

アルゴリズムの千カラ

～大規模データ処理からAIまで～

九州大学 システム情報科学研究所
准教授 稲永 俊介



自己紹介

自己紹介

氏名：稲永 俊介（いねなが しゅんすけ）

出身：福岡県

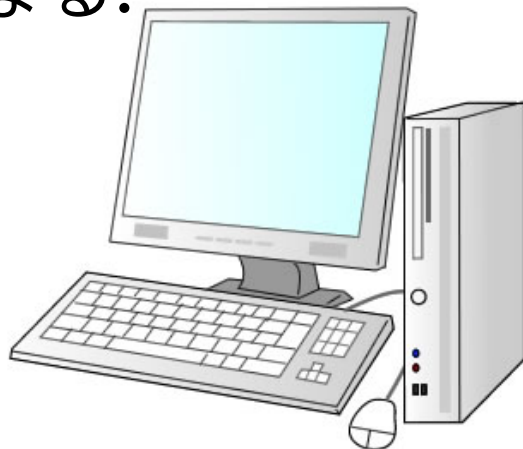
現職：九州大学 システム情報科学研究所 准教授

専門：情報科学

アルゴリズムとは？

アルゴリズムって何？

- ◆ コンピュータ上で行われていることはすべて「計算」である。
- ◆ コンピュータ上でいかなる「計算」を行うときも、その計算方法の基礎である**アルゴリズム**の設計が重要となる。



アルゴリズムって？

- ◆ コンピュータは、四則演算（ $+$ ， $-$ ， \times ， \div ）などの単純な計算を高速かつ正確に行える
- ◆ しかし、複雑な問題を解くための「手順」をコンピュータ自身が編み出すことはできない
- ◆ 単純な計算を組み合わせて、より複雑な問題を解くための手順のことをアルゴリズムという

