(特集論文) 「近未来チャレンジ 2012」

統合環境 TETDM を用いたマイニングツールの開発と利用の実践

Practical Application in Development and Use of Mining Tools with Total Environment for Text Data Mining

砂山 渡
Wataru Sunayama

広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University
sunayama@hiroshima-cu.ac.jp

高間 康史
Yasufumi Takama

首都大学東京システムデザイン学部
Faculty of System Design, Tokyo Metropolitan University
ytakama@sd.tmu.ac.jp

西原 陽子
Yoko Nishihara

立命館大学情報理工学部
College of Information Science and Engineering, Ritsumeikan University
nisiyara@fc.ritsumeik.ac.jp

梶並 知記
Tomoki Kajinami

神奈川工科大学情報学部
Faculty of Information Technology, Kanagawa Institute of Technology
kajinami@ic.kanagawa-it.ac.jp

串間 宗夫
Muneo Kushima

宮崎大学医学部附属病院医療情報部
Medical Informatics, University of Miyazaki Hospital
muneo_kushima@med.miyazaki-u.ac.jp

徳永 秀和
Hidekazu Tokunaga

香川高等専門学校
Kagawa National College of Technology
tokunaga@t.kagawa-nct.ac.jp

keywords: text data mining, total environment, data visualization, graphical user interface

Summary

In this challenge, we develop and distribute an integrated environment to flexibly combine multiple text mining techniques. Text mining techniques include numerous tasks such as salient sentence extraction, keyword extraction, topic extraction, textual coherence evaluation, multi-document summarization, and text clustering. Although tools that individually perform one or more of the above-mentioned tasks exist, it is difficult to integrate and activate multiple tools for a particular task. We attempt to provide the flexibility to integrate numerous tools that exist in the community in our proposed text mining environment. Users can use a customized version of the proposed text mining environment for their specific tasks, thereby concentrating solely on their creative work.

1. はじめに

人工知能学会全国大会の近未来チャレンジテーマとして進められている TETDM (テトディーエム)^{*1} は、複数のテキストマイニング技術を柔軟に組み合わせて使える統合環境を構築し、社会的創造的活動を支援できる環境としての提供を目指している [砂山 11, 砂山 13]。近未来チャレンジでは、5 年以内の目標達成を掲げ、人工知能学会の全国大会において 5 年間セッションを続けることで卒業となる。現在 3 年目 (2013 年の全国大会で 4 年目

に入る) の TETDM は、統合環境の正式公開を 2015 年 4 月と目標を定めて進めている (図 1)。

これまで主に、複数のモジュールを統合的に扱う環境の枠組み作り、ならびに環境内で動作するモジュールの仕様の策定を行い、2011 年 12 月に試用版、2013 年 6 月に版の公開を行い、バージョンアップを続けている。また全国大会や研究会^{*2}での活動を進めている。加えて、2012 年 4 月に TETDM サイトを立ち上げ、統合環境のダウンロードやモジュールのアップロードの手順

^{*1} TETDM サイト: <http://tetdm.jp/>

^{*2} 2012 年 4 月から、人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会でも活動している

図1 近未来チャレンジ TETDM のスケジュール

を含め、TETDM とその関連情報についての公開を行っている。今後は、開発者支援と利用者支援に向けて、構築した統合環境を生かしていくことが課題となる。

TETDM では「テキストマイニング」という用語を「入力として与えられるテキストに関連して行われる処理全般」という意味で定義する。すなわち、キーワード抽出や要約など従来から広く行われている自然言語処理をはじめ、テキストを理解するためにテキストに関連する情報を他のリソースから取得する処理、また学術目的でない、あるいは学術的根拠をもたないテキストの分析結果を出力する処理も含んでいる。

TETDM は以下の3つを達成目標として掲げている。

1. 幅広い利用者と開発者の参入
2. モジュール間での相互インタラクション
3. 知識創発のための基盤環境の構築

1. では、テキストマイニングの研究を直接活用する人だけではなく、一般のネットを利用する人たちや、プログラミングの学習をしたい人たちなど、利用者と開発者の両面で、初心者、初学者の参入が可能な環境づくりを目指している。2. では、統合環境内のモジュールとして、多くのマイニングツールや可視化ツールが集められたときに、それらが独立に動作するだけでなく、互いに連携して動作する環境を目指す。3. では、多くのツールを連携させて使う中で、利用者側の発想を促すプロセスの明確化と積極的な思考支援を行っていく。

これらの目標に向け、本稿では利用者と開発者の分類を行い、その間の関係を明確にし、現在までに実践された内容について述べる。すなわち、目標の2. で挙げたツール間の連動についての実装内容を含め、2章でTETDM 統合環境の構成について述べる。3章で現在までに実践したマイニングツールの開発と利用について、4章で利用者と開発者のスキルアップに向けた既存研究とTETDM の方向性について述べた後、5章で本稿を締めくくる。

2. TETDM 統合環境の構成

本章では、TETDM 統合環境の構成(図2)について述べる。統合環境は「基本テキスト処理部」「テキストデータ/連動処理データ」「連動制御処理部」から構成される

図2 TETDM 統合環境構成図

制御処理部と「マイニング処理^{*3} モジュール」と「可視化インタフェースモジュール」のペアによる出力パネルからなる。

本統合環境は、多くの開発者によって作成されるモジュール集合を、単にそれぞれが独立に動作するだけでなく、それらを協調的に連動させることができる点が新しい。そのため、多様なモジュールが動作可能となること、またモジュール作成のしきい(プログラミングに必要な知識、作成の手間、仕様上の制約)を低くすること、を考慮した構成となっている。

特に、一般的に出力に用いられる可視化インタフェースモジュールを、利用者対話的な操作を行うための入力インタフェースとしても用いることができ、任意のモジュールが、モジュール間で共通の観点(2.3節、2.4.1節で後述のFocus型変数の値)に基づいて、協調的な再処理動作を行える点が新しい。すなわち、利用者がインタフェース上でマウスで触れたりクリックしたデータを観点として、他のモジュールが観点を中心とした、あるいは観点に関連する情報を即座に提示する事で、容易に情報を掘り下げて閲覧できる。これによりインタラクティブな分析作業に没入できる環境を目指している点が、既存のソフトウェアでは実現できていない特徴となっている^{*4}。

以下で、統合環境の各部の詳細について述べる。

2.1 入力: テキスト

本環境には単一のテキストを入力として与える。複数のテキストを入力する場合には、フォルダを指定することで、フォルダ内のテキスト集合が1つのテキストに結合された後に入力となる。

入力されたテキストは「セグメント」「文」「単語」の3つに分割して扱う。「単語」へ区切る際は、形態素解析器を用いて単語に分割する。この際、指定した品詞の単語だけを、キーワードとして取り扱う「文」に区切る方法は、テキスト中の句点記号(「。」や「.」)をもとに分

^{*3} 本稿では、1章で定義したテキストマイニングを行う処理の意味で用いる。

^{*4} 既存のソフトウェアとの関係は、[砂山11]を参照頂きたい。

表 1 テキストデータ (TextData 型) のメンバー変数の例

メンバー変数	意味
originalText	入力テキスト
segment[]	セグメント情報
segment[].segmentText	セグメントのテキスト
segment[].sentenceNumber	セグメントの文数
segment[].wordNumber	セグメントの単語数
sentence[]	文情報
sentence[].sentenceText	文のテキスト
sentence[].word[]	文内の単語
sentence[].wordNumber	文内の単語数
keyword[]	キーワード情報
keyword[].partOfSpeech	キーワードの品詞
keyword[].frequency	キーワードの頻度

