

負例に着目した嗜好解析手法の提案と そのフィードリーダーへの応用

井下雄樹^{†1} 藤田 聡^{†1}

本研究では、ユーザの web コンテンツに関する嗜好をタグを用いて明示的に示す手法の提案と、それに基づくコンテンツの選別と推薦を行うフィードリーダーを実現する。近年、インターネット上の多くのサイトでタグが用いられている。一般に、タグを用いたコンテンツの分類では同じタグの有無により判断されるが、これは必ずしも正しいとはいえない。本手法では特定のタグと異なる事を示すためのタグを導入することで、異なるコンテンツであることをユーザが明示的に指示することを可能とする。これにより、より鮮明にユーザの嗜好が解析可能となり、これを用いることで精度の高いコンテンツの選別と推薦を実現する。

A New Approach to the Analysis of Users' Preference and Its Application to Feed Readers

YUUKI INOSHITA^{†1} and SATOSHI FUJITA^{†1}

In this paper, we propose a tag-based preference analysis scheme for content classification on feed readers. The proposed scheme extends a current scheme, which is used in many modern websites with the users' positive feedbacks, in such a way that also takes negative feedbacks to improve the accuracy of the analysis. Additionally, a filtering and recommendation system based on such analysis is proposed. A prototype filtering system is also implemented to show the validity of the proposed schemes.

1. はじめに

近年、ブログやニュースサイトなどのリアルタイム性の高いコンテンツを提供する Web サイトが急速な勢いで増えつつある。それに伴い、サイトの更新情報を自動的に取得することのできるフィードと呼ばれる技術が広く用いられるようになってきた。フィードとは、Web コンテンツの概要を記述したもの、もしくはそのフォーマットのことを指し、フィードで記述されたコンテンツの概要は、フィードリーダーと呼ばれる専用のアプリケーションを用いて閲覧することができる。具体的には、各ユーザが関心のあるサイトをあらかじめフィードリーダーに登録し、フィードリーダーに更新情報を適宜自動収集させることによって、登録されたサイトから発信される最新情報のリアルタイムな配信が実現される。

Web コンテンツに対する情報フィルタリング手法の開発は、インターネット上に氾濫する大量の情報の中からユーザにとって有益な情報をすばやく効率的に取得する上で、避けては通れない重要な課題のひとつである。多くのフィードリーダーには情報推薦機能が備えられており、多数のユーザが注目しているホットな記事や、そのユーザが過去に閲覧したコンテンツに関連した記事など、さまざまなレベルでのコンテンツ推薦がおこなわれている。しかしそれらのコンテンツは、あらかじめフィードリーダーに登録されたサイトからの配信に限定されるのが一般的であり、そのためたとえそのユーザが非常に強い関心をもっていたとしても、未登録サイトからの情報は見逃してしまう可能性が高い。この問題を解決するため向井と青野は、フィードリーダーにおける未登録サイトからのコンテンツ推薦手法を提案した⁸⁾。この手法では、各ユーザの嗜好や特徴を各コンテンツに含まれるキーワードなどと比較し、協調フィルタリングによって推薦されるコンテンツを決定している。ただし、向井らの研究ではユーザが閲覧した全てのコンテンツはそのユーザの興味に適合していると仮定している。これは閲覧したコンテンツが実際にはユーザの興味に適合していなかった場合にも自動的にそのコンテンツからユーザのプロファイルを形成してしまうなどの問題がある。

本研究では、各ユーザの Web コンテンツに関する嗜好を、閲覧時にコンテンツに付与されるタグの集合によって明示的に識別する手法を提案する。本手法の特徴は、特定の属性をもつことをあらかず (ポジティブな) タグのみでなく、そのコンテンツが特定の属性に関係していないことをあらかずネガティブなタグを新しく導入している点である。本稿では、それらのタグをそれぞれ正例、負例と呼ぶ。ネガティブなタグ (負例) は、ユーザがフィードリーダーから推薦された情報をもとにあるフィードを閲覧したとき、閲覧した情報が期待

^{†1} 広島大学
Hiroshima University

していた内容とは異なっていた場合に、ユーザ自身の手で付与される*1。負例を明示的に表現することで、タグが付与されていないニュートラルなコンテンツを特徴空間内の距離に基づいて分類することが可能となり、「YES 以外はすべて NO」という考え方でコンテンツを分類する従来手法に比べて、より適切な情報推薦が可能となる。

