

## TETDM を利用した情報系専門教育の実践例

## Practice of Information Education using TETDM

梶並知記<sup>\*1</sup>

Tomoki Kajinami

<sup>\*1</sup>神奈川工科大学情報学部

Faculty of Information Technology, Kanagawa Institute of Technology

This paper introduces an example of practice which combines learning with practice for supporting specialized subject utilizing TETDM (Total Environment for Text Data Mining). Moreover, this paper discusses a method for expanding TETDM's community from a perspective of gamification. TETDM has a feature which the system is constructed by mining tools and visualizing tools. Common use of TETDM is text data mining for data analyzers. Application of the system to support that students learn specialized subjects in the field of education, is not considered. This paper introduces a seminar on information engineering utilizing system's feature and discusses elements required for expanding use of TETDM.

## 1. はじめに

本稿では、TETDM (Total Environment for Text Data Mining) を用いた、情報系専門教育の実践例を紹介するとともに、TETDM のコミュニティ発展へ向けた方法論について検討する。

TETDM は、Java で構築された、統合的にテキストデータマイニングを行うための環境である [砂山 13]。基本的には、私用、業務利用問わず、テキストデータマイニングに従事するユーザを対象にしている。

本稿では、TETDM のもつ特徴と情報系専門教育現場との親和性が高いと考え、著者が実践した TETDM を利用したセミナーを紹介する。本稿は、実践例の紹介と議論をとおし、教育現場での TETDM 利用の可能性を示し、また TETDM コミュニティ発展のためのヒントとなることを期待するものである。

## 2. TETDM の特徴

ここでは、TETDM を専門教育で利用する際に有効と考える、TETDM のもつ特徴について述べる。TETDM は、テキストデータマイニングのための統合環境であり、Java で構築されている [砂山 13]。本稿で関連する特徴は、以下の 3 つである。

特徴 1 プリミティブなテキスト情報を格納する変数の存在

特徴 2 役割に応じたツールの分離

特徴 3 マイニング処理ツール同士の連携

特徴 1 は、TETDM へ入力されたテキストの、単語数、文数、段落数、文同士の類似度などを格納した変数が用意されていることである。ツール開発者が、自分でテキストの基本的な情報を算出するコードを記述する必要はない。特徴 2 は、テキスト処理を行うマイニング処理ツールと、マイニング結果を可視化する可視化ツールに、ツールの役割を分離していること

連絡先: 梶並知記, 神奈川工科大学情報学部, 〒 243-0292 神奈川県厚木市下荻野 1030, 046-291-3235, kajinami@ic.kanagawa-it.ac.jp

である。テキストデータマイニングを行う分析者にとっては、マイニング処理ツールと可視化ツールの組み合わせを、タスクに応じて選択できる利点がある。ツールの開発者にとっては、個人の興味や得意/苦手に応じて、作成するツールをマイニング処理ツールと可視化ツールのどちらか一方に絞ることができる利点がある。特徴 3 は、あるマイニング処理ツールが、他のマイニング処理ツールによって出力されるデータを利用して、マイニング処理を行うことである。ツール開発者にとっては、必要に応じて他のマイニング処理ツールの出力結果を利用できるので、実装の負荷が減少できる利点がある。

## 3. 授業デザイン

本節では、授業の目的や想定している受講者を述べたあと、授業のデザインの概要を、2. 節で述べた TETDM の特徴と絡めて述べる。

## 3.1 授業目的と想定受講者

本稿で紹介する専門教育の形態は、情報系学部の 3 年次前期/後期に配置されているセミナーである。受講者は、自分の興味と担当教員の専門分野や興味などを考慮し、受講するセミナーを選択する。大雑把な授業目的は、4 年次で行う卒業研究を視野に入れて、受講者に専門分野について深く学んでもらうことであるが、詳細や具体的な目的は、担当教員によって異なる。

本稿で紹介する授業の目的は、以下の (1) ~ (3) のとおりである。(1) と (2) は受講者に対する教育上の目的だが、(3) に関しては、TETDM コミュニティ発展のための方法論を構築するヒントとするための、著者 (紹介する事例における教授者でもある) にとっての目的である。

目的 1 テキストデータマイニングに関連する理論の理解

目的 2 利用目的志向のソフトウェア作成の経験

目的 3 TETDM コミュニティ参加への動機付け方法の検討

受講者は、2 年次までに Java 言語の基本<sup>\*1</sup>を履修済みである。一方、テキストデータマイニングに関する知識は、テキスト

<sup>\*1</sup> クラスとインスタンスの関係、インスタンス変数/メソッドへのアクセス、継承など。

表 1: 授業の教育上の目的と関連する TETDM の特徴

授業の目的 (3.1 節)	関連する TETDM の特徴 (2. 節)
テキストデータマイニングに関連する理論の理解	・プリミティブなテキスト情報を格納する変数の存在 ・役割に応じたツールの分離
利用目的志向のソフトウェア作成の経験	・役割に応じたツールの分離 ・マイニング処理ツール同士の連携