割する「セグメント」に区切る方法は、特定の文字列をテキスト中に挿入し、その文字列をもとに分割する*5。

2.2 基本テキスト処理部

基本テキスト処理部では、環境内のデータ構造として扱うテキストデータを生成する。現時点では、形態素解析、単語の出現情報や頻度情報の計算、キーワード、文や段落間の関連度計算を行っている。

2.3 テキストデータと連動処理データ

入力されたテキストに前処理を施した後のテキストデータを、TextData 型のインスタンスに格納する。このインスタンスが、各モジュールへの入力となる。表 1 に TextData 型のデータ構造の一部を示す。

合わせて、モジュール間の連動処理に関わるデータ構造を用意する。現時点では、あるモジュールで注目しているデータを共有するためのデータ構造 (Focus 型) (表 2) を、TextData 型のサブクラスとして実装している。

2.4 連動制御処理部

連動制御処理部では、統合環境上で動作するモジュール間の連動を制御し、また連動を実施する処理を行う。これにより、多様なモジュールの柔軟な組合せを実現し、各モジュールの単体としての使用にとどまることなく、組み合わせた使用を前提に、スムーズに深い考察が行える環境としての利用を見込んでいる。連動動作には「フォーカス情報による連動」「オプションによる連動」「データ取得による連動」の 3 種類が実装されている。連動処理制御部では、これらの呼び出しに際して、ループ等の不整合が起こらないように制御を行う。以下で、これら 3 つの連動処理について述べる。

§1 フォーカス情報による連動

利用者が、統合環境の出力内で着目 (マウスで触れるまたはクリックした) しているデータの情報を、TextData.Focus 型の変数に保存して、モジュール間で共有す

*5 現時点では、指定された文字列 (デフォルトでは「スナリブラフト」) によりセグメントに分割している。

表 2 フォーカス情報 (Focus 型) のメンバー変数の例

メンバー変数	意味
focusKeywords[]	キーワード集合
focusSegments[]	セグメント集合
focusSentences[]	文集合
mainFocusKeyword	第 1 キーワード
subFocusKeyword	第 2 キーワード
mainFocusSegment	第 1 セグメント
subFocusSegment	第 2 セグメント
mainFocusSentence	第 1 文
subFocusSentence	第 2 文

ることで、モジュール間の協調動作を実現する。フォーカス情報として、モジュール間で共有できるデータの一覧を、表 2 に示す。これにより、フォーカスされている情報を各モジュールがハイライトして表示することなどが可能となっている。

§2 オプションによる連動

可視化インタフェースモジュール、ならびにマイニング処理モジュールにおいては、それぞれの処理を指定のメソッド内に、オプション番号とともに記述する仕様となっている*6。そこで、各モジュール内において、他のモジュールの処理を、そのモジュールの ID とオプション番号を用いて呼び出して動作させることができる。これにより、あるモジュール内の処理と同時に、他のモジュール内の処理を実行することが可能となっている。

§3 データ取得による連動

マイニング処理モジュール内において、他のモジュールの処理結果を利用したい場合、他の処理モジュールを、モジュール ID とオプション番号を用いて実行させ、その処理結果を取得することができる。これにより、複数のモジュールを組み合わせた処理を実現することが可能となっている。

2.5 マイニング処理モジュール

マイニング処理モジュールは、統合環境内のテキストデータをもとに、テキストの理解や分析に役立つ情報をテキストから抽出する。また 1 章で述べたように、入力として与えられるテキストに関連して行われる処理全般を対象とする。現在までに 20 以上の処理モジュールが作成、公開されている (表 3)。マイニング処理モジュールの処理結果は、可視化インタフェースモジュールに渡されて出力される。

2.6 出力：可視化インタフェースモジュール

可視化インタフェースモジュールは、マイニング処理モジュールによる出力結果を可視化する*7。可視化インタフェースモジュールでは、入力として受け取れるデータ

*6 switch, case 文により処理を選択的に記述し、その case 番号をオプション番号と呼んでいる。

*7 マイニング処理モジュールの結果によらず、統合環境のテキストデータを可視化するモジュールであっても良い。

表3 マイニング処理モジュールの例

名称	処理内容 (ペアとなる可視化モジュール番号)
1. マイニングなし	処理なし (8,11)
2. エディタ	テキストエディタ (1)
3. 長文チェック	文章中の長い文を抽出 (4)
4. 光と影データ	各文の主題との関連度を評価 (2,3)
5. 要約 (展望台)	文章要約 (展望台システム)(1,5)
6. 関連チェック	2つのセグメント間の関係性を評価 (4)
7. 単語間関連度	キーワードマップのデータ作成 (6)
8. タグデータ	文のタグクラウド用データ作成 (7)
9. キーフレーズ	タグクラウド用データ作成 (API を利用)(7)
10. 川下りラベル	文章の主題についての一貫性を評価 (9)
11. アノテーション	文章中の着目すべき文を抽出 (10)
12. 単語チェック	特定の単語の出現をチェック (4)
13. テキスト分析	テキストの評価結果をまとめる (4)
14. 主語抽出	テキストの各文から主語を抽出 (2)
15. 国語辞書	オンライン辞書から単語の意味を取得 (14)
16. Twitter	twitter からつぶやきを抽出 (2)

表4 可視化インタフェースモジュールの例

名称	可視化内容
1. テキスト	テキストを表示
2. テキスト (カラー)	文の背景色とともにテキストを表示
3. 分布	横向きの棒グラフを表示
4. Html テキスト	html 形式でテキストを表示
5. キーワード (展望台)	展望台システムによるキーワード表示
6. キーワードマップ	キーワード集合を関連度に応じて表示
7. タグクラウド	各文の単語をタグクラウド形式で表示
8. セグメント独自性	独自性の高いセグメントを明示
9. 川下り	文章の一貫性を川の流れに喩えて表示
10. アノテーション表示	注視すべき文を明示
11. 連動可視化	注目する単語, 文, セグメントを表示
12. セグメント木	セグメント間の関係を木構造で表示
13. 段落関連度	セグメント間の関連度を表示
14. デュアルテキスト	2つのテキストを上下に表示
15. 文選択	注目する文を選択するチェックボックス
16. レーダーチャート	データをレーダーチャートで表示

型 (boolean, int, double, String 型とその一次元配列, 二次元配列 (String 型以外) の 11 種類) を定め, そのデータの意味を表す整数型変数との組合せにより, マイニング処理モジュールからデータを受け取ることができる. 現在までに 20 以上の可視化モジュールが作成, 公開されている (表 4) .

