

研究紹介スライド

原口誠

情報・類似性・
アナロジー

全ての情報は類似性・アナロジーにより組織化され知識として利用できる形態に進化する

居室： 情報科学研究科棟6F東

講義(電子の専門科目)：

「プログラミング言語」(必)、

「知識ベース工学」(選)

大学院講義 知識発見特論、知識ベース特論

行うべきこと

タスクに依存して「重要な情報」を得て、
それを知識化(言語化)すること

1. データマイニング: データから重要な情報を得ること。
頻度情報に基づく統計処理 + 組合せ論的アルゴリズムの手法

本研究室: これらの成果も最大限に活用しながら、
データ - 情報 - 知識 の関係を明らかにする。つまり

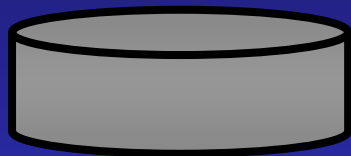
2. 知識獲得: 情報(データ記述言語)からルール等の
形式で記述された知識を得ること。

キーは「重要な情報」：タスク依存



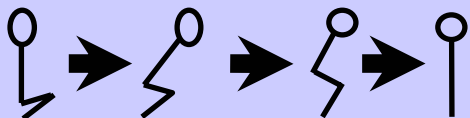
文書化された知識

膨大な文書・Web ページ
管理のための物語データ
ベース



データベース

データマイニング
応用領域：Web検索や
遺伝子情報



モーションのような動き
のある情報

モーションDB + モー
ション合成・編集



データマイニング

- 予兆の発見
- 潜在的に重要な相関関係の発見

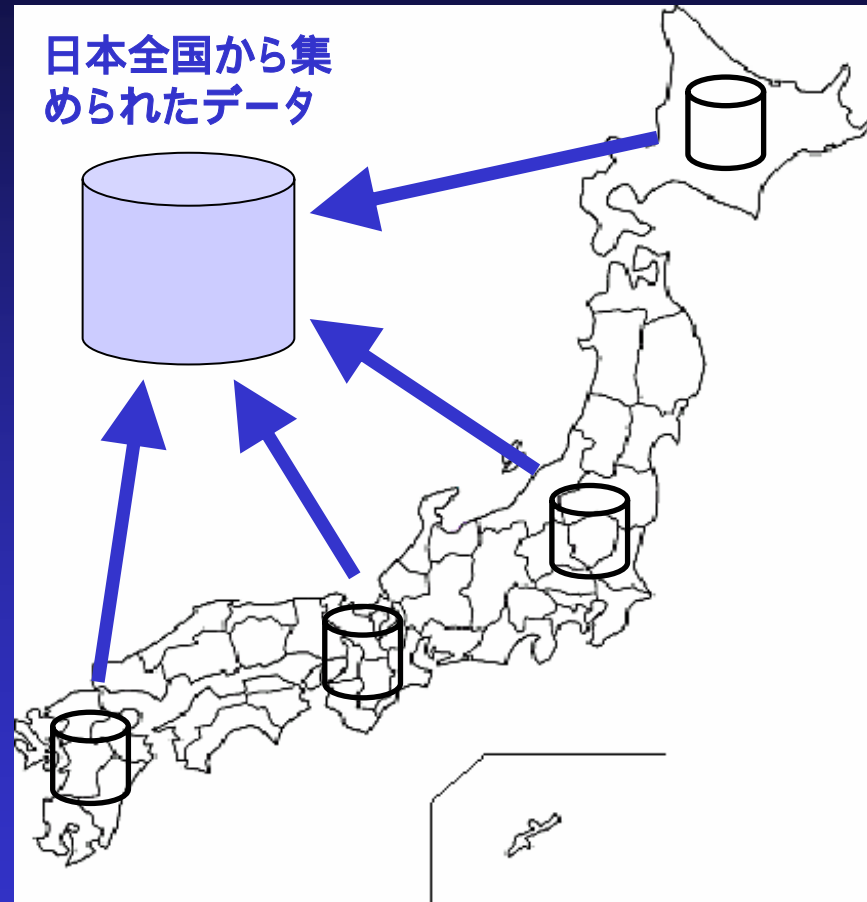
データマイニング: 予兆の発見

POSデータの利用法としてのデータマイニング

全体と部分において、相関(品目間の関連性)に著しい差異があれば、たとえ部分で顕著でなくても、部分で何かが起きている可能性がある。

北海道地区に限定した、関連商品の抱き合わせ販売を促進する、等々...

