

ルールモデル型eポートフォリオシステムにおける 文書解析を適用したマルチキャリアパス支援



背景・目的

少子高齢化による労働力の減少

キャリアパスを多様化させることにより、
潜在労働力の活用を図る
⇒ **マルチキャリアパス支援**



多様化するキャリアパスの指南として
ルールモデルの役割は大きい

ルールモデル型eポートフォリオ(RMP) によるマルチキャリアパス支援

本学卒業生の学生時代の修学状況、就職後の
キャリアパスをシステマティックに示すだけ
でなく、**キャリアデザイン支援**を目指す。

Ref.小川賀代他, 日本教育工学会論文誌.31.1.51-59(2007)

ルールモデル型eポートフォリオシステム

eポートフォリオとは

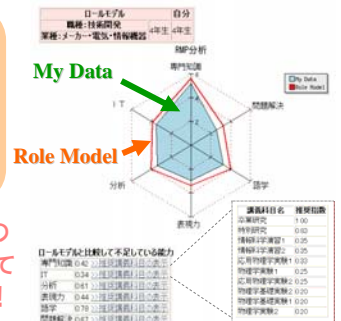
◆ 点在していた個人の学習成果物・履歴などが
一元化され、理解の程度・思考過程が可視化
できる。評価にも活用される



【ルールモデル型eポートフォリオ (RMP) システムの機能】

◆ 卒業生の蓄積データ (成績, 業績, 取得資格) を活用

- ◆ 比較対象の選択
 - ・業種・職業別
 - ・企業別
 - ・自分自身の過去と現在
- ◆ 分析結果の表示
- ◆ 履修推奨科目の提示



ルールモデルの蓄積データの
数値化可能な情報を利用して
自己分析・適職診断を実施!

更なるキャリアデザイン支援を目指して

eポートフォリオには、成績、業績、取得資格の他に
就職活動データ(エントリーシート、履歴書など)や
就業データ(職歴、キャリアパスデータなど)の
文書データも豊富に蓄積されている。



【RMP機能の拡張】

エントリーシート、自己アピール文などの文書データからの
キャリアデザイン支援機能の構築

≪ 検討項目 ≫

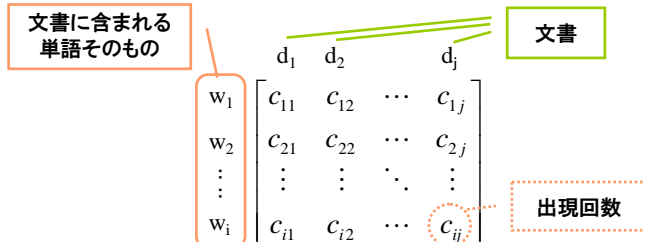
- 文書データによる適職診断
- エントリーシート作成支援(希望企業とのマッチング診断)

ルールモデルの文書データから評価基準を決定

文書解析

【ベクトル空間モデル】

文書内の単語とその頻度を元に作成した行列から
単語や文書に対応するベクトルを定義



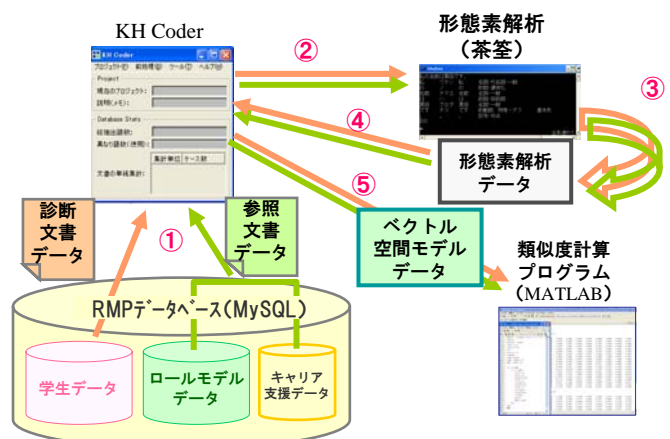
● 行列要素 c_{ij} は単語 w_i が文書 d_j に出現する頻度を表す

【類似度計算】

$$\cos \theta = \frac{d_i \cdot d_j}{|d_i| \times |d_j|} \quad (d_i, d_j: \text{文書ベクトル})$$

● 文書間の類似度はコサイン尺度を用いて算出する
1に近づくほど類似度は高く、-1に近づくほど類似度は低い。

【システムフロー】



■ KH Coder

データベース (MySQL) から文書を抽出し、
茶筌を動かし、ベクトル空間モデル化
したデータをデータベースへ蓄積

■ 茶筌

文書を最小構成単位である形態素
に分割する形態素解析プログラム

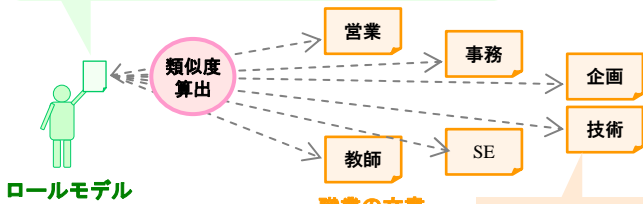
■ ■ 文書データによる適職診断機能の検討 ■ ■

【活用文書データ】

- ◆ 診断文書データ：ロールモデル7名の自己アピール文 (平均400文字)

<自己アピール例文>

私の強みは、向上心が強くフットワークが軽いことです。目指していることに対して知識を増やし技術を磨く努力を惜しみません。…………



- ◆ 参照文書データ：6つの職業の文書 (平均1600文字)