トマイニングやデータマイニングといった用語を聞いたことがあったり、なんとなくテキスト/データを分析する技術と理解していたりする程度である<sup>\*2</sup>。

### 3.2 授業のデザインの概要

表 1 は、3.1 節で述べた専門教育上の目的 (1)(2) と、2. 節で述べた TETDM の特徴を対応付けたものである。

本稿では、3.1 節で述べた目的 1、テキストデータマイニングに関連する理論の理解を促すために、受講者が座学で理論を学習するだけでなく、プログラミングを行う実験課題を通して、体験的に学ぶことが重要と考えている。プリミティブなテキスト情報を扱いやすく、またプログラミングに習熟していない受講者にとって比較の実装が困難な可視化に既存ツールの利用で対応可能な TETDM を、実験ツールとして利用する。

本稿では、目的 2、目的志向のソフトウェア作成を経験させるために、比較の実用的なタスクを想定したマイニング/可視化ソフトウェアを作成してもらうことが重要と考えている。役割に応じてツールが分離しており、マイニングと可視化の両方を受講者自身が実装する必要がなく、また他のツールから出力されるデータを利用することで実装の負担を減らすことができる TETDM を、開発環境として利用する。

## 4. 授業の実践

本節では、15 回のセミナー形式授業 (受講者 4 名) の前半回を導入部、後半回を発展部の 2 つに分けて実施した授業から、TETDM 利用例を紹介する。導入部は、目的 1 に対応し、受講者にテキストデータマイニングに関連する理論を理解してもらう内容である。発展部は、目的 2 に対応し、受講者に目的志向のソフトウェア作成を経験してもらう内容である。4.1 節で、導入部の事例を、4.2 節で、発展部の事例を紹介する。なお、利用した TETDM のバージョンは、0.42 以前のものである。

### 4.1 事例：文書ベクトルと類似度の理解

授業で採用した話題の 1 つは、情報検索に関連する内容である。具体的には、教授者が文章をベクトルで表現することの意味を図表を用いて受講者に示し、受講者に簡単な文章の cos 類似度を手計算で行ってもらい、その後マイニング処理ツールを実装、理論通りの類似度となっているか確認してもらう、理工学系の実験に類する流れである。教授者が例を示す、受講者が手計算する、その後プログラムを組んで計算結果を確認する、といったサイクルを繰り返す。授業時間 (1 回 90 分) の関係上、このサイクルを実施する際に重要視すべきことは、実装の容易さと、受講者に対するフィードバックの早さである。

受講者 (お手本や例を見せる教授者も) TETDM であらかじめ用意されている cos 類似度の変数を利用し、マイニング処理ツールを作成する。実際には、1 からプログラミングする

表 2: 紹介する授業における課題・報酬・交流

課題	レーダーチャートに適するマイニング処理ツール作成 ・文章のどのような特徴を可視化要素とするか考える。 ・要素の属性値の算出手法を考える。
報酬	グレードの高い成績と Java ソフトウェア開発経験 ・4 年次の卒業研究に向けて実践的なプログラミングを経験する。 ・就職活動に役立つ (?) ツール公開のチャンスがある。
交流	受講者同士のアイデア交換 ・ツールを紹介する。 ・新しいツールのアイデアを得る。

のではなく、ひな形として用意されているマイニング処理ツールを改造する形になる。結果の表示は、既存の可視化ツールを利用する。これにより、受講者が「学んだ理論をすぐに試せる (= 自分の理解が正しいか確認できる)」ようになった<sup>\*3</sup>。

本事例では、受講者の学習の支援方法としてどこまで有効か、即ち、どれだけ受講者の理解を助けたかについて定量的な評価は行っていない。しかしながら、実験や演習といった時間を別途とらない座学中心の情報系専門科目において、受講者に体験的な学習環境を提供する 1 つの方法として、TETDM の活用は価値あることではないかと考える。

### 4.2 事例：マイニング処理ツールの作成

受講者の授業参加意欲の向上と、TETDM コミュニティ発展のための方策検討も考慮し、ゲーミフィケーションにおいて重要な課題・報酬・交流のトライアングル [岡村 12] を意識した授業デザインとした。