そうした変化の顕著さを求める

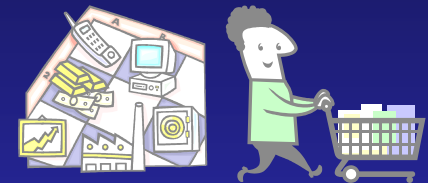


潜在的に重要な相関関係の発見

相関の違い(変化)に注目。
タイムスタンプ、地域、年齢等による条件付け

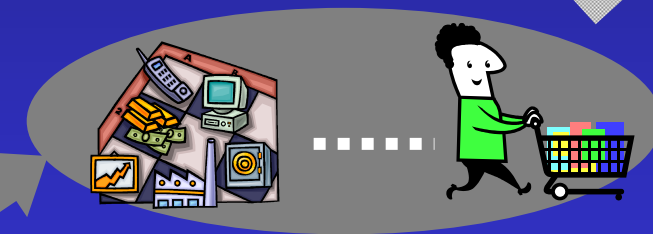


スーパーマーケットの
マネージャー



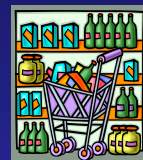
特徴的な関係
になる可能性
もある。

将来において...



顧客のニーズが変わりはじめている商品：
高い相関が観測されとは限らない。
しかし、相関の大きな変化は観測される。

相関：高



一般的な、商品と顧客の関係(定番商品等)：
どの期間でも高い相関が観測されることが多い。

相関：低



ある時系列マーケットDB

最近のタイムスタンプ

物語データベース

オンライン文書を話の筋や展開構造
からデータベース化する試み

- テキスト間の構造類似性
- 結束性と文脈性に基づく文の評価

単一物語から抽出可能な複数の類比関係

- スターウォーズ
 1. 父親は悪組織のリーダーである。
 2. 息子はその悪組織と戦う。
 3. 父親は息子を助ける。
 4. 悪組織は息子に敗北する。

- インディ・ジョーンズ
 1. インディは宝を探す。
 2. インディは悪組織と戦う。
 3. 父親はインディを助ける。
 4. インディは宝を見つける。

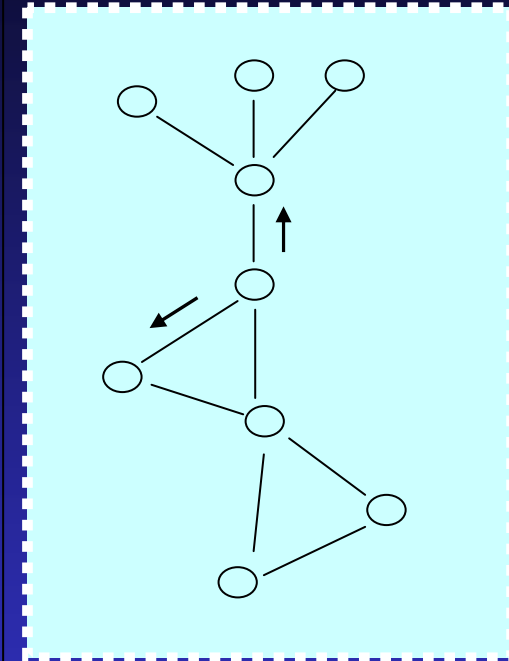
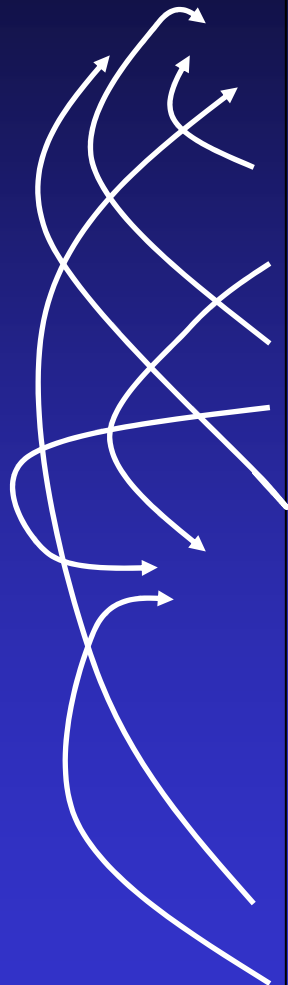
- 007:
 1. ドクター・ノオは、悪組織のリーダーである。
 2. 007 はその悪組織と戦う。
 3. 悪組織は007に敗北する。