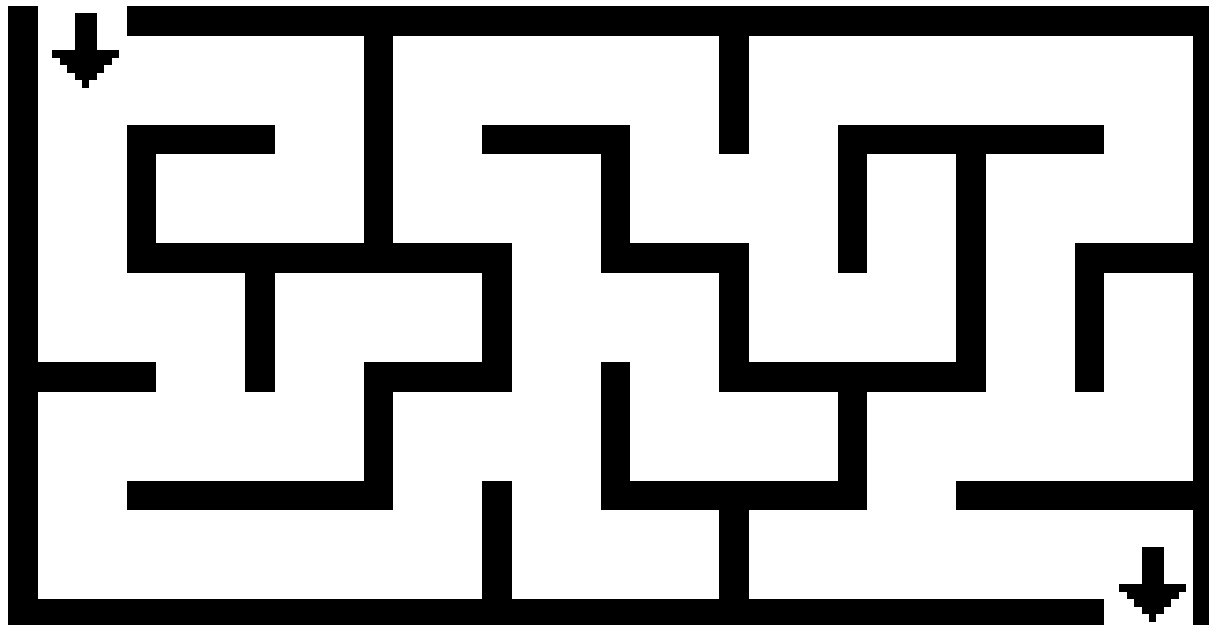
アルゴリズムの例（その1）

問題：迷路を脱出せよ！

コンピュータは前後左右に動くことはできるが、
いつどの方向に行けばいいか知らない



スタート



ゴール

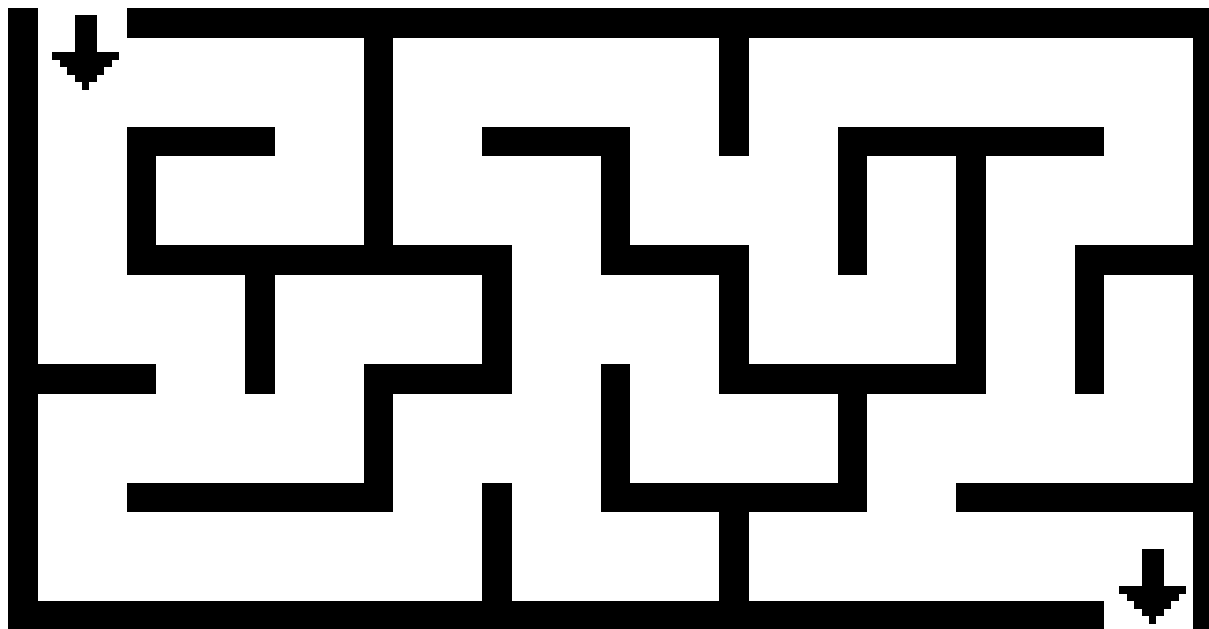
アルゴリズムの例（その1）

問題：迷路を脱出せよ！



行き止まりになったら
どうしたらいいんですかー？

スタート



ゴール

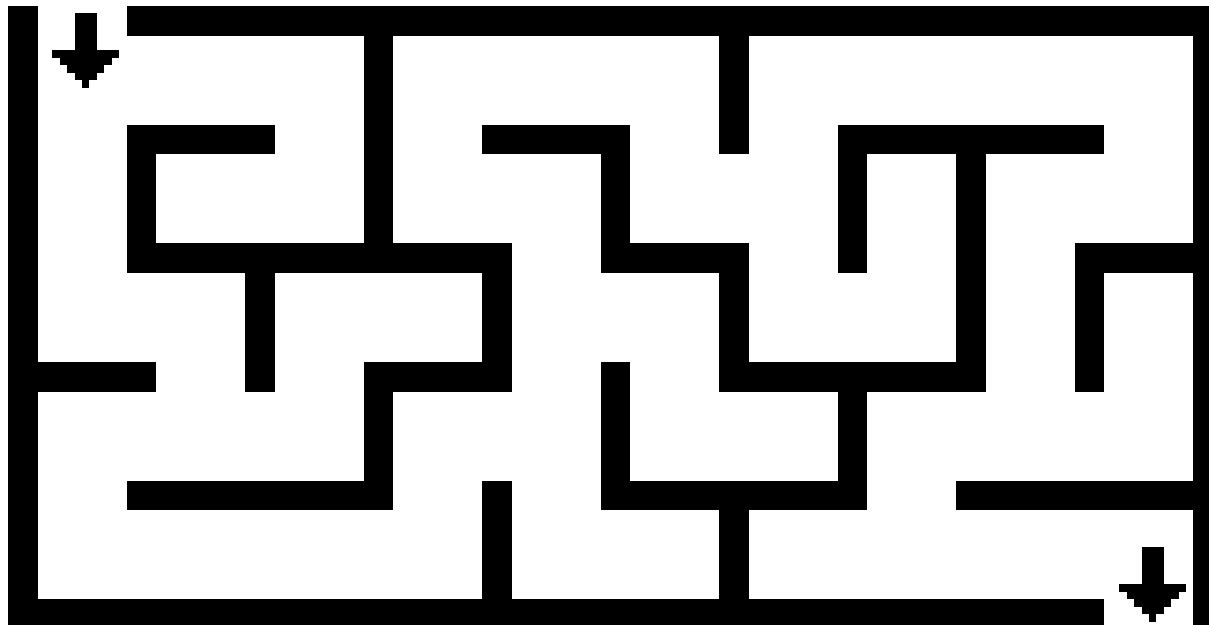
アルゴリズムの例（その1）

問題：迷路を脱出せよ！

壁に右手をついた
まま進みなさい



スタート



ゴール

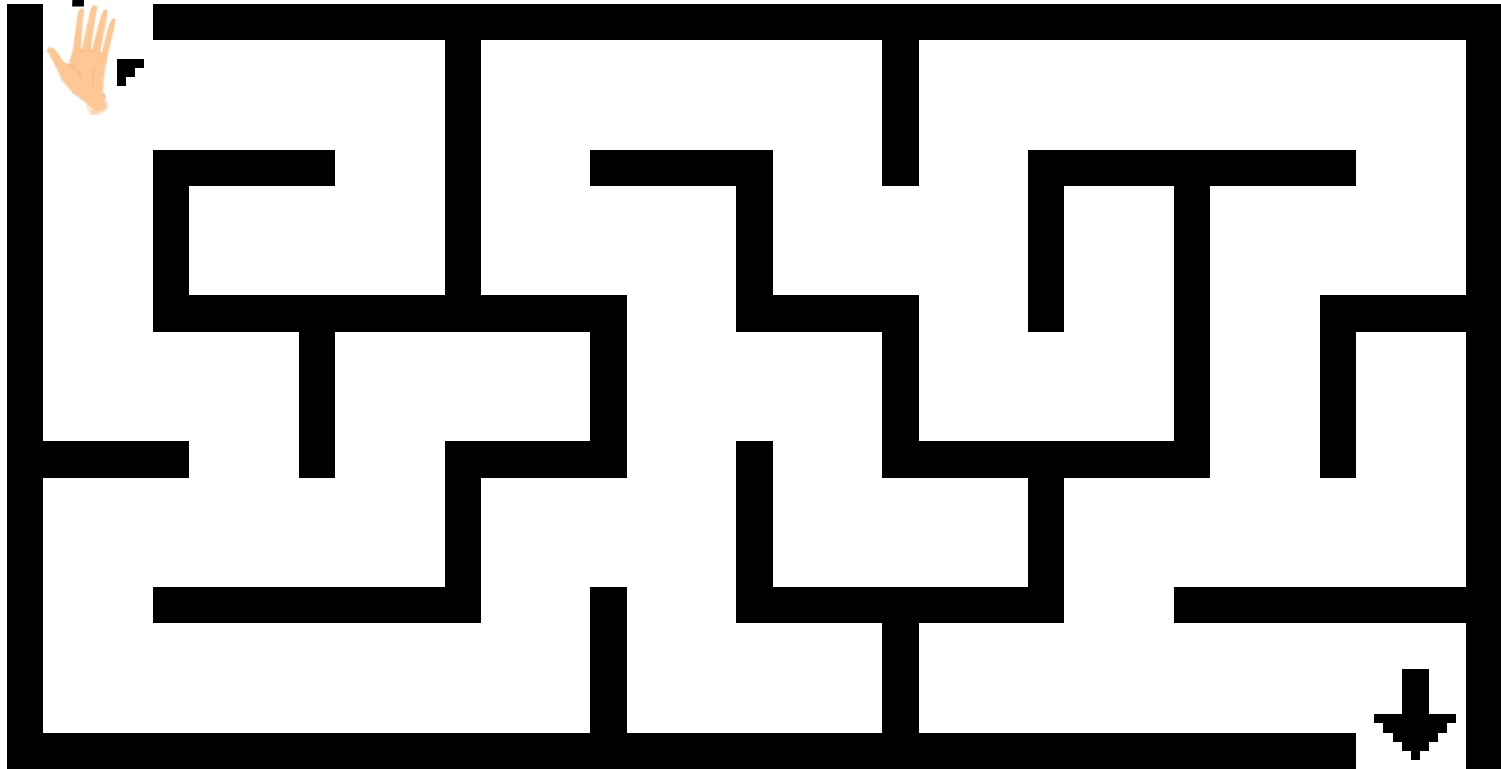
アルゴリズムの例（その1）

問題：迷路を脱出せよ！

いきまーす！



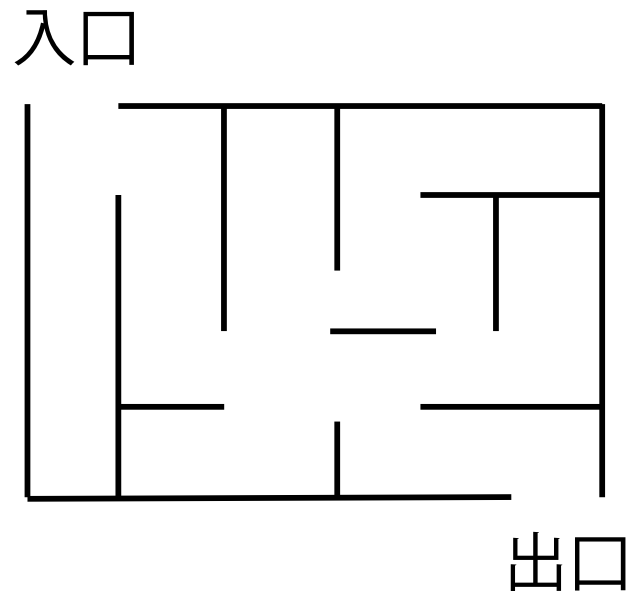
スタート



ゴール

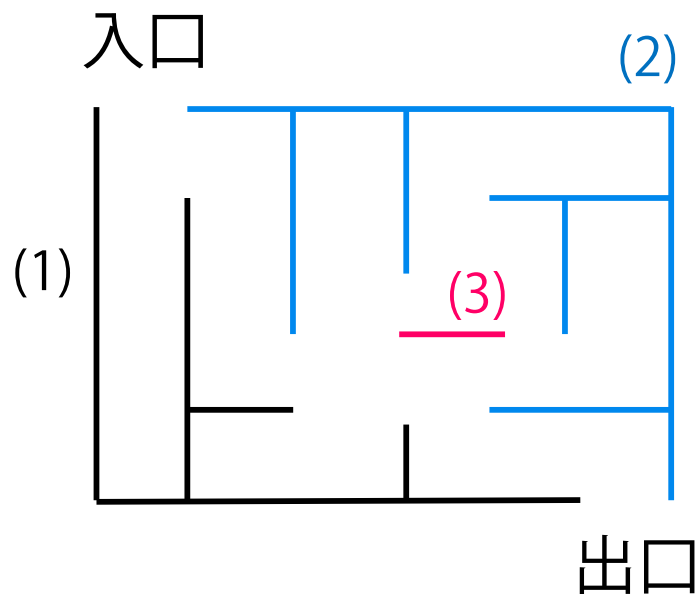
なぜ迷路を抜けられるのか？

- 迷路を以下の3つのパーツに分割して考える.



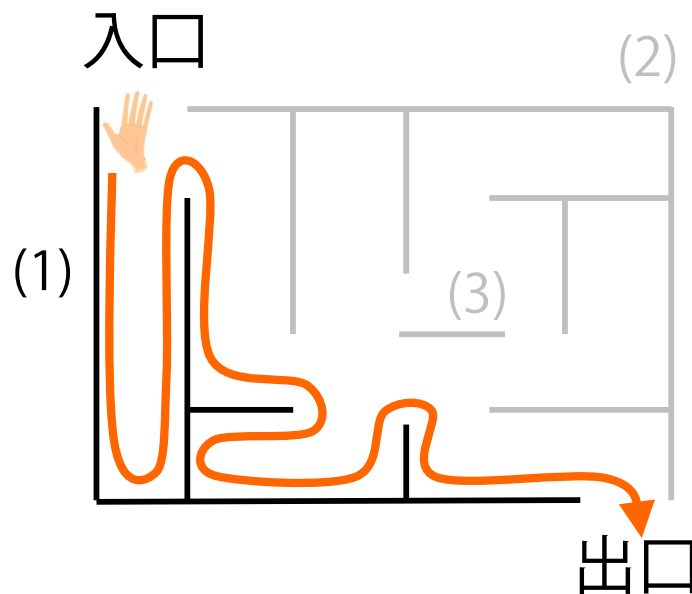
なぜ迷路を抜けられるのか？

