

主専攻実験A 第3回発表会

岡部 純弥

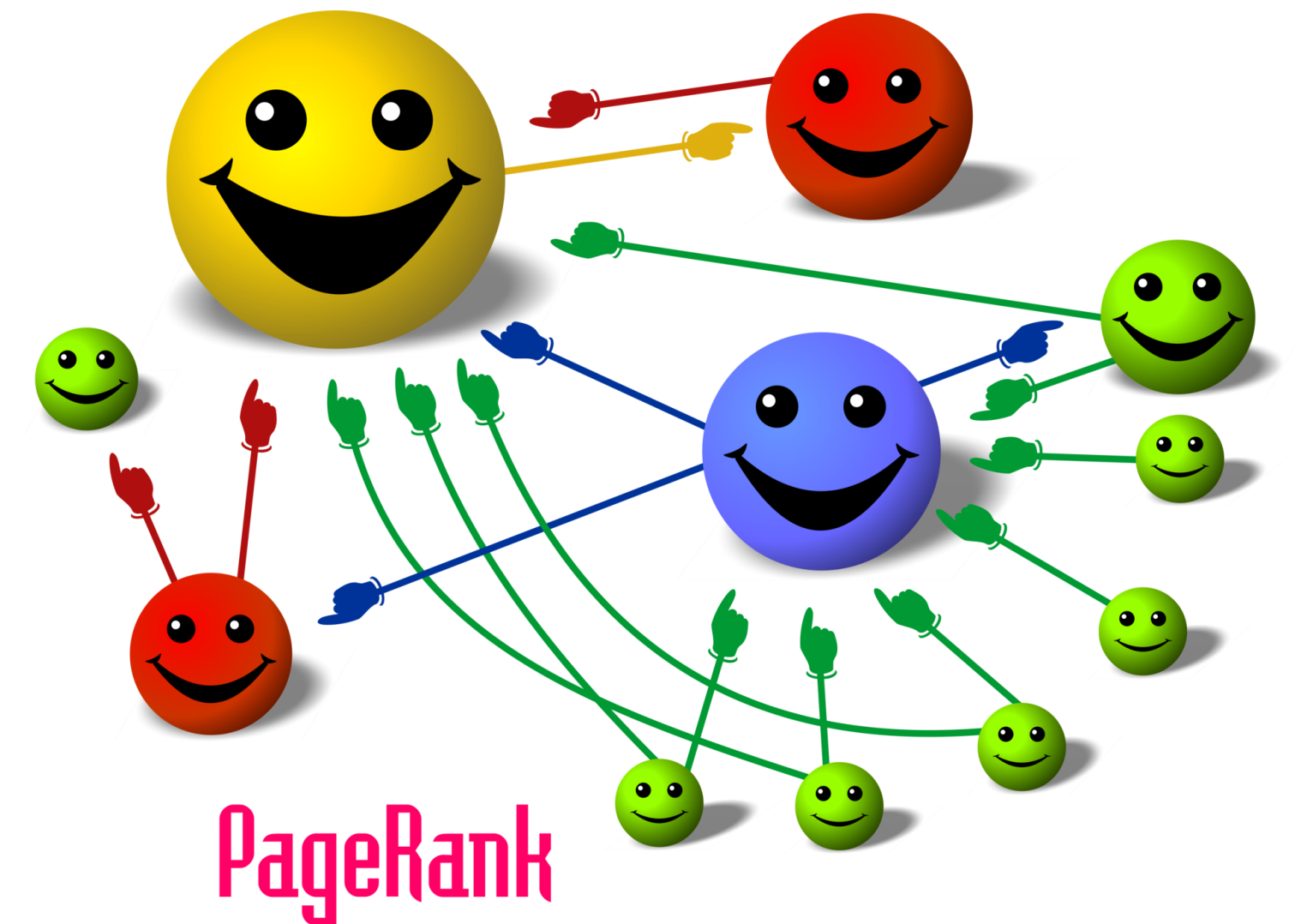
概要

- ・ **ページランクアルゴリズム** を用いた空港の重要度評価
 - ・ L. Page, S. Brin （Googleの創始者）によって提唱
 - ・ 日本の主要7空港の重要度を評価

ページランクとは

概要

- Googleの検索で用いられているアルゴリズム
 - 「良いWebページは別の良いWebページからリンクされている」
- 論文の重要度評価の考え方
- 様々な分野で応用されている
 - SNS分析, 社会科学, 生物学 etc...



ページランクのイメージ図 (Wikipediaより)

ページランクとは

もう少し厳密に・・・

- ・ 各Webページ i は固有の評価値を持っている (P_i と表記する)
- ・ Webページ i の中でリンクが貼られているページ (全部で m あるとする) にそれぞれ O の評価値を与える
- ・ 他のWebページからWebページ i へのリンクが存在する場合, Webページ i に I の評価値が与えられる.
- ・ 評価値が大きいほど重要なWebサイト

ページランクとは

もう少し厳密に・・・

- ・ Webページの総数 n , Webページ i の評価値 P_i
- ・ Webページ i **に**リンクするWebサイトが k つある（それぞれから, I_1, I_2, \dots, I_k の評価値）
- ・ Webページ i **から**リンクするWebサイトが m つある ($m \leq n$)

$$P_i = \sum_{j=1}^k I_j$$

$$O = \frac{P_i}{m} = \frac{\sum_{j=1}^k I_j}{m}$$

ページランクとは

もう少し厳密に・・・

- ・ 前ページのスライドの P_i を固有値問題に帰着させて求める
 - ・ ネットワークフローの問題として解くアプローチもある

$$P_i = \sum_{j=1}^k I_j$$

$$O = \frac{P_i}{m} = \frac{\sum_{j=1}^k I_j}{m}$$

今回の実験

- ・ ページランクのアルゴリズムを用いて空港の重要度ランキングを分析する
 - ・ 空港は 羽田, 成田, 新千歳, 大阪（伊丹）, 関西, 福岡, 沖縄（那覇）の7つ
- ・ 各空港間の年間旅客数を用いる



結果

空港名	羽田	成田	新千歳	伊丹	関西	福岡	沖縄
ランク	0.295	0.083	0.147	0.101	0.076	0.160	0.136

羽田 >> 福岡 > 新千歳 > 沖縄 > 伊丹 > 成田 > 関西

結果

空港名	羽田	成田	新千歳	伊丹	関西	福岡	沖縄
ランク	0.295	0.083	0.147	0.101	0.076	0.160	0.136
総旅客数	20,606,398	1,984,001	6,436,335	5,812,333	2,051,220	6,485,437	6,588,217

総旅客数のデータはe-statより

結果

空港名	羽田	成田	新千歳	伊丹	関西	福岡	沖縄
ランク順位	1	6	3	5	7	2	4
総旅客数 順位	1	7	4	5	6	3	2

考察

- ・ ページランクの重要度の順位 ≡ 総旅客数の順位
 - ・ 那覇以外は2つの順位差が1以下
- ・ 羽田がハブのように機能していた
 - ・ 羽田以外の順位は羽田との往来人数が大きく影響している??
 - ・ 福岡：2位, 成田：6位, 関空：7位
- ・ ハブアンドスポーク / PtoP での結果の差

今後の課題

- ・ 特定の空港を排除したときの変化
- ・ 大規模データの分析への拡張
 - ・ データ収集の難易度
- ・ ライブラリの内部を理解する
 - ・ 高速化の検討