- MA1
 1. ある男は、悪組織と戦う。
 2. 父親がその男を助ける。

- MA2
 1. ある男は、悪組織のリーダーである。
 2. ある人物は、その悪組織と戦う。
 3. 悪組織はその人物に敗北する。

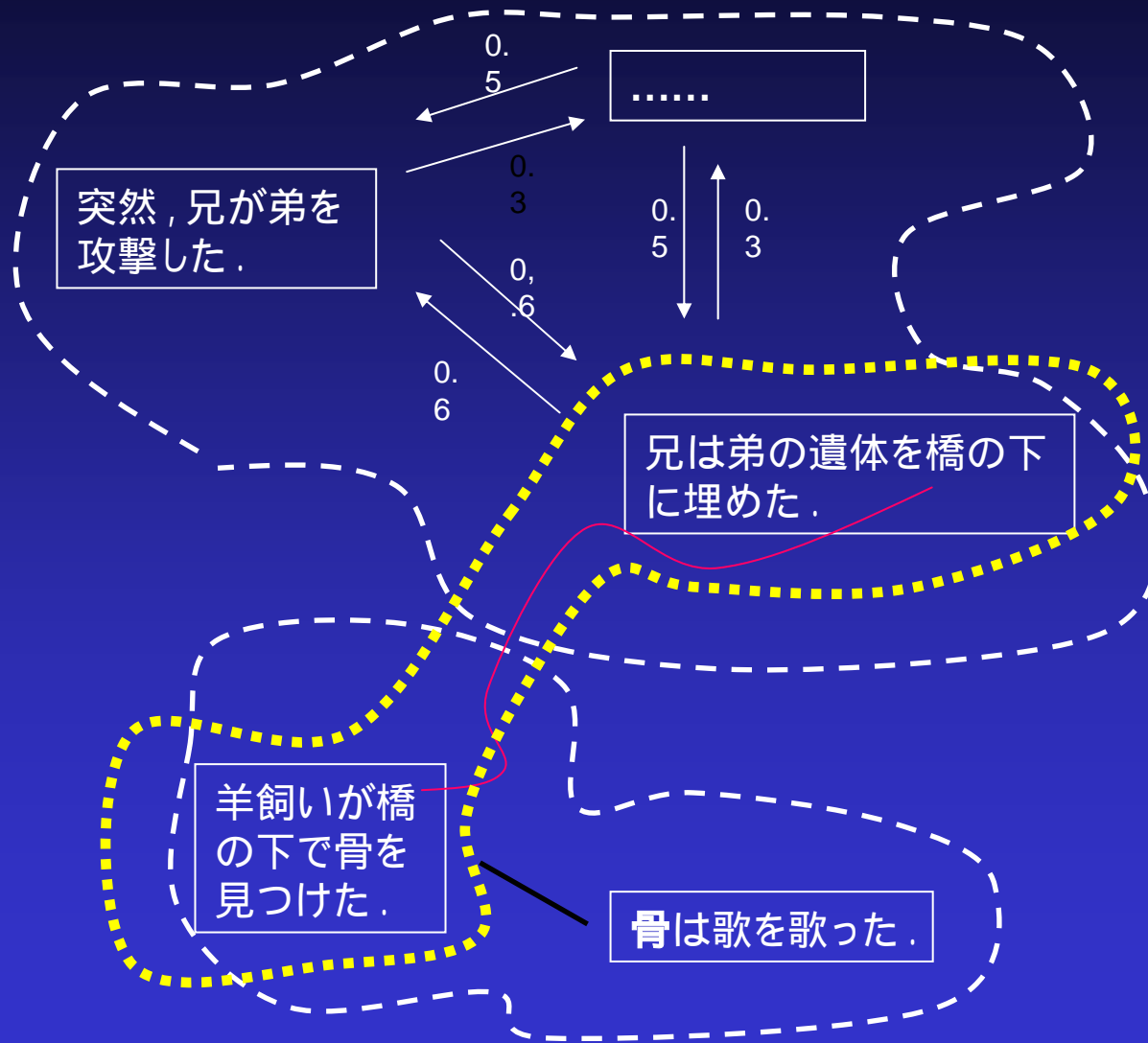
文の情報伝達モデルと物語の構造解析

猪が国を荒らしました。
王は猪をを倒した者に娘をあげると約束をしました。
兄弟は姫を手に入れようと思いました。
弟は森の中に入っていました。
小人は黒い槍を手を持っていました。
弟はその黒い槍で猪を突きました。
弟は意気揚々と家に帰りました。
兄は弟が猪を背中に背負ってやってくるのを見ました。
兄は弟の命を奪ってやろうとたくらみました。
兄は弟を殴り殺しました。
兄は弟の骨を橋の下に深く埋めました。
兄が弟の猪を奪いました。
兄は王のところに猪を持っていきました。
兄は自分が殺したと嘘をつきました。
兄は姫を妻にしました。
羊飼いは骨を拾いました。
羊飼いはその骨で自分の角笛を作りました。
弟の骨が角笛となりひとりでの歌を歌いました。
羊飼いは角笛を王のところに持っていきました。
王は橋の下を掘らせました。
まもなく骸骨が掘り出されました。
兄は自分の行いを白状しました。
兄は水の中に投げ込まれました。
弟の骨は教会の墓地のきれいな墓に埋められました。