- 迷路を以下の3つのパーツに分割して考える。
 - (1) 入口から見て右側の壁に繋がっている部分.
 - (2) 入口から見て左側の壁に繋がっている部分.
 - (3) その他の部分.



なぜ迷路を抜けられるのか？

- (1) のみに着目し, (2)(3) は無視する.
- (1) の壁沿いに歩けば, 出口に通じるのは明らか.



右手法

アルゴリズムの研究

- ◆ 問題を解く手順を与えれば、
コンピュータは人間よりもずっと高速かつ
正確に問題を解くことができる
- ◆ ただし、問題を解く手順（アルゴリズム）を
考案するのは、人間の役割である
- ◆ 備考：アルゴリズムをコンピュータが
理解できるように書いたものがプログラム

アルゴリズムの「良さ」

- ◆ 同じ問題でも、それを解くためのアルゴリズムは何種類もありえる
 - 解き方は一通りとは限らない
- ◆ 「良い」アルゴリズムの条件
 1. 正しく問題を解ける
 2. 高速に問題を解ける

アルゴリズムの例（その2）

問題：9つのコインがあり、その中に一つだけ偽物のコインがある。偽物は本物より重い。天秤を使って、偽物を発見せよ。



天秤を何回使えば偽物を発見できるだろうか？

アルゴリズムの例（その2）

適当に選んだ1個と残りの8個を順番に比べれば、最悪の場合でも8回天秤を使えば偽物が見つかる

より少ない回数でできるだろうか…?