表 2 は、課題・報酬・交流の内容をまとめたものである。課題として、特定のツールと連携させることを想定したツールの作成タスクを受講者に課した。具体的には、教授者は可視化ツール「レーダーチャート (RadarChart)<sup>\*4</sup>」を受講者へ提示し、受講者はレーダーチャートで可視化するのにふさわしいと考えるデータを出力するマイニングツールを作成する。漠然と、マイニング処理ツールを作成せよとの課題ではなく、組み合わせる可視化ツールを明確にし、マイニング処理ツール作成の方向性を定めやすくした。受講者のタスクを細分化すると、レーダーチャートの軸に対応させる要素と、その要素の属性値の算出方法を考えることとなる。単にツールの作成を求めただけでは、今一つ受講者のツール作成意欲がわからない可能性があるため、報酬として、ツールの出来次第で優良な成績や自作プログラムの公開経験を得ることができることを示した<sup>\*5</sup>。交流は、受講者同士 (または教授者と受講者) でアイデアを比較したり、ツールの改造例を考えるなどで、競い合うものではなく協力的な対話である。

図 1 は、ある受講者が作成したマイニング処理ツール「主題語含有率」によりサンプルテキストである浦島太郎の物語を分析した結果を、レーダーチャートで可視化した例である。主題語含有率ツールは、入力テキストに含まれる主題語が、入力テキストの全段落のうちどれだけの数の段落に含まれているかや、入力テキストの全文章のうちどれだけの数の文章に含まれているかなどを算出する。さらに、主題語が主語として現

\*3 受講者の理解をより促すために、ベクトルで表現する文章の内容、類似度を算出する文章の内容などを受講者の興味に合わせたものにするといった工夫も、合わせて行っている。

\*4 現在、TETDM サイト <http://tetdm.jp/pukiwiki/index.php?TETDM> から配布されている。

\*5 後日、教授者は、ある受講者から、就職活動の面接で TETDM のツール作成の話題を出したら面接官に興味をもってもらえた、といった話を聞いた。

\*2 著者の担当するセミナーの受講者が偶然もっていた特徴である。3 年次生全体に共通しているわけではない。

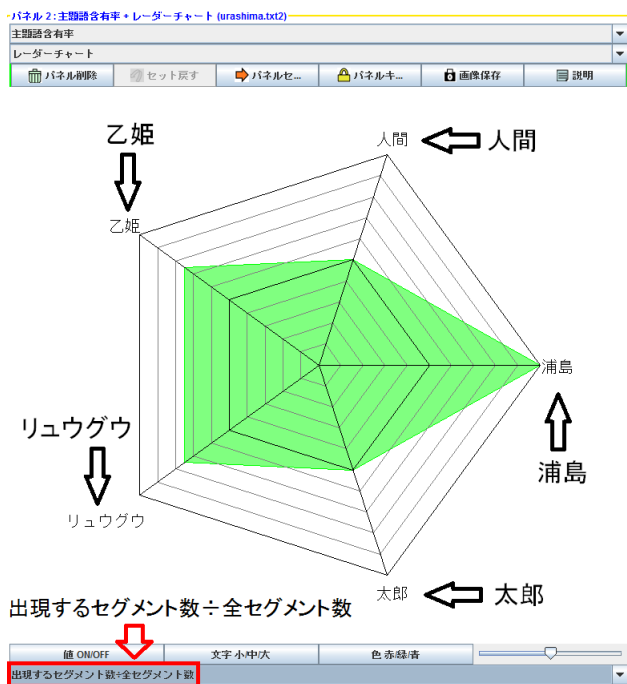


図 1: 主題語含有率をレーダーチャートで表示した例

れている文章の数が、全文のうちのどれだけあるかなども算出する。主題語については既存ツールによって抽出できるため、受講者はツール同士の連携を行った。段落数や文章数などは、TETDM に用意されているプリミティブなテキスト情報を格納する変数を利用している。図 1 では、主題語含有率を、主題語が出現するセグメント数を全セグメント数で割って求めており、主題語が「人間」「浦島」「太郎」「リュウグウ」「乙姫」の 5 つある（レーダーチャートの軸に対応）ことを示している。なお、主題語含有率ツールは、TETDM サイトで公開中である。

また、2013 年 4 月 12 日時点では公開に至っていないが、ある受講者は、TETDM に用意されている、主語を取り扱う変数に着目し、受講者自身の考える日本語らしさ（文章における主語の省略）を考慮したマイニング処理ツールを考えた。このアイデアは、前述した交流の段階で出てきたものである。

## 5. コミュニティ発展の方法論の検討

本稿で述べたセミナー形式授業における発展部で採用した、課題・報酬・交流のトライアングルは、TETDM コミュニティ全体を発展させるための方法を検討するヒントになる。

TETDM に関係するユーザは、私用/業務利用問わずテキストデータマイニングを行いツール作成を行わない「利用者」と、ツールの作成も行う「開発者」の 2 つに大別できる。利用者側と開発者側の双方に、課題・報酬・交流のトライアングルが形成されるような仕組みを整えると良いのではないかと考える。課題・報酬・交流が相互作用することで、コミュニティの活動が維持・発展すると考える。