擬PageRank による重要度計算

話題連結: 繋ぐ情報



文の塊 (話題) を繋ぐ機能を持つ話題連結語に着目した重要度の付与

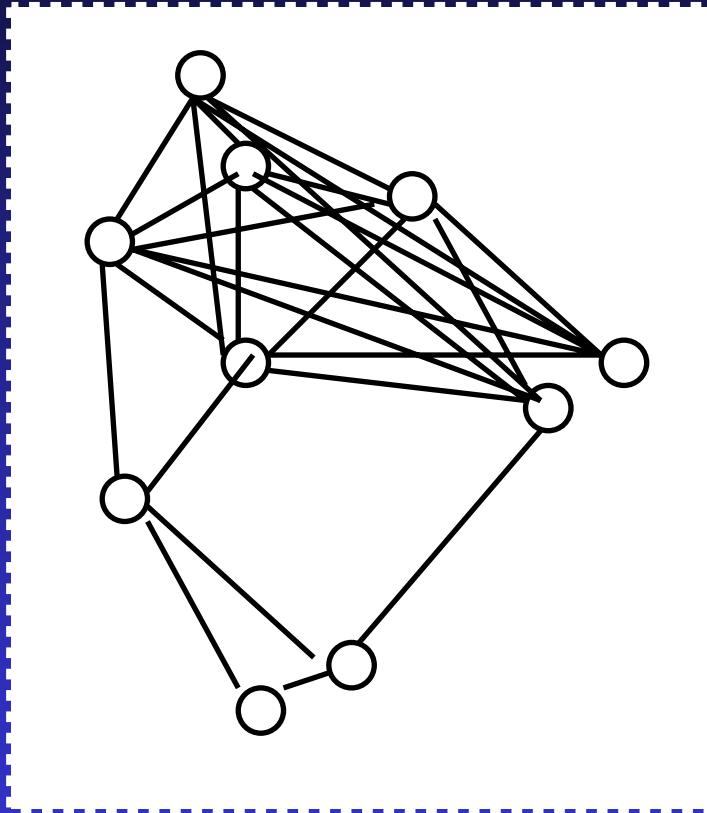
ピンポイントクラスタリング

トップNクリーク探索アルゴリズム

Web ページのクラスタリング

時系列データのトップNクラスタリング

クリーク探索：クラスター抽出のための手法



グラフのクリーク：ノードが密結合になった特徴的な箇所