本稿の構成は以下の通りである。まず 2 節でフィードの説明と関連研究について述べる。3 節で提案システムの詳細について説明した後、4 節でアンケート結果について述べる。提案手法で用いられている関連技術は 5 で概観される。最後に 6 節でまとめと今後の課題を述べる。

2. 関連研究

フィードとはブログやニュースサイトなどの概要を記述したもの、もしくはそのフォーマットのことであり、RSS¹⁾⁻³⁾ や Atom⁴⁾ などの具体例が知られている。RSS フォーマットの例を図 1 に示す。RSS は XML 形式で記述され、一つの channel 要素と複数の item 要素から成る。ここで channel 要素はフィードの配信元サイトの情報をあらわしており、title, link, description の要素を用いて配信サイトのタイトル、URL、そのサイトの説明などを記述する。また item 要素は個々のコンテンツに関する情報をあらわしており、channel 要素と同様、title, link, description などの要素をもつ。フィードは現在多くのサイトで利用されているが、これらのコンテンツを閲覧するためには、フィードリーダーと呼ばれるフィードの管理や閲覧するためのアプリケーションが必要となる。フィードリーダーのうち RSS を対象とするものは RSS リーダー⁵⁾ とも呼ばれる。実際のアプリケーションとしては Google リーダー⁶⁾ などが有名である。ユーザはフィードリーダーに自分が注目するサイトをあらかじめ登録することで、そのサイトを訪問することなく、タイトルやサマリーなどのコンテンツの一部を取得することができる。またユーザは、フィードに含まれているリンクをたどることで配信元のサイトを訪問し、コンテンツの本文を閲覧することもできる。

フィードに関する既存研究として、フィードリーダーに登録しているサイトから配信されるコンテンツを自動的に選別する手法⁷⁾ と、未登録のサイトのコンテンツを推薦する手法⁸⁾ が知られている。これらは両方とも、コンテンツのタイトルや本文中のキーワードに基づいて構成された特徴ベクトルを利用してコンテンツの特徴やユーザの興味をあらわし、それに

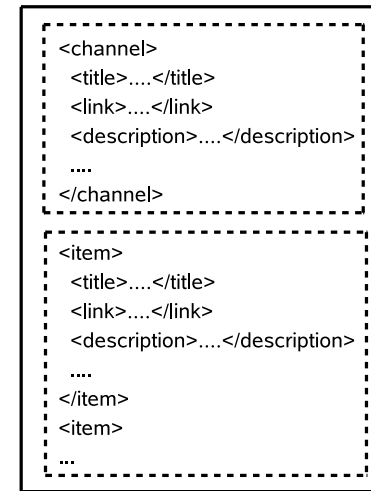


図 1 RSS の構造

基づいたフィルタリングをおこなうことによって、コンテンツの推薦や選別を実現するものである。なお本研究でもこれらの研究と同様、コンテンツのキーワードを用いた特徴ベクトルを利用しているが、ユーザの否定的な意思も利用しているという点で異なっている。

3. 提案システム

3.1 基本機能

提案システムの主な機能は以下の 3 つである：

分類： 明示的なタグの付与によるコンテンツの分類と、ユーザプロファイルの作成

選別： 登録済みサイトから配信されたコンテンツの選別

推薦： 未登録のサイトが配信したコンテンツの推薦

以下では、それぞれの機能の詳細について述べる。

3.1.1 タグ付けによるコンテンツの分類とユーザプロファイルの作成

提案システムでは、フィードリーダーが提示した複数のコンテンツの中からユーザが適切なものを選択して閲覧し、実際に閲覧して興味を持ったコンテンツに対してそのコンテンツの内容をあらわすタグを付与するという基本動作を前提としている。ここでタグとは、ユーザが自由に作成してコンテンツに付与することのできる文字列 (キーワード) であり、ソー

*1 同じ「サッカー」というキーワードにかかわるコンテンツの中でも、「高校サッカー」に関心があるユーザと「プレミアリーグ」に関心があるユーザとは、購読したいと考えるコンテンツにずれがあり、そのようなずれは、ユーザの個人的な心象に起因していることが多いと我々は仮定している。