- ・ (独) 労働政策研究・研修機構が情提供 総合的職業情報データベース「キャリアマトリックス」から抜粋
- ・ 本学の卒業生が就いている主な6職種を選択

URL: <http://db.jil.go.jp/>

<職業文書>

精密機械の設計・製作を行うのが精密機械技術者である。…………新しい技術を習得するための学習意欲が求められる。

【類似度算出結果】

文書数 d_j が13, 単語数 w_j が1052の文書単語行列から類似度算出

| 職種 | 営業 | 事務 | 企画 | 技術 | SE | 教員 |
|--------|-------------|-------------|-------------|-------------|-------------|-------------|
| ロールモデル | | | | | | |
| A | 0.1634 5 | 0.1635 4 | 0.2073 2 | 0.1961 3 | 0.2288 1 | 0.1499 6 |
| B | 0.4392 4 | 0.4284 5 | 0.4683 3 | 0.5485 1 | 0.5422 2 | 0.382 6 |
| C | 0.3226 4 | 0.2952 5 | 0.3237 3 | 0.3588 2 | 0.3833 1 | 0.2931 6 |
| D | 0.4259 4 | 0.4101 5 | 0.4529 3 | 0.5587 1 | 0.5041 2 | 0.3703 6 |
| E | 0.4468 5 | 0.4488 4 | 0.4877 3 | 0.4887 2 | 0.5341 1 | 0.3828 6 |
| F | 0.2744 5 | 0.2811 4 | 0.3046 3 | 0.3242 2 | 0.3453 1 | 0.2596 6 |
| G | 0.4118 4 | 0.3852 5 | 0.4147 3 | 0.4512 1 | 0.4504 2 | 0.359 6 |

上段：類似度結果 下段：類似度の高い順の順位 就いた職業

ロールモデルの就いた職業の順位が高いことから自己アピール文と職業データの類似度の高い順位の職種が、職業適性が高いと提示可能。

■ ■ エントリーシート(ES)作成支援機能の検討 ■ ■

【活用文書データ】

- ◆ 診断文書データ：ロールモデル7名のES 47企業分 (平均800文字)
- ・ 通過ES：29企業
- ・ 不通過ES：18企業

<内容>

- ・ 志望動機
- ・ 学生時代力を入れたこと
- ・ 自己アピール文
- ・ 入社して行きたいこと
- 等

- ◆ 参照文書データ：47企業分のデータ
- ・ 各企業のHPの採用情報のページから抜粋
- ・ 抜粋する文書量を2パターン用意

<内容>

短文(400~500文字)

- ・ 求める人物像
- ・ 就職活動者へのコメント

長文(1700~1800文字)

- ・ 求める人物像
- ・ 企業理念
- ・ 就職活動者へのコメント
- ・ 企業の目指すもの
- 等

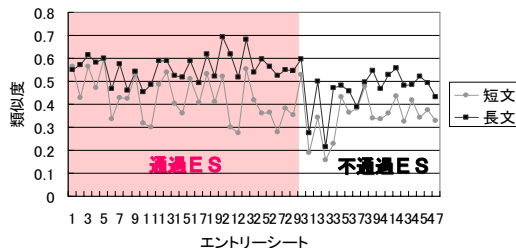
企業1に出したES



<企業データ例文>
向上心が強く、柔軟かつ継続的に学び続けることが出来る人。…………

【類似度算出結果】

文書数 d_j が94, 短文の単語数 w_j が1412, 長文の単語数 w_j が1665の文書単語行列から類似度を算出



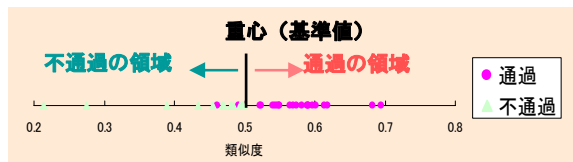
個人ごとの通過・不通過別類似度

| | A | | B | | C | | D | | E | | F | | G | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| | 短文 | 長文 | 短文 | 長文 | 短文 | 長文 | 短文 | 長文 | 短文 | 長文 | 短文 | 長文 | 短文 | 長文 |
| 通過 | 0.451 | 0.537 | 0.436 | 0.535 | 0.45 | 0.544 | 0.412 | 0.628 | 0.39 | 0.569 | 0.364 | 0.564 | 0.339 | 0.54 |
| 不通過 | 0.309 | 0.424 | | | 0.379 | 0.485 | | | 0.381 | 0.511 | 0.452 | 0.545 | | |

通過したESの方が不通過のESよりも類似度が高い傾向

【判別分析を用いた評価】

類似度結果を用いた判別分析



1. 2グループ(通過・不通過)に分けられた類似度から重心の値を算出
2. 重心の値(基準値)より大きいか小さくかでデータの推定グループを示す
3. 実際の各類似度のグループと推定グループの一致度を調べ判別率から分析の精度をみる

【判別分析結果と考察】

判別率的中率結果

| | 短文 | 長文 |
|-----|-------|-------|
| 全体 | 66.0% | 78.7% |
| 通過 | 64.3% | 85.7% |
| 不通過 | 72.2% | 72.2% |

考察

- 長文の類似度結果は78.7%の判別率の中率が得られ、分析精度はやや良いという結果を得た
- 長文での不通過の判別率の中率は通過に比べ低い結果となった要因は、専門や業績が考慮されて選考された結果もあることが考えられる

ロールモデルのESと各企業データの類似度を用いた判別分析から算出した基準値を使用することで、ES作成支援に活用できる可能性を得た。

■ ■ まとめ・今後の課題 ■ ■

- 文書解析による文書間の類似度をRMPシステムへ適用することにより、ロールモデルの蓄積文書データを活用し、キャリアデザイン支援機能として、適正職種診断、エントリーシート(ES)作成支援が行える可能性が得られた
- ES作成支援機能の検討において、企業データの文書量、職業内容の追加などを検討した文書解析を行い精度向上を目指す