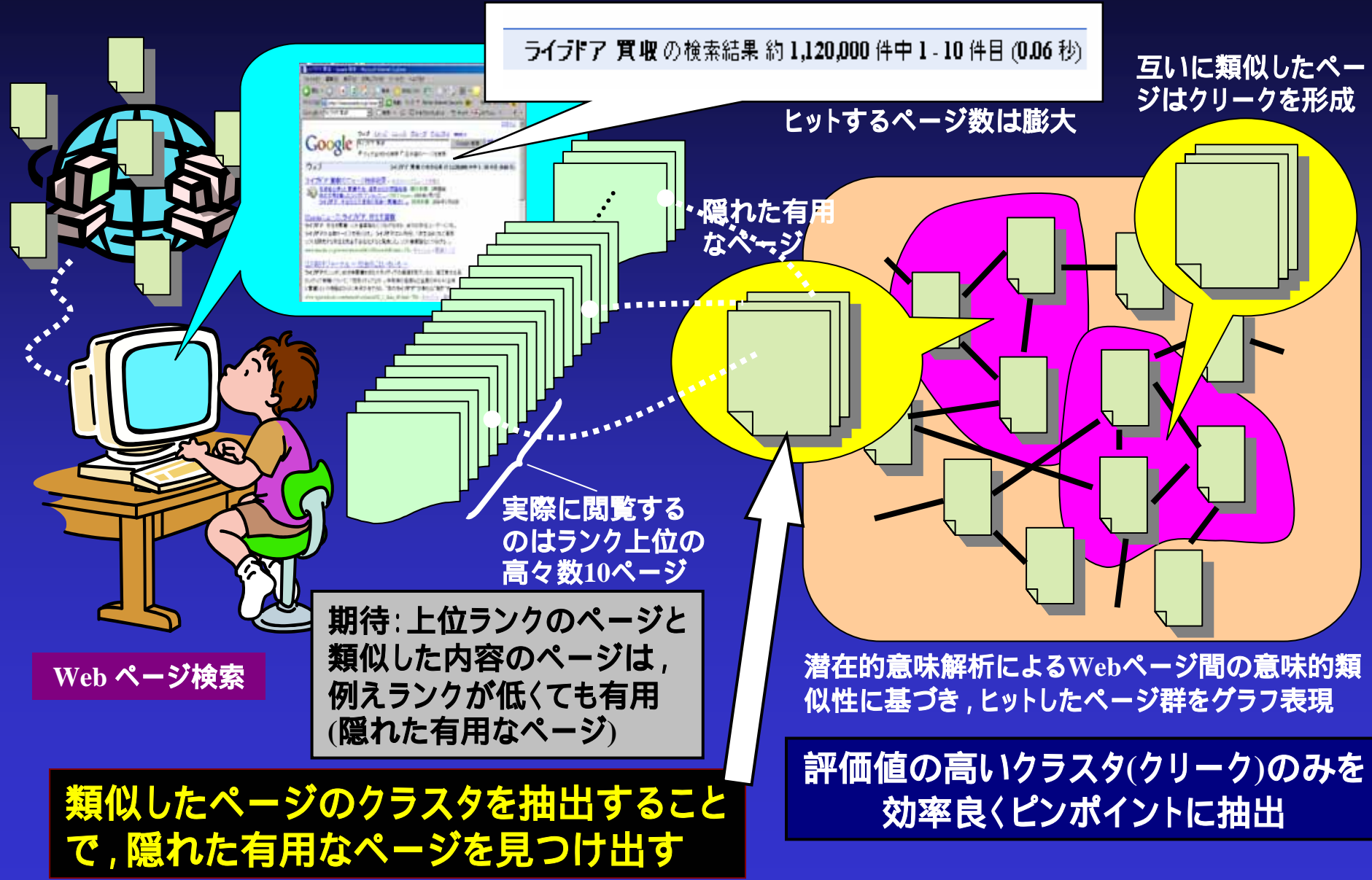
分岐限定・最小彩色数近似などを用いた優れた近似アルゴリズム(アルゴリズム工学)

データをグラフ化してアルゴリズム工学の成果を活用する

応用例：Web 文書の近さ(情報検索で用いる語彙分布の近さ) + リンク構造(as in PageRank)

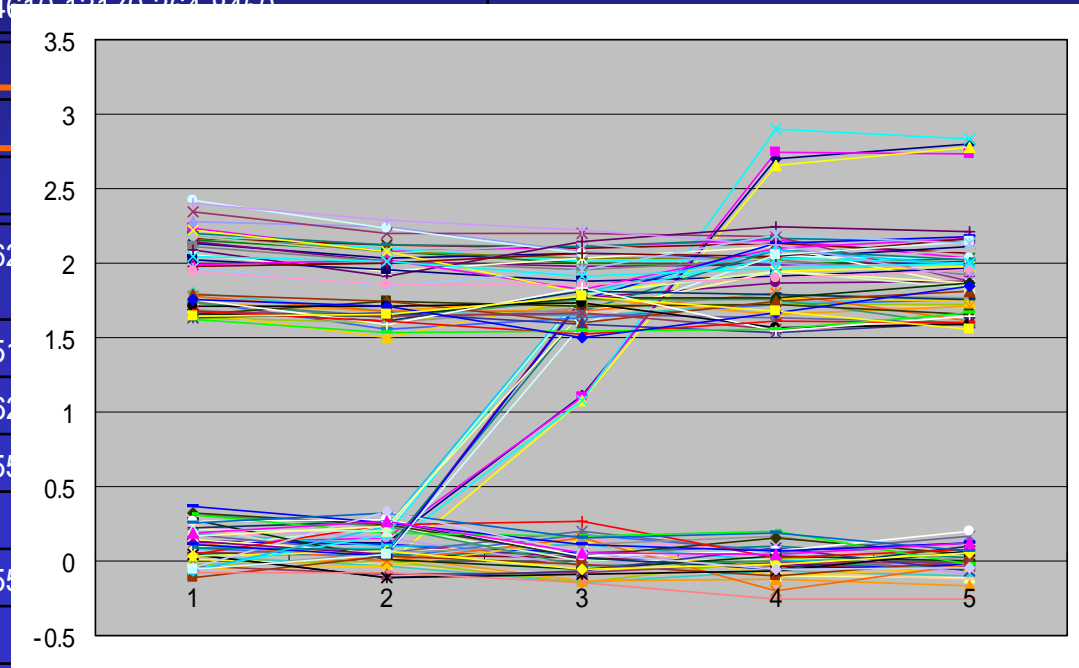
時系列データ：パターンの近さ + 求めたいパターン集合に対する制約を、ノード重みで記述