シャルブックマークやフォークソノミーなどでブックマークサイトの特徴を表したりサイトを分類したりする際に用いられているタグと基本的には同じものである。

あるユーザが異なるコンテンツに同一のタグを付与したとき、そのユーザはそれらのコンテンツが類似していると判断したことになる。類似したコンテンツの集まりをクラスと呼ぶことにしよう(より正確には、コンテンツのクラスは後述のベクトル空間上で定義される)。通常のクラスタリング手法では、あるコンテンツに他のコンテンツとは異なるタグが付与されたり、あるいはそもそも閲覧されなかったりした場合などには、タグが付与されなかったことを根拠として、そのコンテンツは他のコンテンツとは異なるクラスに分類される。しかしそれらのコンテンツはすべてユーザが自身の意志でフィードリーダーに登録したサイトのコンテンツであり、実際には同一サイトの内容の似通ったコンテンツである場合も多い。そのような分類のあいまいさを解消するための方法のひとつとして、提案システムでは通常のタグ付けによって得られる正例(positive example)に加えて、あるクラスに属していないことを示す負例(negative example)もユーザが明示的に指定できるようにする。

ここでの負例であるコンテンツとは、負例タグを付与したコンテンツとする。ここで、負例タグという概念を新たに導入する。負例タグとはある正例に対して負例となるコンテンツに付与するタグとする。本システムでユーザに提示されるコンテンツはユーザの興味に基づいてあらかじめフィルタリングにかけられているため、ユーザは提示されるコンテンツの内容を閲覧する前にある程度予想できると予測される。そこで、実際にユーザが閲覧した際にそのコンテンツがユーザの興味とは異なっていた場合に、予想していた興味のタグの負例タグをそのコンテンツに付与してもらう。この行為でユーザはそのコンテンツが負例である事を明示的に示すことができる。以上のようなコンテンツが負例のコンテンツとなる。

このように正例タグと負例タグを併用してクラス分けをおこなうことで、従来手法のように正例タグのみを用いた場合に比べて、より鮮明なクラス間の境界を求めることが可能となる。以下では、ユーザによって付与された正例と負例に基づくコンテンツの分類のことを、そのユーザのプロファイルと呼ぶ。

3.1.2 登録済みサイトからのコンテンツの選別

フィードリーダーに登録されるサイト数の増加に伴って、フィードリーダーを介してユーザに配信されるコンテンツ数は急速に増加していく。そのため、配信されたコンテンツを各ユーザのプロファイルに基づいて適切に選別し、ユーザの興味に適合しているコンテンツのみを提示させることが、フィードリーダーの可用性を向上させる上で重要な鍵となってくる。本節では、フィードリーダーに登録済みのサイトからユーザに向けて配信されたコンテ

ンツの選別方法について述べる。フィードリーダーは、登録済みサイトからコンテンツを受け取ると、そのコンテンツをユーザプロファイルのベクトル空間上に配置し、配置されたポイントから最も近くに配置された既受信コンテンツを特定する。最近接点の特定は、ボロノイ分割をデータ構造として用いることで効率よくおこなうことができる(5.3節参照)。もし特定された既受信コンテンツに対してそのユーザが正例のタグを過去に付与していれば、フィードリーダーはそのユーザが新しいコンテンツにも興味があるとみなしてそれを提示し、そうでなければ提示することなくそのまま破棄する。

3.1.3 未登録のサイトからのコンテンツの推薦

既存のフィードリーダーでは登録されたサイトからしかコンテンツが配信されないため、ユーザの興味に適合した有益な情報が未登録サイトから配信されたとしても、それを見逃してしまう可能性が高い。提案手法は、正例と負例によって明確化されたプロファイル情報を利用することで、フィードリーダーに登録されていないサイトから配信される有益なコンテンツを、そのコンテンツに興味をもつ可能性のあるユーザに向けて効果的に推薦する。提案手法におけるコンテンツ推薦の手順は以下の通りである：フィードリーダーはまずはじめに、他のユーザのプロファイルに含まれている各クラスの正例と負例に対し、自身のプロファイル中の各クラスの正例と負例をその座標を基に配置した場合の類似度、すなわち自分の正例が他ユーザのあるクラス中の正例の領域に含まれる割合と、自分の負例が他ユーザの負例の領域に含まれる割合をそれぞれ計算する。もし計算された類似度があらかじめ決められた閾値よりも高ければ、それらのクラスは類似した興味をあらわしているとみなされ、相手のもっている類似クラスに属しているコンテンツの中から、未登録サイトから配信されたコンテンツがそのユーザの興味に適合するコンテンツとして推薦される。

