ量の比で考えられる。例えば、過去 7 日間のﾃﾞｰﾀから、次の 7 日間を予測するのと、過去 7 年間のﾃﾞｰﾀから、次の 7 年間を予測することの確からしさは同等である。単位が変わっただけで、どちらも、未知のﾃﾞｰﾀ数(7)/{既知のﾃﾞｰﾀ数(7)+未知のﾃﾞｰﾀ数(7)}=50%までしか予測することができない。単に、遠い未来ほど予測が難しいのではなく、予測に使用するﾃﾞｰﾀによる。この方法なら、根拠なく未来を軽視・重視することなく、時間のｽｹｰﾙが異なる複数の仕事を与えても、汎用的にこなすことが出来る。

フレーム問題の解決

洞窟から時限爆弾を運び出すﾛﾎﾞｯﾄの例を考えてみよう。あらゆる可能性を考慮しようとすると、計算が終わらずに止まってしまうのがﾌﾚｰﾑ問題だ。関係があることだけ考慮すればよいと思えるが、何が関係するのか考えるのに同様に時間が掛かってしまい解決にならない。しかし、人間だって全てを考慮しているわけではなくどこかで計算を打ち切っている。どこまで考慮するべきかというﾌﾚｰﾑをうまく設定しているというのが一般的な考え方だ。しかし、適切なﾌﾚｰﾑというのは一定ではないだろう。人間だって時間に余裕があれば、ﾌﾚｰﾑを拡げてさらに良く考えるだろう。徐々にﾌﾚｰﾑを拡げていくということは、ﾌﾚｰﾑの内側ほど優先順位が高く、優先順に考慮をしているだけで、あらかじめ考慮する・しないを決めているわけではない。とはいえ、優先順位を決めるのにも同様に時間が掛かってしまうだろう。しかし、無限に時間が掛かるのは正確に求めようとしたときだけだ。あるところで優先順位の計算を打ち切ってしまえばよい。RUSAGI のように、次にどれを計算するべきかという集合の中から、良いものを徐々に絞り込んでいき、計算が打ち切られたら、残った中から均等に選べばよい。また、計算を打ち切るには、今計算を打ち切るか否かという選択も必要になる。どう行動するかの選択、何を考えるかの選択、優先順位を付ける選択と、選択しなければいけないことが無数にある。しかし本当に選択しなければいけないことは 1 つだけだ。次に何を計算するかだけを選択すればよい。先に

出たあらゆる選択のうち、次にどれについて計算するのが最良か、選択すればよい。ﾛﾎﾞｯﾄは、爆弾処理の最良の手段を選択するのではなく、爆弾処理するため最良の計算を選択すればよい。

予測ﾌｪｰｽﾞによる学習と自動深層化

説明変数を組み合わせて作った新たな説明変数を使った怠惰学習での決定木では、どんな複雑な問題でも解けるが、組合せ爆発のため効率が悪い。効率 UP の方法を考えよう。例えば、組み合わせて作った説明変数が、似た問題で有効だった場合、有効な可能性が高いと推測できる。また、有効な説明変数を組み合せたものは、無効なものを組み合わせるより有効と推測できる。学習ﾌｪｰｽﾞがなくても、過去の予測ﾌｪｰｽﾞの結果を参照すれば、効率よく推測ができる。予測ﾌｪｰｽﾞが学習ﾌｪｰｽﾞを兼ねるともいえる。新たに作成された説明変数は、従来の説明変数と同様に扱われる。例えば、画像の画素ごとの明暗という説明変数があったとして、それらを組み合わせたｺﾝﾄﾗｽﾄ、線の傾きといった説明変数が作られる。自動的にﾃﾞｨｰﾌﾟﾗｰﾆﾝｸﾞのように階層化されていく。層の数のようなﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀがないため、いくらでも複雑なことを扱えるようになる。また、必要以上に層があることによる効率悪化もない。