2.7 統合環境の動作例

「浦島太郎」のテキストデータを入力として, 表 3 の 13 番の「テキスト分析」とこれに関連するモジュールをセットした統合環境を図 3 に示す. 左端のパネルにセットされている「テキスト分析」モジュールは, 右の 6 つのモジュールから「データ取得による連動」により処理結果を得て, その結果を可視化インタフェース「Html テキスト」において, html 形式でまとめて表示する. 左から 2 つ目のパネルにセットされている可視化インタフェース「キーワード (展望台)」においては, 山の上部にある線より上に, 利用者が主題と考えるキーワードをフォーカ

表5 利用者の分類基準と特徴例

基準	尺度	特徴例
PC	低	キーボードをスムーズに打てない トラブル発生の際, 自力で解決しようとしにくい
	中	特殊キーを除いてタッチタイプができる トラブル発生の際, 自力で解決を試みる
	高	ショートカットキーや便利な操作方法を知っている トラブル発生の際, 問題の切り分けを行い解決を図れる
TM	低	テキストマイニングという言葉を知らない テキスト処理に関する知識がない テキストマイニングの利用を必要としない仕事をしている
	中以上	テキストマイニングの活用例が想像できる テキストマイニングを仕事などで利用したことがある テキストマイニングの積極的な活用が求められている

図4 利用者の分類

ス情報として配置することができ (各キーワードはマウスで移動可能), 配置後にマウスをクリックすることで, 「フォーカス情報を用いた連動」により連動に対応しているパネル (左から 1,3,5 番目) の出力が更新される. このように, 複数のパネルを用いて, 利用者の興味に応じた出力結果をインタラクティブに取得できる環境となっている.

3. マイニングツールの開発と利用の実践

本章では, TETDM 統合環境を用いて現在までに実践した内容を, 想定される利用者と開発者の分類に関係づけて述べる.

3.1 利用者と開発者の分類とその関係

本節では, TETDM 統合環境に関わる人々を, 統合環境とツールを利用する「利用者」と, 統合環境上で動作するツールを作成する「開発者」の 2 つに大別し, それぞれのスキルや能力に基づく分類を示す.

§1 TETDM 統合環境の利用者の分類

TETDM 統合環境の利用者の分類において, 「パソコン操作力 (PC)」と「テキストマイニング活用力 (TM)」の 2 つを基準とし, 利用者を 5 つに分類する. 分類基準の特徴例を表 5 に, 利用者の分類を図 4 に示す.

図 4 の縦方向の矢印は, パソコン操作力に関するグルー

図3 TETDM 統合環境 (Ver.0.46) の出力画面 (7つのパネルにセットされているマイニング処理と可視化インタフェースモジュールは、左から順に表3と表4の番号のペアで、(13,4),(5,5),(4,2)(14,2)(10,9)(12,4)(3,4))

表6 開発者の分類基準と特徴例

基準	尺度	特徴例
PG	低	プログラムのアルゴリズムを考えることができない
	中	逐次・分岐・反復処理の仕組みがよくわかっていない
	高	アルゴリズムを考えて、プログラムを作成できる ポインタ、クラス、構造体を用いたプログラムが作成できる 実行効率を考えたアルゴリズムやプログラムが作成できる オブジェクト指向について理解している
RD	低	業務・個人問わずシステム開発の経験がない 何かを作るより、既存のものを使うことを望む 新しいものを想像・創造する意欲に欠ける
	中以上	システムや製品の開発経験がある アイデアの発想を必要とする仕事をしている 新しいものを想像・創造する意欲がある

プログラムの移行 (スキルアップ) を表し、横方向の矢印は、テキストマイニングの活用に関するグループの移行 (スキルアップ) を表している。次節以降で、この各利用者のグループに対して実践した内容、ならびにグループ移行のためのスキルアップに関する支援と関連研究について述べる。

§2 TETDM 統合環境の開発者の分類

TETDM 統合環境の開発者の分類において、「プログラミング力 (PG)」と「研究・開発力 (RD)」の2つを基準とし、開発者を5つに分類する。分類基準の特徴例を表6に、開発者の分類を図5に示す。

図5の縦方向の矢印は、プログラミング力に関するグループの移行 (スキルアップ) を表し、横方向の矢印は、研究・開発に関するグループの移行 (スキルアップ) を表している。次節以降で、この各開発者のグループに対して実践した内容、ならびにグループ移行のためのスキルアップに関する支援と関連研究について述べる。

図5 開発者の分類

3.2 利用者のスキルアップにつながる実践

本節では、図4のU-1グループに属するTETDMの利用者が、U-3グループに移行する手助けとして実践した内容について述べる。

§1 ツール選択支援インタフェースの開発

C大学では、特別研究 (卒業研究) として、テキストマイニング初心者のためのTETDMにおけるツール選択支援を目指したインタフェースの開発を行った*8。TETDM

*8 TETDMは全体がオープンソースとなっているため、各自の使用状況に応じて、独自のインタフェースを開発することも可能となっている。

図6 ツール選択支援インタフェース

(Ver.0.46)においては、処理モジュールと可視化モジュールがそれぞれ20以上あるが、それらの選択は各パネルごとにそれぞれをセットするようになっている。ツール選択はテキストマイニングを行う上で最も基本的なスキルになると考えられるため、全てのパネルにツールをセットするための環境として、ツールの一覧を確認でき、可能な処理ツールと可視化ツールの組合せを確認できるインタフェースの開発(図6)を行った。

図中のノードが各モジュールを表し、内側に処理ツール、外側に可視化ツールが並べられている。また、組合せ可能な処理ツールと可視化ツールとの間にリンクが表示される。ユーザは、このノードをインタフェース上部の白い四角のエリア内にマウスでドラッグして移動させることで、対応するパネルにツールをセットできる。

類似する従来の環境では、1つのデータに対して複数の可視化処理を行って並列に提示するものが多かった。たとえば、orange[Plaisant 08]やWeka[Hall 09]などが挙げられる。TETDMでは、1つのデータに対して、データそのものに対する処理や、そのデータに関係する処理など、複数のマイニング処理を並列に行い、処理結果を並列に提示できる。

この研究を行った学生は、引き続き、モジュール間の入出力関係や、複数のモジュール間の連動処理について、モジュールの組合せと処理の全体像を直感的に把握できるインタフェースになるよう研究を進めており、利用者のスキルアップを図りつつ、自身のテキストマイニングスキルの向上にも努めている。

§2 学部3年次学生を対象にしたセミナー

A 大学情報学部の3年次学生(4名)を対象にした、情報検索やテキストマイニングに関するセミナーにおいて、TETDMを活用している。このセミナーは、情報検索やテキストマイニングに関する理論やアルゴリズムを確認

することを目的としており、実際にプログラムを作成して確認する利用者についての実践例を以下で述べる。

題材の1つとして、ベクトルを用いた文章の表現を学び、2つの文章間のコサイン類似度を求めるプログラムを作成してもらった。作成には、用意されているテキストデータ(TextData型)のメンバー変数のうち「文」をベクトル「キーワード」をベクトルの要素としたベクトル間の関連度を求める変数(RelationData型 sentenceRelation)とそのメンバー変数であるベクトル間のコサイン類似度(double cos[][]: 各ベクトルの要素は0または1としたときの値)を利用した。これにより、2つの文章からなるテキストファイルを読み込み、作成したモジュールによって求められたコサイン類似度と、手計算で求めたコサイン類似度が一致するか確認することで、テキストマイニングに関する理解を深めてもらうことができた。