3.2 提案システムの構成

本節では提案システムの構成を述べる。提案システムはインタフェース、プロファイル作成部、プロファイルDB(データベース)、選別部、推薦部から構成される。提案システムの構成を図2に示す。それぞれの詳細については以下で述べる。

3.2.1 インタフェース

インタフェースでは、ユーザに対して各コンテンツのタイトルやその本文の一部を提示する。ここで提示されたコンテンツのうち、ユーザが実際に閲覧して興味をもったものにタグが付与される。

3.2.2 プロファイル作成部

プロファイル作成部は、インタフェース上で実行されたコンテンツの閲覧やタグの付与な

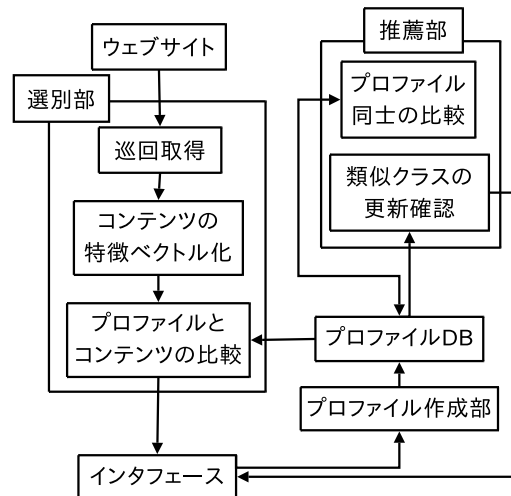


図 2 提案システムの構成

どのアクションを受けて、そのユーザのプロファイルを作成する。プロファイルには、ユーザに提示された全コンテンツの特徴ベクトル、閲覧対象となった Web ページの URL、プロファイルへの追加日時、タグの付与されたものはそのタグ名、そして他のユーザと類似したクラスが存在する場合はそのユーザの ID と類似クラス名が含まれる。なお各コンテンツの特徴ベクトルは、後述の選別部が配信コンテンツを受け取った時点で作成される。

3.2.3 プロファイル DB

プロファイル DB には、全ユーザのプロファイルが保管される。プロファイル作成部で作成されたユーザのプロファイルはここに送られ、選別や推薦でプロファイルを利用する場合は、ここから自分や他のユーザのプロファイルを参照する。

3.2.4 選別部

選別部は、登録済みサイトを定期的に巡回して更新されたフィードを取得し、取得されたフィードに含まれる URL を参照してコンテンツの本文を取得する。その後、取得したコンテンツの特徴ベクトル化をおこない、ユーザプロファイルを参照して、そのフィードを対応するユーザに提示すべきかどうかを判断する。

コンテンツのベクトル化には MeCab⁹⁾ という形態素解析ソフトを用いる。具体的には、取得したコンテンツの本文に対して MeCab を適用して名詞を抽出し、抽出された名詞を要

素とする特徴ベクトルを作成する (5.2 節参照)。作成された特徴ベクトルは、プロファイル DB に保管されているユーザプロファイルと比較される。前述のようにユーザプロファイルには、各ユーザにこれまで配信されたフィードがその特徴ベクトルとともに保管されており、各ベクトルには、ユーザによって明示的に付与されたタグが付加情報として与えられている。前述のように、受信したフィードがユーザに提示されるのは、そのフィードに対応する特徴ベクトルの最近接ベクトルが正例のタグを付与されているときのみであり、提示されることが確定したフィードは、そのままインタフェースに送られる (図 2 参照)。