表 3 は、利用者にとっての課題・報酬・交流をまとめたものである。利用者にとっての課題は、開発者へ課題を出すことであり、作成して欲しいツールの概要を伝えたり、既存ツールの問題点を報告することである。報酬として、新規ツールを早めに使えたり、お試し版を利用できたりといった、優先的な

表 3: 利用者の課題・報酬・交流

課題	開発者の課題を提示 ・欲しいツールの要望をだす。 ・既存ツールの問題点を報告する。
報酬	優先的なサービス利用権と貢献度可視化 ・新規ツールを他ユーザより先にお試しできる。 ・課題提示回数や頻度、交流の度合いに応じた貢献度を公開する。 ・ツールを活用するスキルが上がる。 ・コミュニティ内での発言権が増す。
交流	利用者同士の情報交換 ・ツールの活用場面を公開する。 ・他の利用者のトラブルを解決する。

表 4: 開発者の課題・報酬・交流

課題	新規ツールの作成や既存ツールのメンテナンス ・自身が作成したい/利用者の要望に応じたツールを作成する。 ・既存ツールの不都合修正や改良を行う。
報酬	貢献度可視化と利用者からのフィードバック ・作成ツール数やツール作成頻度などに応じた貢献度を公開する。 ・利用者からのお礼メッセージなどを受け取れるようにする。 ・ツールを作成するスキルが上がる。 ・コミュニティ内での発言権が増す。
交流	開発者同士の情報交換 ・ツール作成時のトラブルを相談する。 ・他の開発者のトラブルを解決する。 ・新しいツールのアイデアを得る。

サービス利用権を与える。また、作成ツールの要望、問題点報告の回数や頻度に応じて TETDM コミュニティへの貢献度を測り、それを他のユーザに見える形で提示することも考える。交流は、主に利用者同士で、ツールの活用例に関する情報交換である。交流を通して、欲しいツールが閃き課題となることがあると想定する。交流に関する報酬は、TETDM の使用スキルが向上することによる、自身のタスクの効率化や、もっと便利なツールの概要を開発者へ提示できるようになることである。また、他の利用者のトラブル解決の回数や頻度に応じた、TETDM コミュニティへの貢献度上昇となる。多大な貢献度が可視化されることで、他のユーザから相談を持ちかけられたり、自身の影響力が増していったりすることも報酬といえる。

表 4 は、開発者にとっての課題・報酬・交流をまとめたものである。開発者にとっての課題は、新規ツールの開発や既存ツールの改良・修正である。報酬は、利用者の場合と似ており、TETDM コミュニティへの貢献度の可視化である。利用者の要望に応えた数や頻度に応じて貢献度を測ることを考える。なおここで、利用者から開発者へのフィードバック（便利なツールを作成した開発者へのお礼コメントなど）も取り入れ、コミュニティ貢献度上昇に影響を与えるものとする。交流は、開発者間で情報交換を行うことである。関連して、開発スキルの向上や、新規ツールのアイデアを得ることが報酬となる。利用者と同じく、コミュニティへの貢献度が大きくなるほど、発言権を得られるのも報酬である。

以上、利用者と開発者にかけて検討したが、利用者側と開発者のほかに TETDM のコアメンバグループを加えて、これら 3 つの立場のユーザを 3 つの頂点として描くトライアングルの上で、前述した課題・報酬・交流のトライアングルを構築するようなイメージの、コミュニティのあり方を今後検討したい。

## 6. おわりに

本稿では、TETDM を利用した情報系専門教育の実践例を紹介した。授業の導入部では、TETDM のもつプリミティブ

なテキスト処理用変数を活かし、受講者に、情報検索関連の技術を理論面だけでなく実験的な課題を通して体験的に理解してもらった。授業の発展部では、マイニング処理ツールと可視化ツールに分離している TETDM のツール構成と、ツール同士の連携機能を活かし、教授者が提示した可視化ツールに合わせたマイニング処理ツールを、受講者に作成してもらった。

本稿では、学習効果の検討、他の授業デザインとの比較などは行っていないため、TETDM を教育現場に用いることの有効性に関する議論の余地は多分に残されているが、今後、事例を蓄え考察を深めていきたい。

また、今回授業の発展部で取り入れた、課題・報酬・交流のトライアングルに基づき、TETDM コミュニティ発展のための方法論の検討を行った。今後、実現可能性も考慮しつつ、コミュニティのありかたを模索していきたい。

## 参考文献

- [岡村 12] 岡村健右：第 4 章 ゲームフィクションを社員教育に活かす、ゲームの力が会社を変える、日本実業出版社、pp.143-182 (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)