また同セミナーでは、「マイニング結果をユーザにどのように提示するべきか」「ユーザへの提示の仕方に適するマイニングは何か」を学習してもらうための実践を行った。教授者が、可視化モジュールとしてレーダーチャートを表示するモジュール(RadarChart)を構築して学習者へ提示し、学習者には、このレーダーチャートでの可視化に適するマイニングとその処理結果について考えさせた上で、マイニング処理モジュールの作成を試みてもらった。モジュール作成の際には、他のモジュールと連携し、他のモジュールで得られたマイニング結果を利用することも許可した。結果として、テキストの主題語*9が含まれている文やセグメントの割合、主題語を主語とする文の割合など、主題語の出現割合を抽出するものが作成された。

この実践例におけるTETDMの特徴は、テキストマイニングのための処理に特化したプログラミングにより、テキストマイニングの処理を実体験的に理解できる点となっている。すなわち、前処理、処理モジュール、可視化モジュールが分離されているため、テキストマイニングに不可欠な前処理である形態素解析部分や、処理結果を視覚的に出力する部分をプログラミングすることなく、必要なマイニング処理のみを記述して動作させることが可能となっている。

3.3 開発者のスキルアップにつながる実践

本節では、図5のD-2グループの開発者が、D-4グループに移行する手助けとして実践した内容について述べる。

§1 大学院博士前期課程の学生を対象にした演習

B大学の大学院博士前期課程の演習で、6名が1チームになり、「Twitterから社会をセンシングするツールを作る」というコンセプトで、実質3ヶ月程度でツール(モジュール)の仕様策定と作成を行ってもらった。

結果として、標準語 vs. 方言や、同じ事を表す二つのスラングなど、対となる単語を含むツイートを集めて、

*9 既存の処理モジュール、Panoramicの出力結果を利用した。

どちらが多いかを表す割合を可視化する、というツールの作成に取り組んだ。作成には、ツイートを取得する既存ツール (TweetExtraction) を利用した。また、TETDM では複数のパネルを並列に表示しての利用が可能のため、(a) キーワードを指定し、それを含むツイートを検索するパネル、(b) 複数の (a) のパネルから検索結果件数を受け取り、それぞれの割合を可視化するパネル、の組合せで動作する環境を作成している。

C 大学の大学院博士前期課程の講義の演習では、この講義を履修している人が個別に「文章の作成に役立てられる、テキストを評価するツールを作成」する課題に取り組んだ。

結果として、文章の段落ごとの文字数バランスの調整を支援するツールや、文章の語尾や曖昧表現の使用量によって文章のタイプを動物で表すツールなどが作成された。なお前述の、ツイートを取得するツールの原型は、前年度の本演習で作成されたものとなっている。

これらの演習における TETDM の特徴は、処理モジュールと可視化モジュールとが分離していることにより、共通の可視化モジュールの利用を前提として、作成した処理モジュールの内容、面白さ、アルゴリズムを比較しやすい点が挙げられる。ただし、複数人が複数のモジュールを個別に作成し、それらを組み合わせて動作させる場合、従来 1 つのシステム内にまとめて処理を記述できていたところを、明確に処理モジュールと可視化モジュールに分離させる必要があり、各処理をどちらのモジュール内に記述するかを事前に十分に設計する必要がある。TETDM では可視化モジュールの作成について「原則として汎用性を重視して、可視化処理のみを実装する」という指針を設けており、開発に不慣れな学生は初めは戸惑う可能性があるが、モジュールの独立性を高めることや、汎用性を考慮することは、開発スキルの向上には有益になると考えられるため、今後このモジュールの切り分けを支援できる方策を検討する。

§2 探索型検索支援モジュールの開発

D 高等専門学校では、専攻科生 1 名が特別研究 (大学の卒論に相当) として、探索型検索 (exploratory search) を支援する環境の開発を半年間で行った。

得たい情報が不明確で、検索結果より知識を得ながら検索を進めていく探索型検索においては、検索結果のスニペットやリンク先のページを読み、次の検索クエリを採り出すことが行われる。この読むという手間をなくし、ユーザの目的にあったキーワードを推薦するシステムを TETDM のモジュールとして開発した (図 7)。

キーワードの推薦は、自己組織化マップと決定木の処理の組合せとして実装した。すなわち、まずキーワード検索の結果を Web API により取得し、各検索結果の URL とスニペットのペアをセグメントとしたテキスト (図 7 左下) を入力とする。統合環境制御処理部により TextData 型インスタンスが生成された後、その URL とスニペット

図7 探索型検索の支援環境

が含む名詞集合を用いて SOM 処理モジュール (図 7 左パネル) が単語ベクトルを生成する。その結果をもとに、R の java インタフェースを利用して、自己組織化マップが MAP 表示パネル (図 7 中央パネル) に出力される。そのマップ上で、ユーザに興味のある単語と興味のない単語に分類してもらい、興味のある単語が多い URL を正例とした決定木を、決定木処理モジュール (図 7 右パネル) と Weka により作成し、決定木の根に近い部分が含む単語が推薦キーワードとして、可視化モジュール (TextDisplay) 上に表示される。

この実践例における TETDM の特徴は、複数の処理を並列に並べたパネルに割り当て、複数の可視化モジュールを利用して、複数の処理をインタラクティブに行える環境を、低コストで実現できる点が挙げられる。特にコスト削減に関しては、環境の枠組みが利用できる点と、既存のモジュールが再利用できる点が挙げられる (詳細は次節で後述する)。この後者について、既存モジュールが汎用性を十分に考慮した設計となっていないなどの場合、必要なデータを得るための処理や機能を追加したり、データ型を合わせるなどの手間がかかる可能性がある。そこで今後は、各モジュールの再利用性を高めるために、機能を限定した小さなモジュールの開発を推奨することや、データ型を自動的に変換して柔軟な接続が可能になる仕様を追加していく。

3.4 利用者と開発者のスキルアップにつながる実践

本節では、TETDM を用いて、利用者と開発者のニーズとシーズのマッチング [砂山 13] を実践した例について述べる。現在、ある病院と TETDM の開発メンバーとの間で、看護記録 (電子カルテ) の監査に関する共同研究が行われている。

病院側では、看護師不足と言われる今日、看護師の技能向上を急がなければならないと言われている。そのため、新人や若手看護師が、手本となるような記録と自己

の記録を比較すること、また対象となる疾患や病棟などを配慮した上で、記録の自己評価を行うことができれば、各自の技能を把握し、記録を含めた看護技能をより向上させられる、といったニーズが挙げられている。そこで開発側としては、看護記録の比較評価を行うための単語比較支援ツール、また医療現場に特有の専門用語を利用するための専門用語抽出支援ツールの2つをシリーズとして提供する開発を行った。開発には、テキストデータとして33名の看護師の平成19年4月1日から平成20年3月31日までのデータを利用した。

開発は、図5のD-4グループに属する研究者がリーダーとなり、D-1またはD-2に属する学生等と共同で開発を行っている。本節で述べる2つのツールのベースは、それぞれ異なる大学の学生が特別研究(卒業研究)として行った。作成されるツールの利用は、図4のU-1グループに属する人が行うことを想定している。利用側は、作成されるツールを利用してニーズを満たすとともに、ツールの利用によって、U-3グループに移行することが期待される。