汎用 AI に必要な 3 要件

汎用 AI のｱﾙｺﾞﾘｽﾞﾑが、どんな問題に対しても、計算時間とともに最適解へ近づき続けるためには、次の 3 要件を満たす必要がある。

必要要件	保障内容
ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀなし	・どんな問題へも最適化できる ・余計な事前知識の混入がない
微小な差は、 微小な差しか生まない	・精度の限界が無い ・原理的な帰納推論の妥当さ
情報量増加で 予測精度は単調増加	・計算と共に精度が上がり続ける ・一部の認知バイアスの回避

一つ目の要件は、ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀがないこと。処理する数値に限らず、処理の流れについても、固定されていてはいけない。問題によって最適な数値・流れが変わるため、あらかじめ固定されてはいけない。理想的には無限の可能性（例えば実数全体）の中から、ｱﾙｺﾞﾘｽﾞﾑが数値を選べられるべきである。あらゆる問題に対応できる汎用性のためには、あらゆる状態を取れなければならない。ただし、知能のｱﾙｺﾞﾘｽﾞﾑと無関係な部分（例えばｵﾍﾟﾚｰﾃｨﾝｸﾞｼｽﾃﾑ）にﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀが合っても良い。

二つ目の要件は、ある数値の変化が小さくなるほど、他の数値の変化も小さくなること。y=f(x)の、x の値の差が近いほど、y の値の差も近くなると推測すること。つまり関数に不連続な段差があってはいけない。段差があれば無限小のﾉｲｽﾞで結果が大きく変わってしまうので、ﾉｲｽﾞの影響を減らし続けることができない。ある特徴量が最大の要素を選ぶといった処理も、

段差同様に無限小のﾉｲｽﾞの影響を受けるのでならない。また、この要件は原理的にどういった予測がより良いかという基準が、あらゆる場合について同じであることを意味する。実際には不連続な関数も存在するが、ﾉｲｽﾞがある状態では連続な関数として予測できる。例えば tan(約0)は正か負の頻度によって予測値は連続的な値をとると予測するのが妥当である。

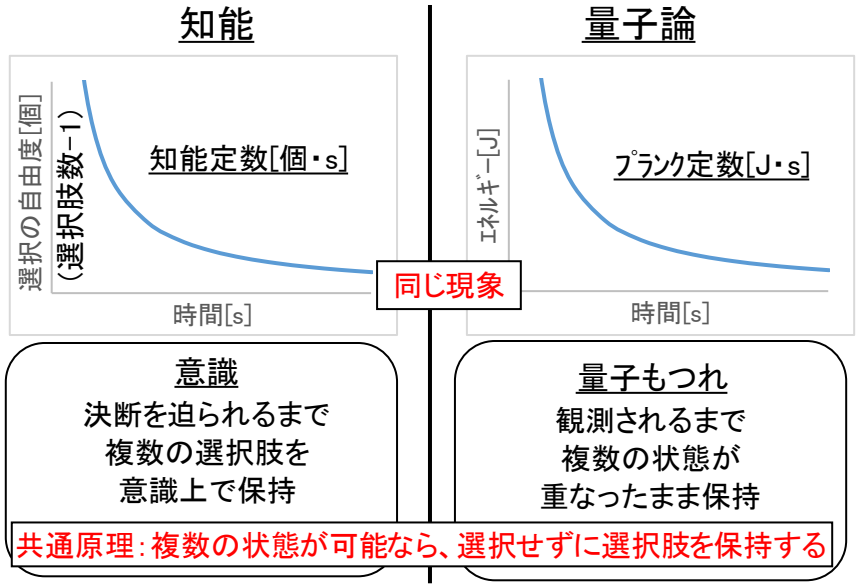
三つ目の要件は、どんな情報が与えられても予測精度は同じか、高くならなければならない。また、ﾌﾟﾛｸﾞﾗﾑは目的の予想精度が高くなるように進行しなければならない。これが満たされれば時間と共により精度が高い結果に向かい続けることができる。これが満たされない場合、例えばこれまでの予測結果に否定的な新しい情報が入ってきたときに、その情報を見て見ぬフリをした方が予測精度が良くなってしまう。情報の取捨選択する権限をﾌﾟﾛｸﾞﾗﾑへ与えるということは、人間同様に認知ﾊﾞｲｱｽの可能性も与えることになる。最適解を目指すなら、どんな情報も捨てずに保持し続けなければならないが、記憶容量の制約は回避できない。