1個ずつ天秤に乗せていく

アルゴリズムの例（その2）

実は、たったの **2回** で
発見できる！



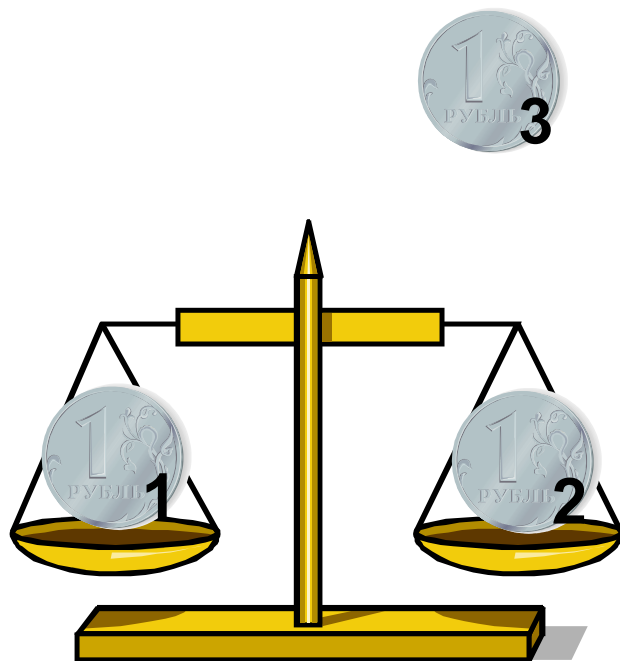
考え方

コインが3個の場合を考える



考え方

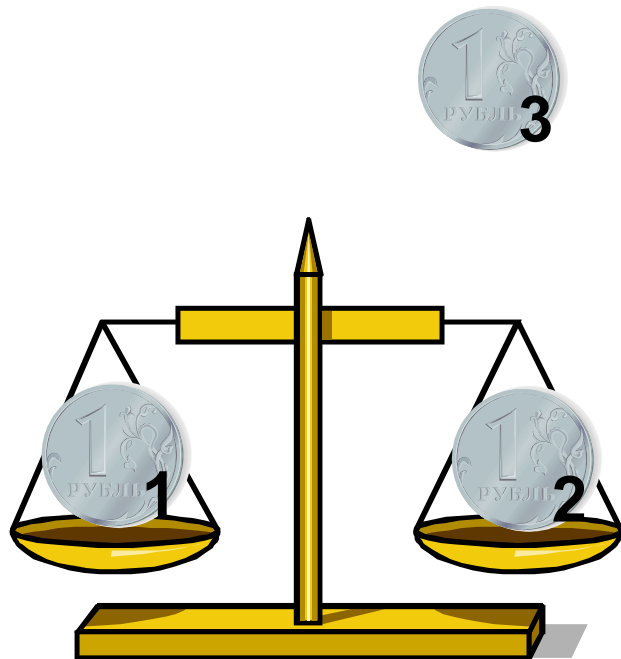
コインを2個選び、天秤にかける



- a. 1のほうに傾けば、
1が偽物
- b. 2のほうに傾けば、
2が偽物
- c. 釣り合えば、
3が偽物

考え方

コインが3個のときは、
天秤を **1回** 使えば偽物を発見できる



- a. 1のほうに傾けば、
1が偽物
- b. 2のほうに傾けば、
2が偽物
- c. 釣り合えば、
3が偽物

考え方

コインが9個のときは、
コインが3個のときのやり方を応用する



考え方

コインを3つのグループに分ける

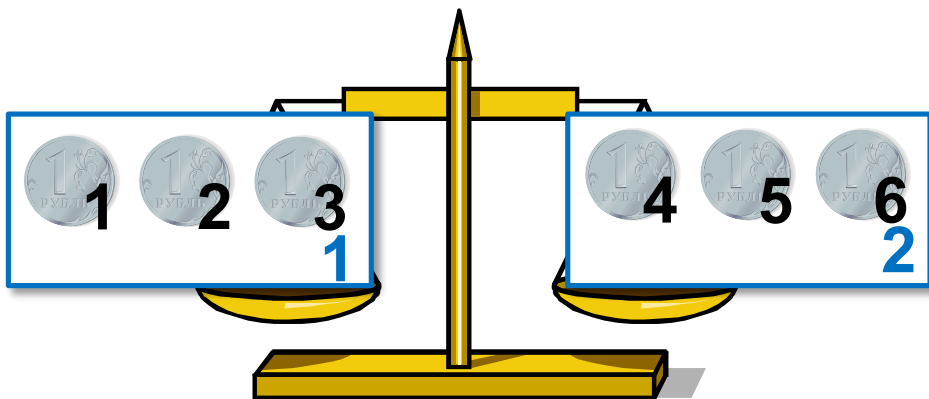


考え方

グループを2つ選び、天秤にかける



- 1のほうに傾いたら、
1の中に偽物がある
- 2のほうに傾いたら、
2の中に偽物がある