Webページのピンポイントクラスタリング



時系列データ からの トップNクラスタ 抽出

No.	要素数	クラスタ重み	得られたクラスタ
1	3	2.666667	1812 19368 2555
2	3	2.666667	1812 19368 21652
3	7	0.420238	20999 20165 9971 19489 6069 18713 12090
4	6	0.418253	20999 20165 9971 19489 18713 16248
5	8	0.416813	20999 20165 9971 19489 3158 6069 5045 5001
45	8	0.406984	20999 20165 9971 19489 46069 5001 13870 9074 8450
46	4	0.253556	6069 3962 12090 6019
47	8	0.242615	4206 14635 6989 14146 14610 12170 264 8450
61	5	0.21294	3158 4313 668 9766 4893
62	4	0.21248	17024 10828 1001 13419
63	4	0.212007	7032 19933 6993 7791
92	10	0.172195	20595 7138 12176 14703 6210 10592
93	9	0.170675	20595 7138 19171 12176 5100 10592
94	10	0.168448	20595 7138 19171 12176 6210 10592
95	5	0.16675	4313 20020 668 9766 10555
96	2	0.165263	10828 5294
97	5	0.158648	4313 668 9766 11233 10555
98	3	0.157576	17024 12805 6581
99	5	0.153972	4313 668 9766 10555 4893
100	4	0.15389	19933 6993 7791 5858

