

## 時系列情報を用いた研究リソース共有・推薦システム

### Research-Resource Sharing and Recommendation System Using Timeline

真木努<sup>1</sup> 越智洋司<sup>2</sup> 井口信和<sup>2</sup>

Tsutomu Maki<sup>1</sup>, Youji Ochi<sup>2</sup>, Nobukazu Iguchi<sup>2</sup>

<sup>1</sup> 近畿大学大学院総合理工学研究科

<sup>1</sup>Interdisciplinary Graduate School of Science and Engineering, Kinki University

<sup>2</sup> 近畿大学理工学部

<sup>2</sup> School of Science and Engineering, Kinki University

#### 1. 概要

研究分野の多様化に伴う学術論文等の研究リソースの増加から、研究活動における情報 DB や検索エンジンの重要性が一層高まる傾向にある。そこで我々は、研究活動支援の一環としてリソース情報を研究者間で共有し、蓄積されたデータを分析・推薦するシステムの開発に着手してきた[1]。本稿ではその推薦手法と実装方法について述べ、有効性の検証を今後の課題とする。

#### 2. 研究活動におけるリソース調査と課題

##### 2.1 想定する支援環境

研究活動は大別して以下の過程で構成される。

- 1) 研究テーマ選択・決定
- 2) 研究遂行(調査・分析・実験)
- 3) 研究成果発表・蓄積(論文発表)

研究活動とはこれら3つの過程を繰り返していくことで構成され、他のメンバーとのコミュニケーションを通して、サークルを拡大させていく渦巻き型のモデルで定義可能である(図1)。

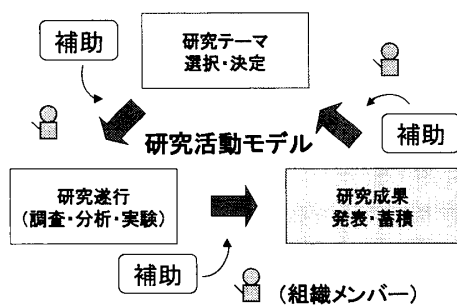


図1: 研究活動モデル

本研究では、一般的な大学の研究室における研究活動を想定しており、対象となるユーザとして研究室の院生・学部生を考える。特に今回開発するシステムでは、2) 研究遂行過程で取り扱われる研究リソースの管理環境を提供する。

##### 2.2 研究リソース

研究遂行過程において調査の対象となる書

籍・論文・学会誌等の文献リソースと補助ツール・ライブラリ等の技術リソースを、本稿では総じて研究リソースと定義する。研究リソースには永続的な増加傾向があり、かつ研究成果として見た場合、研究成果の公開は離散的だが研究の進歩は継続的であるという時間的な特徴を持っている。以下にそのモデル図を示す(図2)。

また、研究分野を研究成果の集合と見た場合の研究リソース(R)と研究成果(RR)、研究分野(C)の相関は次のように定式化することができる。

$$RR = \{R_t\}_{t=1}^n \quad (n: \text{定数}), \quad C \supseteq RR \quad (1)$$

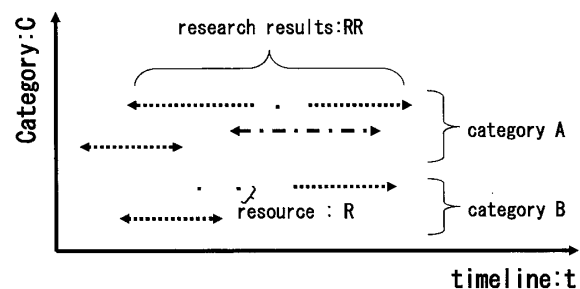


図2: 研究リソースと時系列の関係

##### 2.3 研究リソース調査活動の問題点

研究リソースはその特徴から、将来的に絶対数が莫大な量になりユーザの調査活動に支障を来す可能性がある。本研究の対象ユーザである院生・学部生は研究期間が限られているため、リソース探索には迅速性が要求される。それ故、ユーザは検索要求を正確に情報 DB に伝達する必要があるがキーワードマッチングのみでこれを行うのは困難である[2]。通常、この問題は先行研究者とのコミュニケーションを通して解決を見る場合があるが、必ずしもそのような環境が整っているとは限らない。よって、ユーザの検索要求を情報 DB に伝達することをサポートする枠組みが必要となる。