これまでの利用者が、専門的なツールの開発を依頼する場合、基盤となるシステムがない状態では、一から大規模なシステムを設計する必要があり、利用者と開発者の双方において負担が大きかった。TETDMの利用を前提とした場合、利用者側では既存のモジュールのデモを見ることで、その利用や拡張、実現可能なシステムのイメージを作り、要求する仕様を策定しやすい。開発側も、必要な機能ごとにモジュールを作成すれば良いため、開発内容の切り分けや、機能の追加を行いやすく、一部の機能が完成した段階で、利用者に試用を依頼することも可能となっている。

§1 看護記録からの専門用語抽出支援ツールの開発

看護記録を対象としたテキストマイニングのために、看護記録内で用いられている専門用語の辞書構築作業を支援する目的でツールの開発を行った。開発した専門用語抽出支援ツールは、TETDMの2つのパネルを利用し、マイニング処理モジュールと可視化モジュールのペアとして、各パネルに「用語抽出モジュール+TermDisplay(図8右)」と「TermHighLight+Htmlテキスト^{*10}(図8左)」を割り当てて利用する。

右のパネル内では、用語抽出モジュールによる専門用語の抽出結果を、単語リストとして提示する。単語をダブルクリックすることで、赤(専門用語として適切)、灰色(不適切)、白(未選択)の順に切り替わる。代表的な専門用語抽出手法として、複合語へのスコア付けに基づく下記4種類の手法[湯本01]をそれぞれマイニングモジュールとして実装した^{*11}。

*10 表4の4のモジュールをそのまま流用している。

*11 医療分野では英単語、カタカナ語が専門用語として出現する割合が多いとの予備調査結果に基づき、アルファベット、カタカナの連続をひとまとまりとして取り出す処理も導入している。

図8 専門用語抽出支援ツール

図9 専門用語抽出の評価結果

- C-Value: 構成単名詞数、出現回数等に基づき複合名詞を評価する手法
- MC-Value: 単名詞の場合にスコアが0にならないようにC-Valueを修正した手法
- LR: 隣接する単名詞の頻度やバリエーションに基づき複合名詞を評価する手法
- FLR: 複合名詞が単独で出現した頻度を考慮するようにLRを修正した手法

左のパネルは、抽出された用語が看護記録中でどのように出現するかを確認するために用いられ、右のパネル内でクリックした単語が文中でハイライト表示される。

表7に、看護記録10件から各手法により抽出された上位5件の専門用語を示す。また、抽出結果について大学病院医師に評価してもらった結果に基づき求めたP-R曲線を図9に示す。これより、再現率が低い領域ではC-Valueが、高い領域ではLRの性能が良いことがわかる。

本実践におけるTETDMの特徴は、可視化モジュールが共通で、専門用語を抽出する処理モジュールだけが異なるシステムを実現でき、今後も新たな専門用語を抽出するモジュールを容易に追加できる点、またパネル数を自由に増やせるため、4種類の手法による専門用語を4つ

表 7 看護記録から抽出された専門用語の例

順位	C-Value	MC-Value	LR	FLR
1	潰瘍性大腸炎	潰瘍性大腸炎	性	炎症性腸疾患
2	炎症性腸疾患	炎症性腸疾患	性腸	潰瘍性大腸炎
3	内視鏡	腸	性大腸	内視鏡
4	潰瘍性大腸	内視鏡	腸	mg
5	性大腸炎	腸炎	潰瘍性	状態

図 10 新人とベテランの看護記録比較支援環境

のパネルに並べて出力することができる点が挙げられる。

§2 新人とベテランの看護記録比較支援ツールの開発

新人とベテランの看護記録の比較支援を目的としたツールの開発を行った。新人とベテランの看護記録に含まれている単語を比較することにより、それぞれの書き方の特徴をもとに、看護過程で取り上げられている単語や書き方の違いの分析を支援する。

開発した環境では5つのパネルを併用する(図10)。図中一番左のパネルで、新人とベテランの経験年数をチェックボックスで選んで看護記録を絞り込むことができ、図の中央、および右から2番目のパネルでは、それぞれ新人とベテランの看護記録の集合をクラスタリングして地図形式で表示する^{*12}。また一番右のパネルでは、地図中のマウスで選択されたカルテの詳細を表示する。そして、左から2番目のパネルで、看護記録に含まれる「新人に特有の単語」「ベテランに特有の単語」「同程度に使われている単語」を表示する。

医師の評価に先だって、20人の大学生に新人とベテランの看護記録の比較を行ってもらった実験を行った。被験者に、新人とベテランのカルテの違いを回答(複数回答可能)してもらったところ、大きく「患者の状態についての詳細さ(34件)」「処置、処方についての詳細さ(16件)」「記述の仕方の適切さ(20件)」の3つに分類することができ、医学の知識を持っていない学生でも比較を行うことができていた。今後、先の専門用語の抽出支援モジュールを用いて、専門用語の辞書を構築した後、実際の医師に使用してもらうことを想定している。

本実践におけるTETDMの特徴としては、モジュール間の連動機能により、看護記録の集合をチェックボックスで絞り込むことで、地図、および地図間で共通の単語と異なる単語を随時更新できることや、マウスで触れた

表 8 マイニングを行うプログラムの実装時に必要な処理(表中のType-A, Type-BはTETDMを用いる開発者を表す)

処理	Type-A	Type-B	No TETDM
ウインドウの生成	×	×	
テキストデータ生成	×	×	
マイニング処理の実装			
処理結果と可視化の接続			×
処理結果の可視化	×		

表 9 TETDM 統合環境の主要処理のプログラムの行数

処理	行数
TETDM 統合環境合計	10050
1. ウインドウやパネルの生成と制御	2000
2. テキストデータの生成とデータ保持	2000
3. モジュールの動作の制御	4000
4. モジュール間の連動処理の制御	650
5. 各種設定変更と設定データの保持	1400
マイニング処理モジュール中央値	217
100 行未満のモジュール数:	4
200 行未満のモジュール数:	7
300 行未満のモジュール数:	9
300 行以上のモジュール数:	5
可視化インタフェースモジュール中央値	209
100 行未満のモジュール数:	3
200 行未満のモジュール数:	9
300 行未満のモジュール数:	4
300 行以上のモジュール数:	11

看護記録の内容を右端のパネルに出力することが可能になっており、利用者が対話的にシステムを利用できる点が挙げられる。特に、地図を表示するモジュールは既存システムを再利用しており、他人が作成したシステムとの連動を実現できる点が新しい。

3.5 スキルアップに伴うモジュール開発時のコスト

本節では、スキルアップに伴うモジュール開発時のコストについて述べる。

3.2.2 節で述べた、開発目的ではなくマイニングアルゴリズムの学習目的でTETDMを利用する人(表8のType-A)の場合、マイニング処理と処理結果を渡す部分のプログラム実装を行う必要があるが、表9のマイニング処理モジュールの中央値よりは少ない200行未満のプログラムの記述ですむ。しかしTETDMを利用せずに同様のことを行う場合、ウインドウの生成や結果のグラフィカルな可視化を行わない場合でも、テキストの入力、形態素解析の実行、データ構造へのデータ保存などのプログラム実装が必要となり、最大2000行の記述が必要になる。またTETDMを利用する場合に必要な、処理結果と可視化モジュールの接続プログラムについては、ライブラリ関数のように可視化モジュールが必要とするデータ型を調べた後、必要なデータを専用のメソッドを用いて渡すことができ、数行の記述ですむために大きなコストはかからない。