- 1と2が釣り合えば、
3の中に偽物がある



考え方

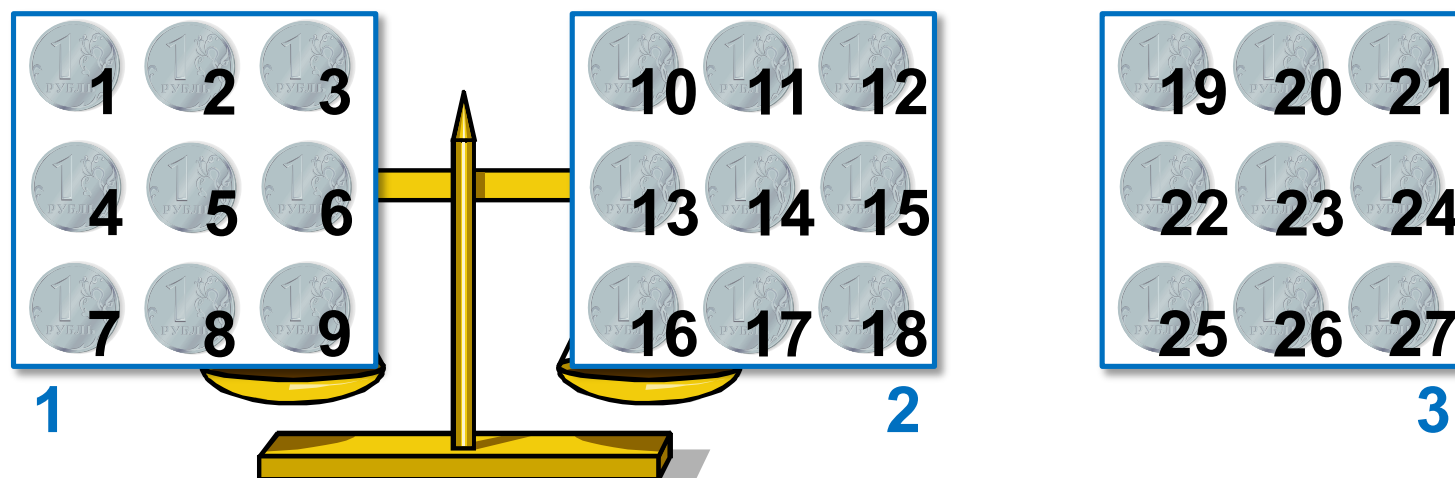
偽物が入ったグループを見つけるのに1回
そのグループから偽物を見つけるのに1回

よって、コインが9個のときの
天秤の使用回数は **2回**



コインが 27 個のとき

3 グループに分け、2 つを天秤にかける



天秤を 1 回使って、偽物が入ったグループを発見
あとは、コインが 9 個の場合と同様の手順
⇒ 天秤を **3 回** 使用すれば偽物を発見できる

天秤の使用回数 ($n = 3^k$ のとき)

コインの個数	天秤の使用回数
3	1
9	2
27	3
81	4
⋮	⋮
n	$\log_3 n$

天秤を1回使うと偽物の候補の数が3分の1になる。

$\log_3 n = n$ を3で繰り返し割って1になる回数



天秤の使用回数（一般の場合）

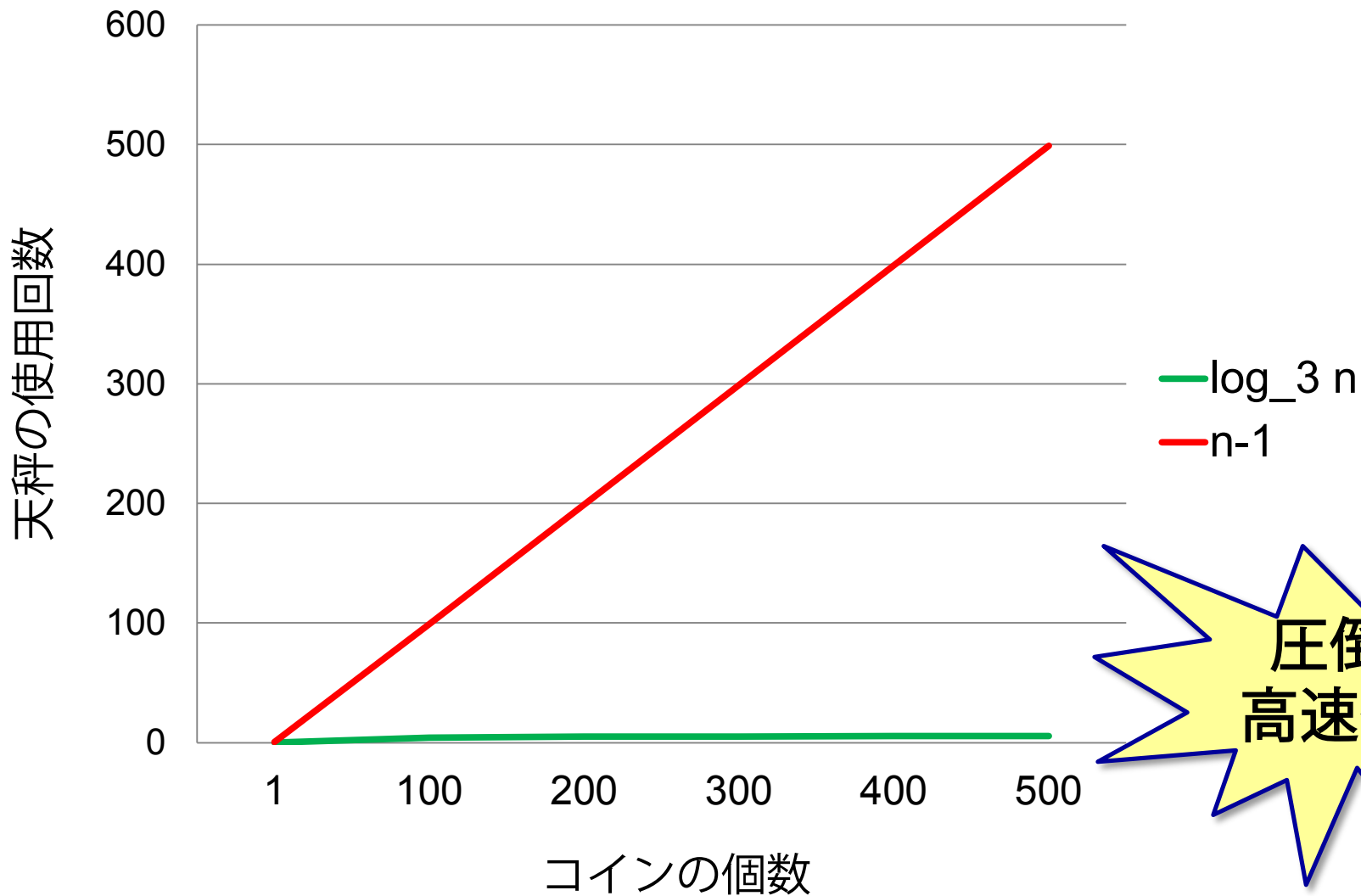
コインの個数	天秤の使用回数
2 ~ 3	1
4 ~ 9	2
10 ~ 27	3
28 ~ 81	4
⋮	⋮
n	$\log_3 n$