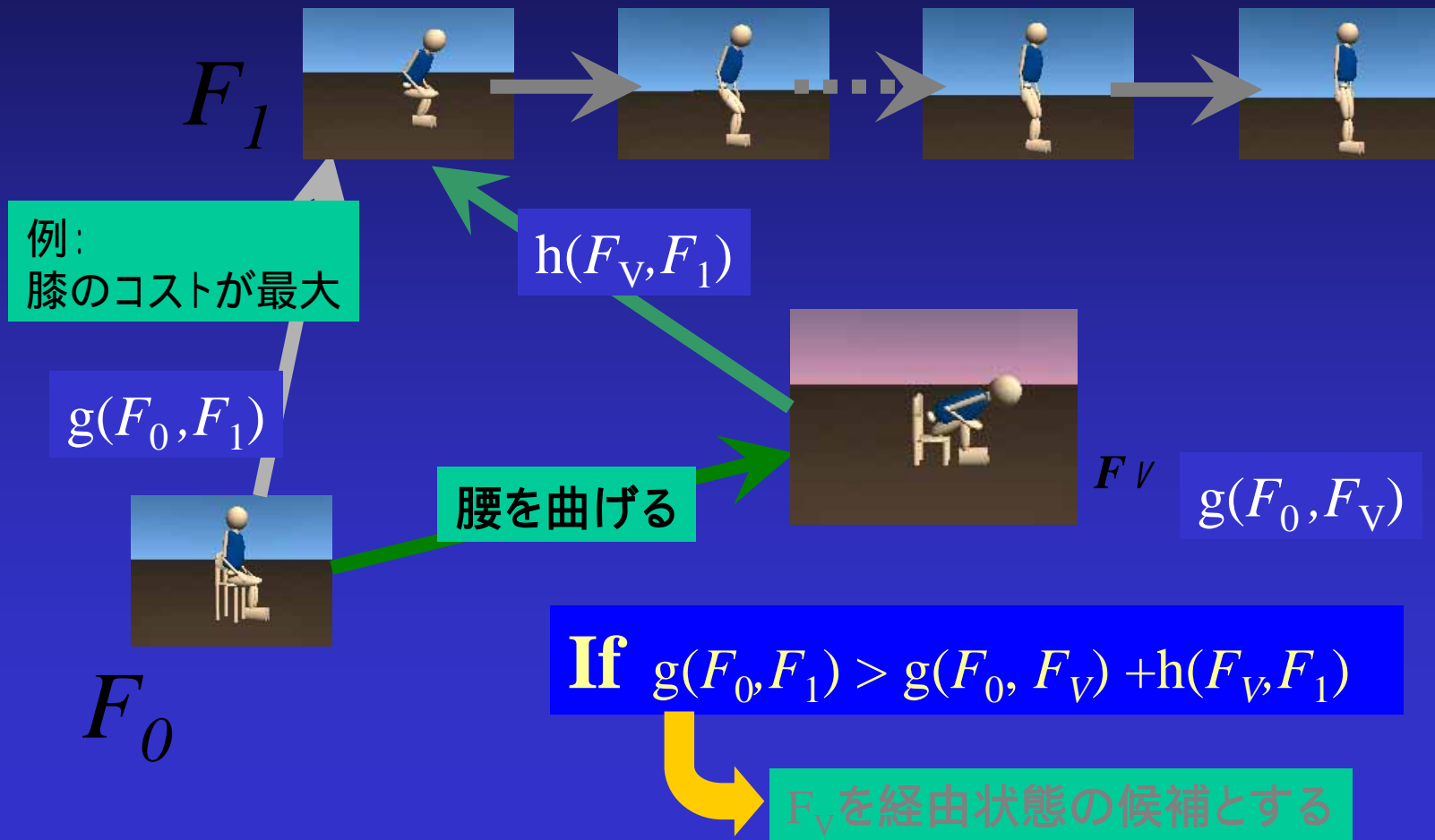


モーションDB

- モーションデータの活用・流通の加速化
- 豊かな表現力を持つ動作バリエーションの生成

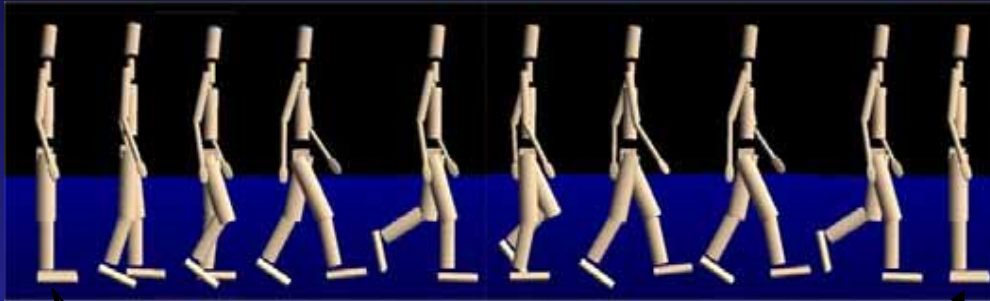
A* (本来AIの記号的探索プログラム)