3.3.1 節や3.3.2 節で述べた開発者(表8のType-B)の

*12 既存システム[砂山12]をモジュール化して利用している

場合、マイニング処理と可視化インタフェースのモジュール1組で、合計600行未満のプログラムを実装することが想定される。しかしTETDMを利用しない場合、モジュールの実装に加えて、表9の1,2,3に相当する最大8000行の統合環境部分の実装が必要となる。またモジュールの実装に際しても、既存のモジュールを再利用できる場合には実装の必要がなく、既存のモジュールをそのまま利用できない場合でも、類似するモジュールのソースコードを参照し、ベースとすることで、開発のコストを軽減できる。

また複数のモジュール間で共通の観点に基づいて協調的な再処理動作を行うなど、モジュール間の連動処理は、表9の4および、1,2,3内の関連部分に記述されている。TETDMの連動処理は、複数の人が独立に作成したモジュール同士を連動させることを想定している。これは個人レベルで、手間や時間をかけて達成させることは現実的ではない。すなわちTETDMを利用しない場合、個別に連動させたいプログラムを入手あるいは実装した上で、それらの連動処理を実装する必要がある。他人のプログラムの解読や実装に大きな時間を割く必要がある。TETDMでは、TETDMサイトを通じて複数の人のモジュールを利用できる事が最大のコスト削減になり、協調的な再処理動作を行うためには、連動させたい変数を指定し、連動のためのメソッドを呼び出すのみで実現できる[砂山13]。

4. 利用者、開発者のスキルアップに向けた既存研究に対するTETDMの方向性

本章では、テキストマイニングのソフトウェアの利用者、開発者のスキルアップにつながる環境についての既存研究を示すとともに、既存研究が取り上げている問題点に対して、TETDMが対処する方針について述べる。

4.1 利用者のスキルアップにつながる環境

利用者のスキルアップにつながる環境、すなわち図4の横軸について、テキストマイニングの活用力を高める既存研究について述べ、取り上げられている問題点に対してTETDMが対処する方針を述べる。

§1 モジュール選択支援

インターネットなどで公開されているテキストマイニングのソフトウェアの多くは汎用性が高く、利用者の目的に合わせてソフトウェアを他のソフトウェアと組み合わせる必要がある[三川07]。目的に合わせてソフトウェアを使用するスキルをアップするためには、過去に類似する目的でソフトウェアを利用した人が、どのような方法で用いたかを知ることが有効と考えられる。現時点では、TETDMのサイト内で活用事例の公開を始めており、それら事例を参考にすることで、モジュールの組合せを選ぶスキルの向上が期待できる。また、3.2.1節

で述べた実践例の内容を発展させ、目的に応じた出力を得るためのモジュール選択を支援する。

§2 テキストマイニング活用力の獲得

教育の分野においては、テキストマイニングのソフトウェアを用いて学生の学習効果の向上を狙うもの[Dickey 05, Romero 08]や、教師の指導能力の向上を狙うもの[森下11]などがある。学習効果や指導能力の向上を図るには、データを元に現在の状態を把握するスキルが必要となる。TETDMの統合環境においては、複数のパネルを用いてデータを並列に並べることができるとともに、あるパネル内で着目したデータの他のパネル内での位置づけを可視化できるため、データ分析の際の着眼点を養う練習ができると考えられる。テキストマイニングの活用力の獲得については、3.2.2節で述べた学部生にテキストマイニングを理解してもらうセミナーなどの実施も有効と考えられる。

§3 テキストマイニングツールを用いた発想支援

マーケティングの分野においては、テキストマイニングのソフトウェアが、ブランド戦略に適用すること[豊田03]、顧客の購買パターンを分析すること[加藤02]、技術の特許テキストを分析して経営戦略を立てること[菰田11]、販売戦略を立てること[Ghani 06]などに利用されている。このような販売戦略の策定においては、データをあらゆる視点から分析することや、一般的でないツールの使用などにより新たな視点を発見することが望まれる。TETDMにおいては、複数のパネルでデータを多角的に分析することを意識した構成になっていることや、最新の研究や評価が確立されていない斬新な研究についても、自由にツールとして追加できるため、利用者のスキルアップにも役立てられる。

TETDMを知識創発のための基盤環境として活用するため、近未来チャレンジの残りの期間で、様々な視点からのデータ分析を可能とする枠組みを構築する。TETDMの知識創発に向けた活用では、以下の2点を軸とする。

1. モジュールの柔軟かつ多様な組合せに基づいた、斬新な結果の取得による発想
2. 異なるデータによる処理結果の比較を支援するツールの開発による発想

1. については、モジュールの開発者が想定するモジュールの組合せに留まるのではなく、多様なモジュールを自由に組み合わせられるようにすることで、開発者が想定していなかったモジュールの組合せにより、新しい結果を得られる余地を作る。また多様なモジュールを集める上では、信頼性が高いモジュールのみならず、面白さや斬新さを重視したモジュールも積極的に集める事で、組合せの自由度を増したいと考えている。

また2. について、データ分析においては、データの量が増すにつれ、全データの解釈だけではなく、データのある観点に基づいて絞り込んだ上での解釈が必要かつ重要となってくる。例えば、あるテキストの前半と後半の違

いを見ることや、3.4.2 節で述べた新人とベテランの看護記録を比較する場合などが挙げられる。そのため、複数の部分テキストデータを保持する機能、ならびにそれらの部分テキストに対する処理結果を比較できる機能を実装する事で、発想を支援したいと考えている。

4.2 開発者のスキルアップにつながる環境

開発者のスキルアップにつながる環境、すなわち図 5 の縦軸のプログラミング力を高める既存研究、ならびに横軸の開発力を高める既存研究について述べ、取り上げられている問題点に対して TETDM が対処する方針を述べる。

§1 プログラミング力の獲得

プログラミングを学び始めた人が大学の講義や演習でプログラミングを学習していく時、課題をこなせず途中でドロップアウトしてしまう人も少なくない。教授する方法と学生のプログラミングパフォーマンスとの間には関連があり [Mandinach 87]、ドロップアウトしそうな学生を早期に発見する方法が提案されている [長谷川 11]。たとえば、学生が講義資料を閲覧しているログを収集し、閲覧のパターンから進捗が遅れがちな学生を発見する方法 [堀口 12] や、e-ラーニングの学習履歴から発見する方法 [高岡 11]、行き詰まる可能性が高い学生を学習状況から推定し、かつ理解状況が似ている学生をまとめることで、指導の効率を上げる方法 [田口 09] が提案されている。TETDM 統合環境では、プログラミングを初めて学んでいる最中の学生、すなわち U-0 グループの学生が自力で学習することには向いていない。しかし、プログラムを教える側が TETDM で学生の学習に使用するモジュールを作成し、そのモジュールの一部を学生に書き換えさせる課題を与えることにより、モジュールを作成する敷居を下げると同時に、学生のモジュール作成の達成感を与えることは可能と考えられる。プログラミングの初心者に対しては、いかに飽きさせずに学習を持続させるかが肝要と考えられ、学んだ少しのことで大きな結果が得られることは学習の動機づけになると期待できる。今回、プログラミングに特化した実践は行っていないが、今後上記の活用を可能にする枠組みを検討していきたい。