天秤を1回使うと偽物の候補の数が3分の1以下になる。

$\log_3 n = n$ を3で繰り返し割って1以下になる回数



$\log_3 n$ VS $n - 1$



実社会のアルゴリズム

実社会で活躍するアルゴリズム

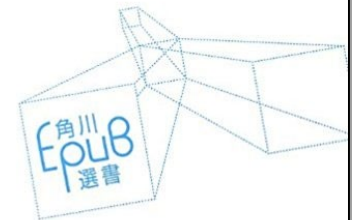
- ウェブ検索 (Google)
- SNS (Facebookの友達候補・広告, twitterのボット)
- 商品推薦 (Amazon など)
- 航空機の自動操縦
- 車の自動運転 (Tesla, ウェイモ)
- 株式市場 (アルゴリズム取引)
- コンピュータ将棋・囲碁 (ボナンザ, アルファ碁)
- クイズ (ワトソン)

商品推薦

➤ Amazon の商品推薦

アルゴリズムに
基づいて商品を推薦

アルゴリズムが
世界を支配する
クリストファー・スタイナー
Christopher Steiner
訳/永峯涼



シリコンバレー発 アルゴリズム革命の衝撃
Fintech, IoT, Cloud Computing, AI...
アメリカで起きていること、これから日本で起ること
ここまで変わる
人間と機械の近未来!
シリコンバレー発 アルゴリズム革命の衝撃
Fintech, IoT, Cloud Computing...
榎田健児
★★★★☆ 13
単行本
¥1,728 ✓prime



AIの衝撃 人工知能は人類の敵か
小林雅一
講談社現代新書
2007
AIの衝撃 人工知能は人類の敵か (講談社現代新書)
小林 雅一
★★★★☆ 52
新書
¥864 ✓prime



アルゴリズム図鑑 絵で見
てわかる26のアルゴリズム
石田 保輝
★★★★☆ 20
単行本 (ソフトカバー)
¥2,570 ✓prime

アマゾンが開発した
商品推薦アルゴリズムの
解説論文



Amazon.com Recommendations *Item-to-Item Collaborative Filtering*

Greg Linden, Brent Smith, and Jeremy York • Amazon.com

Recommendation algorithms are best known for their use on e-commerce Web sites,¹ where they use input about a customer's interests to generate a list of recommended items. Many applications use only the items that customers purchase and explicitly rate to represent their interests, but they can also use other attributes, including items viewed, demographic data, subject interests, and favorite artists.

There are three common approaches to solving the recommendation problem: traditional collaborative filtering, cluster models, and search-based methods. Here, we compare these methods with our algorithm, which we call *item-to-item collaborative filtering*. Unlike traditional collaborative filtering, our algorithm's online computation scales independently of the number of customers and number of items in the product catalog. Our algo-

クイズ

- 2011年, IBM が開発した質問応答システム「ワトソン」がアメリカの早押しクイズ番組 Jeopardy! で人間のクイズチャンピオンと対戦して勝利.
- オンライン百科事典「ウィキペディア」から学習した知識を利用して, クイズに回答する.
- 一方で, ワトソンが回答を出すプロセスは, 人間の思考方法とはかなり異なる.



WIKIPEDIA
The Free Encyclopedia

ワトソンの応用

- 医療アドバイス（東大医科研）
 - がんに関連する約2千万件の論文をワトソンが学習（機械学習）。
 - 60代の女性患者の白血病が治療が難しい特殊なタイプであるとの分析を10分で行い，治療法の変更を提案。
- 人材マッチング（フォーラムエンジニアリング）
 - ワトソンが個人の興味，関心，性格などを数値化して把握し，マッチングスコアを算出する。
2016年4月から実際に運用されている。
- 他にもいろいろ

コンピュータ将棋

将棋とは



- 9×9のマスの中で、動き方8種類、40枚の駒を動かし、玉を取れば勝ち
- 相手の駒を取り、自分の持ち駒として使うことができる
- 将棋人口は約1300万人 (レジャー白書調べ)
- ニコニコ生放送カテゴリ別視聴数ランキングTOP3 (羽生善治 & ドワンゴ社長対談)

コンピュータ将棋の歴史

1974年 コンピュータ将棋開発開始？

...

2005年 「激指」がアマ竜王戦ベスト16

2006年 「Bonanza」登場

(評価関数に**機械学習**を導入)