これらの要件を満たせば汎用 AI となるが、計算の効率の良さまでは保証しない。ｱﾙｺﾞﾘｽﾞﾑに冗長性があれば、計算時間は余分に掛かる。しかし、計算速度は、計算機を並列にすれば上げられるため、重要ではない。汎用 AI の開発においては、この要件を満たしているか確認することによってのみ、真には汎用性があるか確認できる。特定の試験成績が良いかどうかでは、汎用性の評価はできない。試験結果が良かったとしても、それは偶然か、その試験に特化させる調整をしただけであり、汎用 AI の評価に全く意味がなく、ﾊﾟﾌｫｰﾏﾝｽでしかない。

脳と意識

脳と AI を比較する。脳は記憶装置と演算装置を兼ねている。また意識と呼ばれるﾜｰｷﾝｸﾞﾒﾓﾘｰ内で意思決定する仕組みがある。例えば、食物の映像が目に入り、捕食しようとする場合を考える。意識に映像が上がると、ほぼ同時にそれが食物という認識結果と、捕食すれば空腹が満たされるとい予想結果が意識に上がる。ここまでに自由意志はなく、認識したくないと思っても抗うことはできない。その先の、捕食するべきか思索したり、実際に捕食行動を始めるタイミングの決断は、自由意志で決定される。蓄積された学習の記憶を元に、ﾆｭｰﾛﾝの発火の連鎖により関連することを瞬時に想起する仕組みは、ﾃﾞｨｰﾌﾟﾗｰﾆﾝｸﾞに相当する機能である。しかし、それだけでは、意識上に情報がﾛｰﾄﾞされるまででその先がない。今、捕食するべきと十分な精度の予測結果が得られるまで、思索を続ける仕組みが必要である。その仕組みが、RUSAGI のような汎用 AI の仕組みである。意識上では学習済みのことに限らず、自由に物事を考えることができる。意識は怠惰学習である。そのため意識上での思考は直感より遅いが、汎用性がある。

知能の単位と量子論



知能の水準の数値化を考える。あらゆる問題は、解を候補の中から選択するものと解釈できる。時間と共に有効な選択肢を絞り込んでいく。「選択肢数の減少量／時間」が大きいほど優れているが、この値は時刻と共に変化する。時刻 0 では、選択肢数は∞である。また、時刻と共に選択肢数は 1 へ漸近的に近づく。これらの条件を満たすには「(選択肢数-1)×時刻」という指標が考えられる。この値が小さいほど、選択肢数が速く収束するため優れる。また、選択の自由度=選択肢数-1 であるので、「知能定数[個・s]＝選択の自由度×時間」となる。

この知能定数の単位[個・s]は、プランク定数と単位[Ｊ．s]と似ている。物理世界が、ｴﾈﾙｷﾞｰがいくつであるかを選択していると解釈すると、2 つの単位は一致する。知能と不確定性原理を同じ仕組みで考えてみよう。計算機の能力が無限大ではないため、AI は瞬時に予測値を絞り込み切れないのと同じで、この世界もまた素粒子の取りうる状態を瞬時には絞り込めないのである。2 つが同じ仕組みと考えると、知能にも量子もつれのような現象があるはずである。ｼｭﾚｰﾃﾞｨﾝｶﾞｰの猫は、生きているという選択肢と、死んでいるという選択肢の両方が許されるが、そのどちらかの状態であるのではなく、観測するまではどちらでもある重なった状態である。観測によって 2 つの選択肢が 1 つに絞り込まれる。しかし、観測しない限り、いくら時間経過してもどちらかに決まらない。どちらであるべきかという情報がないため、いくら時間があっても絞り込めないのである。これは意識上で考え続けているのと同じである。判断に必要な情報が欠けていれば、無限に計算し続けても選択肢は 1 つに絞られない。ただし、猫は生きていても死んでいても、ただ確率的にそうだといだけなので、観測する前にどちらかに決まってしまっても物理法則に反しない。しかし、現実には決まらない。この世界には、「複数の選択肢をとり得るなら、どれか選択せず、複数の選択肢保持する」という原理が存在すると解釈できる。この原理があるから、ﾃﾞｨｰﾌﾟﾗｰﾆﾝｸﾞのように直感だけで判断せず、意識の中で選択肢を保持する仕組みが存在すると考えられる。