§2 研究、開発力の獲得

より複雑なプログラムを組めるようになるには、プログラムの設計を自ら行い、アルゴリズムを考えるスキルを鍛える必要がある。ここでオーソドックスなプログラムを組む問題を繰り返し解いてはアルゴリズムを考えるスキルは身に付かない [Merrienboer 92]。また研究と開発のスキルにも関連して、研究の進捗管理を行うシステム [真原 10] や研究活動のノウハウを共有するシステム [渡邊 09]、中には研究テーマの決定を支援するもの [岸川 11] もある。TETDM では不特定多数の現場の利用者が見込まれるため、それら実践的なニーズに対応できるモジュールを考えていくことや、既存のモジュール

といかに組み合わせて新しい環境を作り上げるかなど、考えることを刺激できるようになっている。これにより、考えることが必要なアルゴリズムの作成スキル、ならびに研究、開発能力を鍛えていくことが可能と考えられる。今回、3.3.1 節や 3.3.2 節で実践した内容を踏まえ、今後さらに活用の場を広げていきたいと考えている。

4.3 利用者と開発者の両者のスキルアップにつながる環境

利用者と開発者の両者のスキルアップにつながる環境、すなわち、シーズとニーズのマッチングに関する既存研究について述べ、TETDM の今後の方向性について述べる。

§1 シーズとニーズのマッチング

利用者は欲しいモジュールの情報をニーズとして開発者に提供し、開発者は利用者が求める可能性があるモジュールを開発しシーズとして提供する。これら利用者と開発者は独立に活動するのではなく、両者が共同して開発することが望ましい [Hippel 86]。開発者は利用者が求めるモジュールについて、データの形式や必要な機能について深く理解する必要があるが、多くの場合、互いに持つ事前知識が異なるため、モジュールを造り上げていく上でのコミュニケーションに齟齬が生じる。特に大きなモジュールの開発においては、上流工程になるにつれてコミュニケーションを円滑にとることが難しくなる [松澤 08]。またアジャイル開発のように機能を細かく分割して、テストを繰り返し替える中でソフトウェアを造り上げることも、コミュニケーションの齟齬をおさえることに繋がる [Highsmith 01]。そこで、TETDM という一つのフォーマットを前提にするとともに、個別に動作可能な小さなモジュールを組み合わせて大きなモジュールを開発していくことにより、情報の行き違いを避け、お互いが納得しながら開発を進められる可能性が高められる。現在、3.4 節で述べた内容は継続中であり、今後のモデルケースになると考えられるため、現場での活用が可能なレベルまで研究を進めていきたい。

§2 利用者と開発者のコミュニケーション

モジュール開発のためのコミュニケーションを繰り返すことで両者のモジュール開発のスキルがアップできる [肥後 07] と言われている。継続的な開発者と利用者のコミュニケーションについて、TETDM では、TETDM サイト内に、開発者側からのシーズとなるモジュール一覧に加え、利用者側からのニーズ一覧を表示するページを生成した上で、シーズとニーズの双方を評価する機能を設けることで、お互いが質の高いシーズとニーズを提供できるように心掛けられる環境を構築する予定となっている。シーズとニーズの一覧表示、および評価機能は、TETDM 統合環境が完成し、多くの利用者と開発者が見込まれる状況の中で、構築と改善を行う必要があるため、逐次検討を進め、適切なタイミングでの活用を見込んでいる。

5. 結 論

TETDM では、複数のテキストマイニング、ならびにその周辺技術を柔軟に組み合わせる環境を構築し、それらを広く提供することを目指している。本環境により、複数の技術を用いたい利用者の環境が整えられ、ニーズに応じたモジュールを選択し、集中して作業を行えること、また開発者による多くの研究が認知、実用化されていくことを期待している。

TETDM の利用が、単なるツールとしての利用、研究内容の公開のみに留まるのではなく、この統合環境を利用し続けることで、利用者側、開発者側の双方のスキルアップが図れる環境となることを望んでおり、そのための枠組み作りも継続して進めていきたいと考えている。

今後は、図 1 のスケジュールに対応して、「開発者支援」と「知識創発に向けた利用者支援」を行っていく。開発者支援については、チャレンジの目標 1 と関連して多くの開発者の参入を促すことを目的として、3.3 節の実践と 4.2 節の今後の方向性で述べた内容の中でも、特にモジュールの実装方法の曖昧さの解消と実装時のコスト削減を目指す。知識創発に向けた利用者支援については、チャレンジの目標 3 と関連して知識創発の基盤環境を構築することを目的として、3.2 節の実践と 4.1 節の今後の方向性で述べた内容の中でも、特に複数のツールを連携させながら、複数の部分データ間の結果の比較を行う中で、知識創発に結びつくプロセスの明確化と積極的な思考支援を目指す。