2010年 「あから2010」が清水女流王位(当時)に勝利

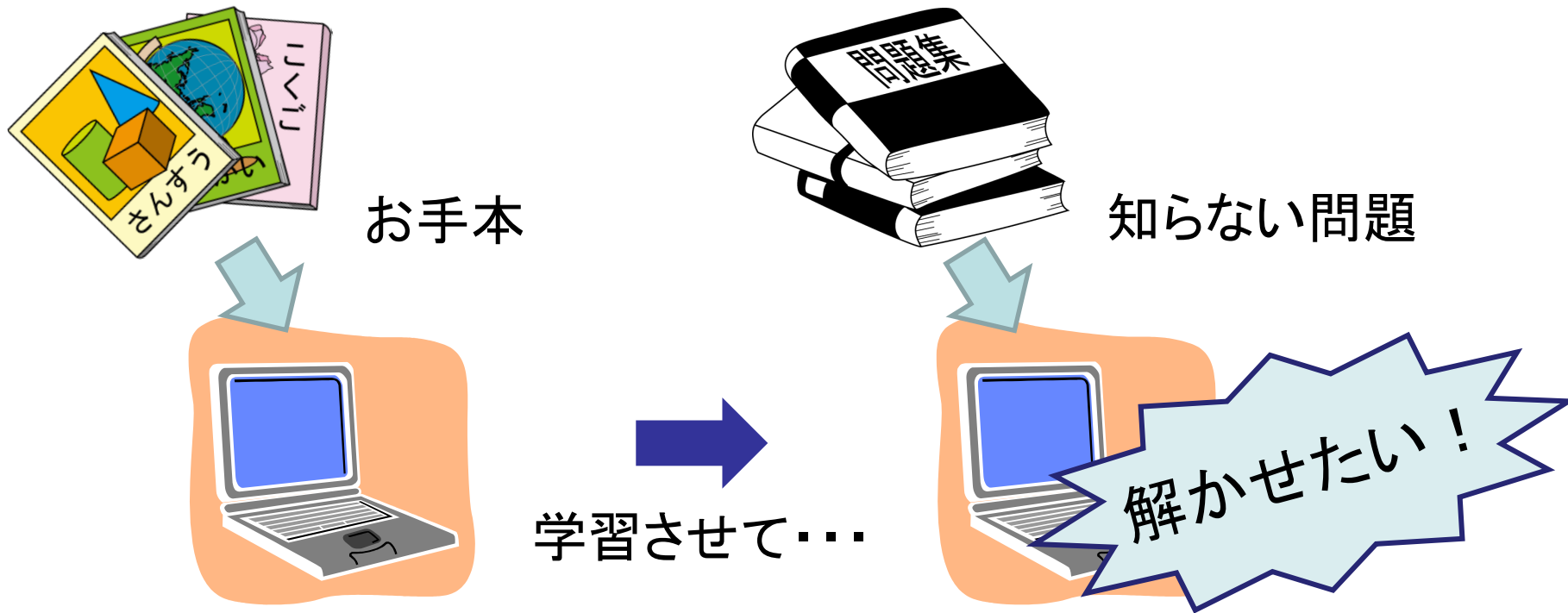
2012年 「ボンクラーズ」が米長永世名人(引退棋士)に勝利

2012年 電王戦にてコンピュータが現役プロ棋士チームに勝ち越し

以後コンピュータ将棋の優勢が続く
(トップ棋士も敗れる)

機械学習

- 機械学習：人間が自然に行っている学習能力と同様の機能をコンピュータで実現しようとする技術 (Wikipedia)



Bonanza 法 [保木, 2006]

- コンピュータ将棋に、機械学習技術を取り入れた最初的手法.
- 現在も評価関数作成の基礎となっている、ブレイクスルー的な手法.
- プロ棋士の棋譜を教師データとして、プロ棋士の手とコンピュータの手が一致するように評価関数を自動調整する手法.

評価関数（評価基準）

- 評価項目：どんな特徴を盤面の点数に反映させるべきか？
- 値の調整：各特徴に、どのくらいの点数を設定すればよいか？

例：駒に点数をつけて評価

飛	角	金	銀	桂	香	歩	龍	馬	成銀	成桂	成香	と
170	140	100	80	60	50	20	185	160	91	82	78	65

コンピュータ将棋

コンピュータ将棋の代表的な評価項目

- 駒の価値
- 玉に対する金の位置
- 玉の危険度(玉周辺の敵駒の利き)
- 各駒の働き(利き, 敵, 味方の駒との位置関係)
 - 桂馬が跳べるか
 - 歩が進めるか
 - ひもがついているか

組み合わせを考慮すると**数億**に及ぶ

コンピュータ将棋

コンピュータ将棋の代表的な評価項目

- 駒の価値
- 玉に対する金の位置
- 玉の危険度(玉周辺の敵駒の利き)
- 各駒の働き(利き, 敵, 味方の駒との位置関係)
 - 桂馬が跳べるか
 - 歩が進めるか
 - ひもがついているか

全部使うわけにはいかない
➤ 計算時間がかかりすぎる
➤ 不要な項目もたくさんある

組み合わせを考慮すると**数億**に及ぶ

評価項目の選定の難しさ

- これらの機械学習アルゴリズムの性能、ひいては人工知能 (AI) の性能は、対象をどんな項目で評価するかに大きく依存する。

職人技

- 既存の機械学習の手法では，評価項目の選定と値の設定は人間が手作業で行っていた。
- 職人技によるチューニングの世界。



飛	角	金	銀	桂	香	歩	龍	馬	成銀	成桂	成香	と
170	140	100	80	60	50	20	185	160	91	82	78	65

職人技

- 既存の機械学習の手法では，評価項目の選定と値の設定は人間が手作業で行っていた。
- 職人技によるチューニングの世界。



この辛い作業を自動化したい...

飛	角	金	銀	桂	香	歩	龍	馬	成銀	成桂	成香	と
170	140	90	80	60	50	25	185	160	91	82	78	65

深層学習 (Deep Learning)

- 蓄積された大量のデータから，コンピュータが自動的に評価項目を作り出す手法。
 - 人間による評価項目の選定，数値の設定（職人技）を必要としない。
- 現在の人工知能ブームは，深層学習の発達・普及によるところが大きい。
- 「人工知能研究における50年来のブレイクスルー」

大規模示一々処理

ビッグデータ

毎日のように大量のデジタルデータが生成され、蓄積され続けている。

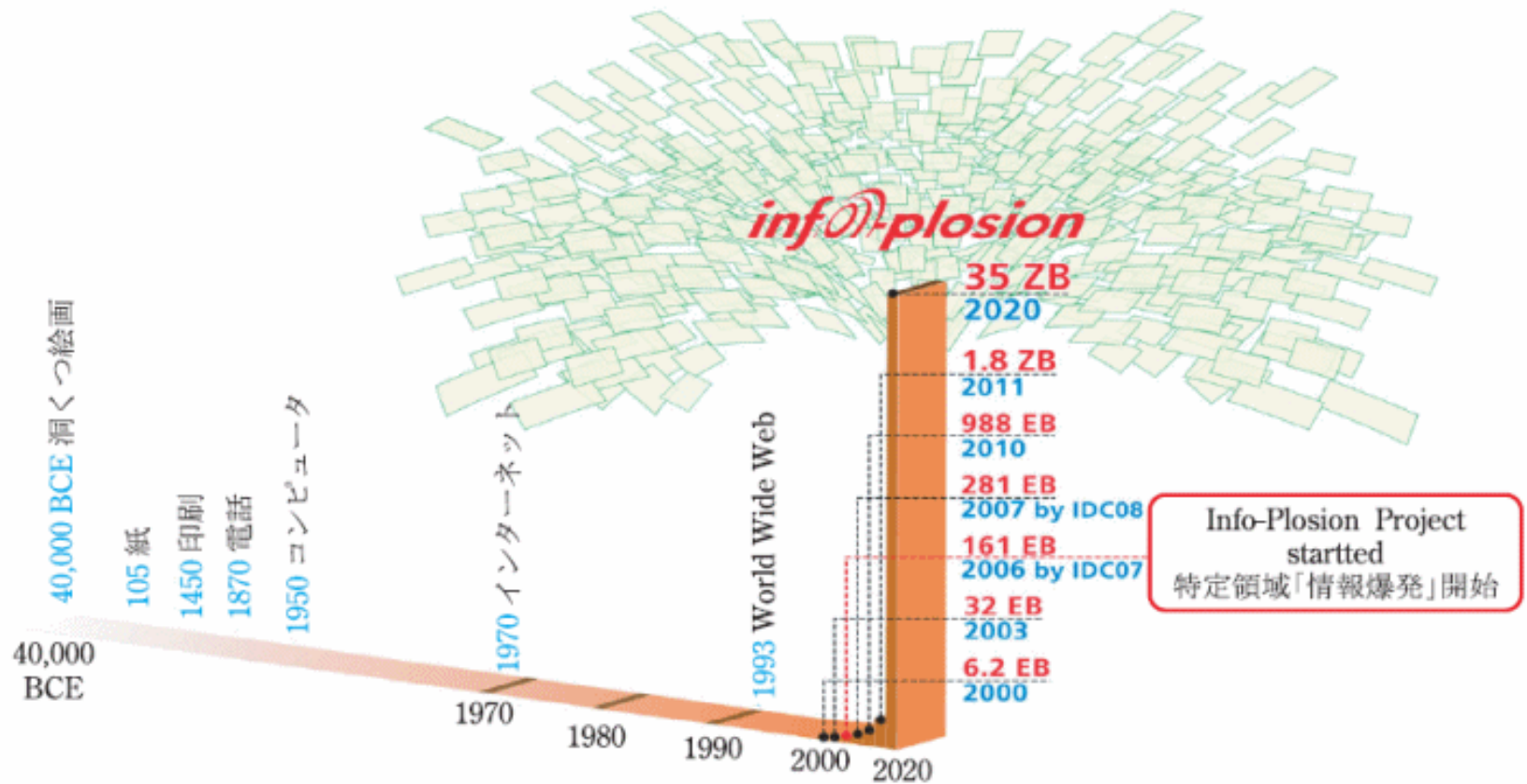
- Google: 60 億件の検索
- Facebook: 43 億メッセージ
- Twitter: 5 億ツイート
- YouTube: 400 万時間分の動画