負荷のかかる関節（関節を動かす筋肉が多くのエネルギーを消費する）に対して、
負荷分散すべく、隣接した関節を少しずつ動かしてみる

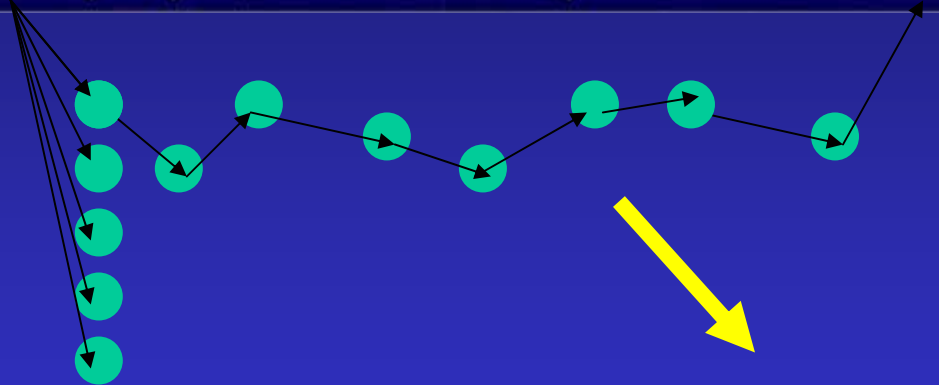


適用処理

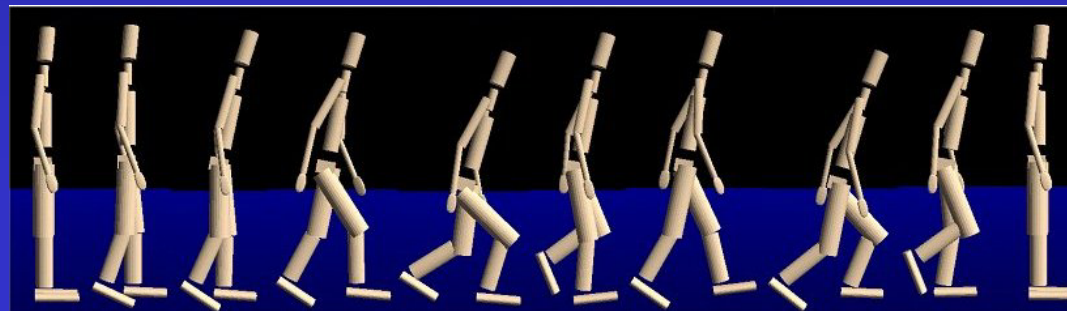
標準キャラクタ

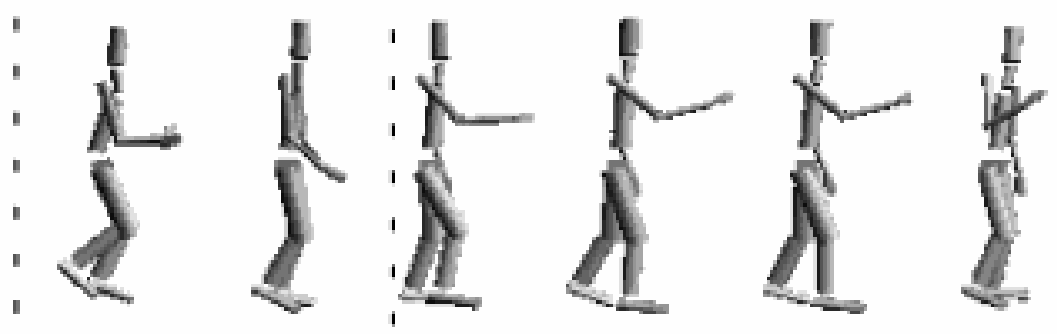
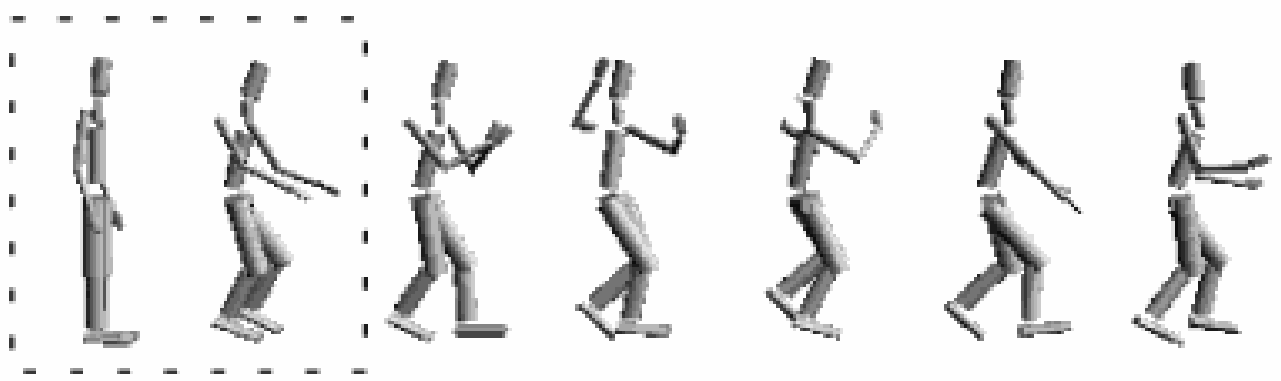


強度パラメータを変えた場合の処理



目標キャラクタ





complementary motion

- モーションDB中の動作を、現在の状況に最適な形で再利用する
- 例えば、右ひざを怪我をした人に、右ひざをかばう動作を生成する。
- 多様で状況依存の動作を、モーションDBの典型的な動作の切り貼り・再利用で実現する