◇ 参 考 文 献 ◇

- [Dickey 05] D. A. Dickey: Teaching Data Mining in a University Environment: Proceedings of the SUGI 30 Conference, pp.1-10, (2005).
- [Ghani 06] R. Ghani, K. Probst, Y. Liu, M. Krema, and A. Fano: Text Mining for Product Attribute Extraction, ACM SIGKDD Explorations Newsletter, Vol.8, No.1, pp.41-48, (2006).
- [Hall 09] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1, (2009).
- [長谷川 11] 長谷川伸, 松田承一, 高野辰之, 宮川治: プログラミング入門教育を対象としたリアルタイム授業支援システム, 情報処理学会論文誌, Vol.52, No.12, pp.3135-3149, (2011).
- [Highsmith 01] J. Highsmith and A. Cockburn: Agile Software Development: the Business of Innovation, Computer, Vol.34, No.9, pp.120-127, (2001).
- [肥後 07] 肥後芳樹, 吉田則裕, 楠本真二, 井上克郎: 産学連携に基づいたコードクローン可視化手法の改良と実装, 情報処理学会論文誌, Vol.48, No.2, pp.811-822, (2007).
- [Hippel 86] E. von Hippel: Lead Users: A Source of Novel Product Concepts, Journal Management Science, Vol.32 No.7, pp.791-805, (1986).
- [堀口 12] 堀口悟史, 井垣宏, 井上亮文, 山田誠, 星徹, 岡田謙一: 講義資料閲覧ログを用いたプログラミング講義進捗管理手法の提案, 情報処理学会論文誌, Vol.53, No.1, pp.61-71, (2012).
- [加藤 02] 加藤直樹, 羽室行信, 矢田勝俊: マーケティングとデータマイニング, システム制御情報学会誌, Vol.46, No.4, pp.190-196, (2002).
- [岸川 11] 岸川善紀, 嶋崎真仁: 経営工学分野における卒業研究管理手法について, 電子情報通信学会技術研究報告 (教育工学), Vol.111, No.213, pp.1-6, (2011).
- [菰田 11] 菰田文男: 「単語セット」の作成と進化に基づくテキストマイニング手法 MOT (技術経営) のためのテキストデータ解析を事例として, 情報管理, Vol. 54, No. 9, pp.568-578, (2011).
- [真原 10] 真原友春, 田隈広紀, 西尾雅年: ロジックモデルとバランススコアカードを活用した卒業研究支援, 国際プロジェクト・プログラムマネジメント学会誌, Vol.5, No.1, pp.161-170, (2010).
- [Mandinach 87] E. B. Mandinach and M. C. Linn: Cognitive Consequences of Programming: Achievements of Experienced and Talented Programmers, Journal of Educational Computing Research, Vol.3, No.1, pp.53-72, (1987).
- [松澤 08] 松澤芳昭, 杉浦学, 大岩元: 産学協同の PBL における顧客と開発者の協創環境の構築と人材育成効果, 情報処理学会論文誌, Vol.49, No.2, pp.944-957, (2008).
- [Merrienboer 92] J. J. G. Van Merrienboer and M. B. M. De Croock: Strategies for Computer-Based Programming Instruction: Program Completion Vs. Program Generation, Journal of Educational Computing Research, Vol. 8, No. 3, pp.365-394, (1992).
- [三川 07] 三川健太, 高橋勉, 後藤正幸: テキストデータに基づく顧客ロイヤルティの構造分析手法に関する一考察, 日本経営工学会論文誌, Vol.58, No.3, pp.182-192, (2007).
- [森下 11] 森下孟, 新村正明: 大学間遠隔講義システムを活用した複数の授業担当教員らによる遠隔授業実施の評価, 日本教育工学会論文誌, Vol.35, pp.65-68, (2011).
- [Plaisant 08] C. Plaisant, J. D. Fekete, and G. Grinstein: Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository, IEEE Trans. on Visualization and Computer Graphics, Vol. 14, No.1, pp.120 - 134, (2008).
- [Romero 08] C. Romero, S. Ventura, and E. Garcia: Data Mining in Course Management Systems: Moodle Case Study and Tutorial, Computers & Education, Vol.51, No.1, pp.368-384, (2008).
- [砂山 11] 砂山渡, 高間康史, Danushka Bollegala, 西原陽子, 徳永秀和, 串間宗夫, 松下光範: Total Environment for Text Data Mining, 人工知能学会論文誌, Vol.26, No.4, pp.483 - 493, (2011).
- [砂山 12] 砂山渡, 濱岡秀平, 奥田澄: 情報収集のためのテキストデータ集合の再帰的クラスタリング, 日本知能情報ファジィ学会誌, Vol.24, No.3, pp.697 - 706, (2012).
- [砂山 13] 砂山渡, 高間康史, 西原陽子, 徳永秀和, 串間宗夫, 阿部秀尚, 梶並知記: テキストデータマイニングのための統合環境 TETDM の開発, 人工知能学会論文誌, Vol.28, No.1, pp.1-12, (2013).
- [高岡 11] 高岡詠子, 大澤佑至, 吉田淳一: e-Learning 学習履歴を用いたドロップアウト兆候者早期抽出手法の提案, 検証および今後の可能性, 情報処理学会論文誌 Vol.52, No.12, pp.3080-3095, (2011).
- [田口 09] 田口浩, 原田史子, 高田秀志, 島川博光: プログラミング学習カルテの分析による人的教育資源の有効活用, 情報処理学会論文誌, Vol.50, No.10, pp.2409-2425, (2009).
- [豊田 03] 豊田裕貴, 森永聡: 企業におけるマーケティング分野でのテキスト活用事例: ブランド・イメージ調査へのテキストマイニング技術の適用, 情報処理, Vol.44, No.10, pp.1028-1031, (2003).
- [渡邊 09] 渡邊岳志, 服部哲, 速水治夫: 研究活動向けナレッジ共有システムの提案, 情報処理学会研究報告 (グループウェアとネットワークサービス), Vol.2009, No.33, pp.127-132, (2009).
- [湯本 01] 湯本紘彰, 森辰則, 中川裕志: 出現頻度と連接頻度に基づく専門用語抽出, 情報処理学会研究報告 (自然言語処理), Vol.2001, No.86, pp.111-118, (2001).

〔担当委員: 阿部 明典〕

2013 年 2 月 8 日 受理

 著 者 紹 介

砂山 渡(正会員)

1995 年大阪大学基礎工学部制御工学科卒業。1997 年同大大学院博士前期課程修了。1999 年同大大学院博士後期課程中退。同年同大学院助手, 2003 年広島市立大学助教授, 2007 年同准教授, 現在に至る。博士(工学)。人間の創造活動を支援する研究に興味を持つ。電子情報通信学会, 日本知能情報ファジィ学会, 言語処理学会, IEEE, 各会員。

高間 康史(正会員)

1994 年 東京大学工学部電子工学科卒業。1999 年同大学院博士課程修了。1999-2002 東京工業大学大学院総合理工学研究科助手, 2002-2005 東京都立科学技術大学助教授, 2005 年より首都大学東京システムデザイン学部准教授。博士(工学)。Web Intelligence や情報可視化, 知的インタフェースの研究に従事。主要著書は「インテリジェントネットワークシステム入門」(コロナ社)。IEEE, 日本知能情報ファジィ学会, 情報処理学会, 電子情報通信学会各会員。

西原 陽子(正会員)

2003 年大阪大学基礎工学部卒業。2005 年同大大学院基礎工学研究科博士前期課程修了。2007 年同研究科博士後期課程修了。博士(工学)。日本学術振興会特別研究員を経て, 2008 年東京大学大学院工学系研究科助教, 2009 年同講師, 2012 年立命館大学情報理工学部准教授, 現在に至る。コミュニケーション支援, 対話デザインに興味を持つ。情報処理学会, 医療情報学会, 機械学会, 各会員。

梶並 知記(正会員)

2004 年 東京都立科学技術大学工学部電子システム工学科卒業。2010 年 首都大学東京大学院システムデザイン研究科博士後期課程修了。首都大学東京システムデザイン学部特任研究員, 東京工科大学コンピュータサイエンス学部助教を経て, 2012 年より神奈川工科大学情報学部助教。博士(工学)。情報可視化, 意思決定支援, インタラクティブシステムの研究に従事。ACM, 日本知能情報ファジィ学会, 情報処理学会, 電子情報通信学会, 情報知識学会, 日本デ

ジタルゲーム学会各会員。

串間 宗夫(正会員)

1987 年宮崎大学大学院工学研究科修士課程修了。2003 年同大大学院工学研究科博士後期課程修了。2012 年同大大学院医学系研究科博士課程医学専攻修了。宮崎大学医学部附属病院研究員, 地方公務員。博士(医学)。博士(工学)。医学系では, 癌治療, 診療情報, 電子カルテ, 地域医療連携, 工学系では, MOS アナログ集積回路, 多値論理回路, 教育工学, に興味をもつ。日本医療情報学会, 電子情報通信学会, 多値論理研究会, バイオメディカル・ファジィ・シ

ステム学会, 各会員。

徳永 秀和(正会員)

1986 年東京工業大学大学院総合理工学研究科修士課程修了。同年新日本製鐵(株)勤務。1993 年高松工業高等専門学校講師。2005 年同校助教授, 2007 年同准教授, 2009 香川高等専門学校准教授, 現在に至る。博士(工学)。Web からの知識獲得, 知識共有の研究に興味を持つ。情報処理学会, 日本知能情報ファジィ学会, 各会員。