データは <https://www.gwava.com/blog/internet-data-created-daily> より

情報爆発

- 電子データの量が「爆発的に」増加している！

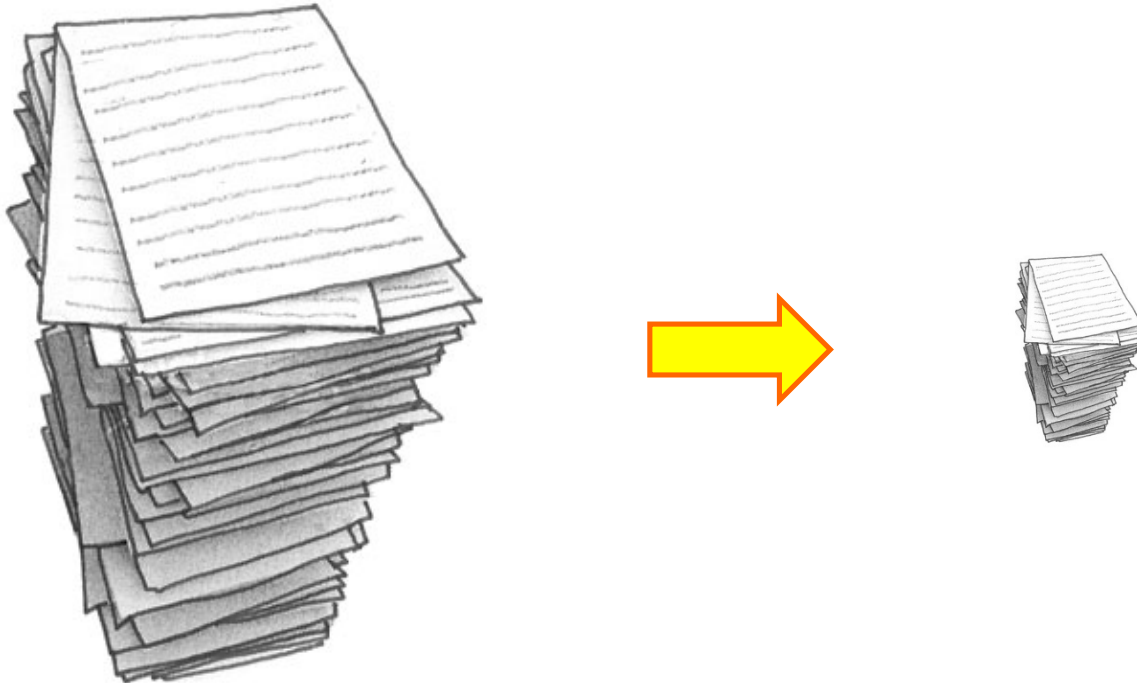


喜連川優「情報爆発のこれまでとこれから」,電子情報通信学会誌, Vol.94, No8, 2011

私の研究

【圧縮文字列データ処理】

- デジタルデータを**文字列**（記号の連鎖）として捉え、
- データを**圧縮**したまま処理する技術。



データ圧縮

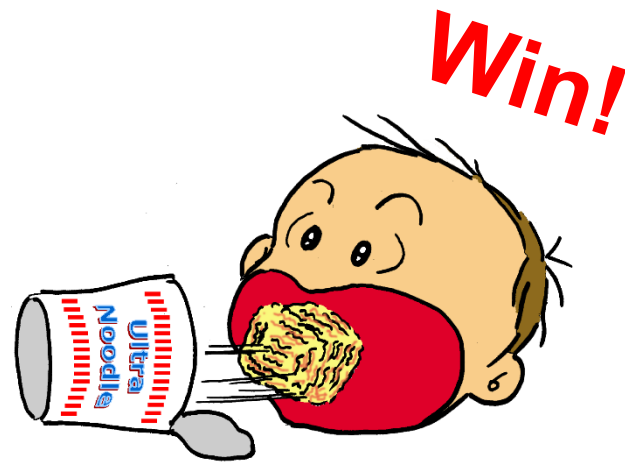
文字列データを圧縮して表現すれば、データの「無駄」を削減できる。

例 1) すもももももももものうち
=すも⁸のうち

例 2) にわにわにわにわとりがいる
=(にわ)⁴とりがいる

研究成果

- 圧縮された文字列データ上で情報検索を高速に行うアルゴリズムを開発した
 - 大量のデータも圧縮すれば小さくなる
 - 圧縮することで、情報検索の速度も向上！



九州大学の情報系学科

九州大学 工学部 電気情報工学科



まとめ

- アルゴリズムとは、計算の手順のことである。
- アルゴリズムの工夫次第で、
計算を高速化できる。
- 実社会の様々な場面でアルゴリズムが活躍中。
- アルゴリズム・プロコン等に興味がある人の
大学編入・大学院進学を歓迎します！