RUSAGI ロードマップ

完成予定	汎用性	ｱﾙｺﾞﾘｽﾞﾑ(製品)
2019 年完成	ﾚﾍﾞﾙ 0	(なし)
2021-23 年	ﾚﾍﾞﾙ 1	Lazy RUSAGI
2025-27 年	ﾚﾍﾞﾙ 2	Deep RUSAGI
2029-31 年	ﾚﾍﾞﾙ 3	Super RUSAGI

2019 年完成 (汎用 Lv0)

ここで述べられている方法で、ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀを消去した、k 近傍法や決定木。予測精度は改善するが、汎用性はない

2021-23 予定 (汎用 Lv1) Lazy RUSAGI

時間さえ掛ければどんな問題でも最適解へたどり着く。ﾊｲﾊﾟｰﾊﾟﾗﾒｰﾀのない k 近傍法の距離を決定木で決定。説明変数のあらゆる組み合わせ試すため、組み合わせ爆発が起こり、複雑な問題は現実的な時間で結果が収束しない。画像認識のような説明変数が大量にあるものは対応困難。これまで、回帰分析、ﾗﾝﾀﾞﾑﾌｫﾚｽﾄ、SVM、層の少ないﾆｭｰﾗﾙﾈｯﾄﾜｰｸ等である程度予測できていたことならば、計算時間とﾄﾚｰﾄﾞｵﾌでどこまでも精度を上げられる。計算時間または目標精度を指定する必要あり。純粋な怠惰学習であり、学習ﾌｪｰｽﾞは存在しない。株価や為替予測のように、限られた情報の中で、時間多少かかってもよいので、できるだけ精度よく予測したい用途に最適。

2025-27 予定 (汎用 Lv2) Deep RUSAGI

Lazy RUSAGI に自動深層化と、過去の予測結果の参照を可能にしたもの。同様の問題なら、予測するたびに学習して高速化する。ﾃﾞｨｰﾌﾟﾗｰﾆﾝｸﾞが得意とする画像認識のような複雑な予測が可能。現在、ﾃﾞｨｰﾌﾟﾗｰﾆﾝｸﾞが活躍している用途で、計算時間と引き換えに、精度をどこまで上げ続けることができる。指定した時間か精度に達するまで処理を継続し続ける。単独の用途

であれば汎用であり、高効率。

2029-31 予定(汎用 Lv3) Super RUSAGI

Deep RUSAGI に、複数の問題の計算時間を自動的に最適割り当てする機能を付与したもの。与えられた問題を解くのではなく、今、何を計算すべきかを解く。強化学習ｴｰｼﾞｪﾝﾄのように、入力、出力、報酬だけを設定して使用する。どうすれば報酬が大きくなるのかを目的として、今、予測するべきことを予測する。学習が進むともに、どこまでも複雑なことを考えられるようになる。人間と同等のことが可能。計算速度・効率を脳と比較するのは難しいが、こちらの方が優れた汎用性を持つだろう。人間と違って、時間を掛けるほど、どこまでも難しいことを、どこまでも正確に予測できるだろう。人間を代替するという用途だけでなく、計算ﾘｿｰｽさえ割ければ、人間に成しえなかったことさえ成せるだろう。完全自律型の超 AI である。