本稿ではこの問題を解決するにあたり、ユーザの検索要求を以下のように定義する。

- i) 研究成果として将来性のあるリソース
  - ii) ユーザの研究に関連するリソース
  - iii) 研究成果で重要な位置づけを取るリソース
- i) および ii) は分野における成果の位置づけの把握につながる。また、iii) は成果におけるリソースの位置づけの把握につながる

### 3. 時系列情報を用いた研究リソース推薦

#### 3.1 研究目的

2.3 節で挙げたユーザの検索要求を満たすためには、リソース情報を個別に取り扱うのではなく研究分野、研究成果との相関を明確にする必要がある。本研究では、時系列情報に着目することで個別の検索要求に対応したリソース情報をユーザに提示する、研究リソース推薦システムの開発を行う。以下にその推薦手法を示す。

#### 3.2 アプローチ手法

##### 3.2.1 加重移動平均法による近似 ARIMA モデルを用いたトレンド推定

本研究では、2.3 節の検索要求 i) の将来性のあるリソースの算出に、その判断材料として研究分野に対する閲覧数の予測に着目した。将来的に閲覧数が増加傾向のある研究分野に属するリソースは、将来性があるリソースであると考えることができる。本研究では、過去の時系列データから将来の予測値を算出する手法である ARIMA モデル [3] を利用する。ARIMA モデルにおける MA 過程については、加重移動平均法を用いて近似することでシステムの高速度を図る。本手法により、将来の閲覧数を予測することができる。その増加傾向から将来性を算出することができる。

##### 3.2.2 協調フィルタリングによる類似リソース推薦

本研究では、2.3 節の検索要求 ii) のユーザに関連するリソースの算出に、ユーザに類似するユーザ集団に着目した。嗜好が似ているユーザ集合を特定することができれば、そのユーザ集合が高い評価を与えたリソースは、ユーザにとって有益な関連リソースであると考えることができる。本研究では、協調フィルタリングの考え方を元に、ユーザのリソース閲覧履歴からユーザ集団を特定し、そのユーザ集団が評価しているリソースを推薦する。

##### 3.2.3 時系列定常性を考慮したリソース推薦

本研究では、2.3 節の検索要求 iii) の重要な位置づけを取るリソースの算出に、時系的にみた研究リソースの定常性に着目した。つまり、重要な研究リソースとは、ブームやトレンドに

依存せず、時代を問わず定常的にアクセスされるものとする。本稿では、交互作用効果を狙った多変量時系列解析により、対象の安定性を基準とした推薦手法を行う。ただし、重要なリソースとは、ただ単にアクセスが定常だけでなく、リソースに対する評価も重要な要因となるので、

R: リソースの閲覧回数

U: ユーザのリソースに対する評価

S: リソースに対するサーベィの存在数

C: リソースに対するコメントの存在数

B: リソースに対するブックマークの存在数

に着目する。一般的には、ある時刻におけるデータとその前後のデータを比較した場合、平均 (E), 分散 (Var) 共分散 (Cov) のそれぞれの値が一定であれば、定常性が高いと言われている。そこで本稿では、ある時刻  $t$  における (E, Var, Cov) の値を微小時間前のデータ  $t-1$  の値と比較することでデータの定常性を調べ、それによってリソース毎の評価値  $I$  を決定する。ここで得られた  $I$  は定常度を表したものであり、この数値の高いものを重要度が高く、安定性が高いリソースとして推薦する。

### 4. まとめ

本稿では、時系列情報を用いた研究リソースの推薦手法について述べた。本推薦手法は現在試作中のリソース共有システムに実装を薦めており、リソース推薦結果の表示には、時系列表示ツール Timeline (Timeplot) [4] を用いる。今後の課題として、システムの実働期間を設けた長期的な実証実験が挙げられるが、今春より 4 年生に進級するゼミ生を対象に半年間の検証を行う予定である。また、推薦の有効性をどう評価するのもかも考察しておく必要があると考える。

### 参考文献

- [1] 真木努, 越智洋司, 井口信和: 研究活動を支援する研究リソース共有・検索システム, 電気関係学会論文集 (2008)
- [2] 阪本俊樹, 北村泰彦, 辰巳昭治: 競争型情報推薦システムとその合理的推薦手法, 電子情報通信学会論文誌, Vol. J86, No. 8, pp. 608-617
- [3] G.E.P.Box, G.M. Jenkins, "Time series analysis. Forecasting and control", Holden-Day (1970)
- [4] (Smile) Timeline, Timeplot  
<http://simile.mit.edu/timeline/>