3.2.5 推薦部

推薦部では、ユーザ間のプロファイルの比較と、類似したクラスをもつ他のユーザのプロファイルの更新確認をおこなう。

前述のようにユーザのプロファイルでは、クラスによってそれぞれ異なる興味を表している。プロファイルの比較ではユーザ間で類似したクラスが存在するかを確認する。この類似度の判定方法を、当該ユーザのプロファイルに存在するクラス A と他のユーザのプロファイルに存在するクラス B を比較する場合を例に説明する。まずクラス A の 1 つの正例コンテンツに注目し、そのコンテンツをユーザのプロファイル上にプロットした場合にクラス B の領域内に含まれるかを確認する。これをクラス A の全ての正例のコンテンツに対して行い、クラス B に含まれたものの割合を求める。次に今度はクラス A の 1 つの負例のコンテンツのコンテンツに注目し、そのコンテンツを他のユーザのプロファイル上にプロットした場合にクラス B の領域外に存在するかを確認し、これをクラス A の全ての負例のコンテンツに対して行い、その割合を求める。これら 2 つの割合の値が閾値を越えた場合にクラス A とクラス B は類似しているとみなす。類似したクラスを発見した場合は他のユーザを識別する ID とクラス B を表すタグを当該ユーザのプロファイルに記録する。一方、類似した他のユーザのプロファイルの更新確認では、ユーザ間のプロファイルの比較によってプロファイルに記録された、類似した他のユーザのクラスについて調べる。具体的には、類似クラスに属するコンテンツのプロファイルへの追加日時を確認し、前回の確認時よりも後に追加されたものをインタフェースへ送る (図 2 参照)。

4. 評価実験

提案手法の有効性を検証するため、提案システムのプロトタイプと、負例タグを用いない同様のシステムを比較対象として作成した。なお作成期間の関係上、今回は選別に関する実験のみ行う。被験者は両システムを利用し、選別された各コンテンツに対して、適切に選別

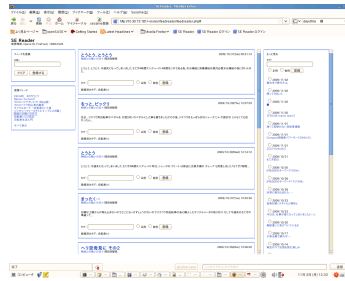


図3 プロトタイプシステム

されたか評価してもらい、その結果から選別の精度を調べる。

4.1 実験方法

被験者は同じ研究室に所属する大学生1名とし、システムの利用とアンケートを行う。用いているシステムは提案システムの選別のみを行うプロトタイプシステムと、プロトタイプシステムと同様の動作をする負例タグを用いないシステムである。後者のシステムは負例タグを用いない代わりに正例タグを付与されなかったコンテンツ全てを自動的に負例タグのついたコンテンツとしてプロフィールに書き込む。なお、本実験では選別によって負例とみなされたコンテンツも、タグを付与してプロフィールを構築するためにタイトルのリストを別枠に表示する。プロトタイプシステムのインターフェースを図3に示す。図の左側がサイト名、中央が選別によって提示されたコンテンツ、右側が選別によって提示すべきでないとしたコンテンツのタイトルである。実験の手順は、以下の通りである。

- (1) 事前の準備として両システムに対して初期プロフィールを作成する。まず被験者に自身の興味を1つ定めてもらい、その興味に関するコンテンツを配信しそうなサイトをシステムに登録する。登録するとそのサイトから配信されたコンテンツが被験者に提示されるので、各コンテンツに対してタグを付与する。以上でプロフィールが完成する。
- (2) 評価用に、自分の興味に類似したサイトを新たに1つ登録する。登録するとプロフィールを基にそのサイトのコンテンツの選別が行われる。
- (3) 選別された各コンテンツに対して正しく選別されたか評価をする。評価基準は「○:適切に選別が行われた」、「×:適切に選別が行われなかった」、「△:コンテンツが被験者の興味に適合しているとしてもしていないともとれる内容であるため適切に選別が行われたかどうか判断し兼ねる」の三段階とした。

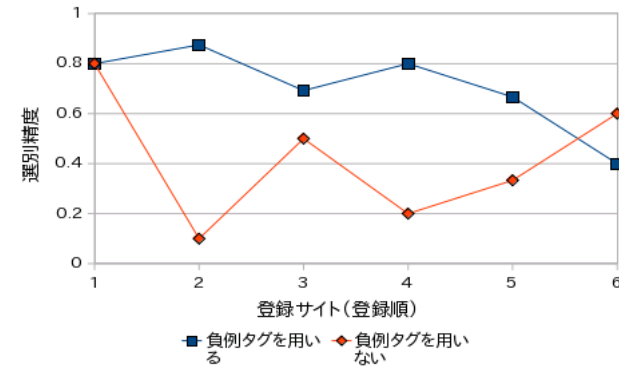


図4 負例の有無による選別の精度

- (4) 選別された各コンテンツに対してタグを付与する。これによりプロフィールが再構築される。
- (5) (2)に戻って同様の作業を繰り返し、全部で6つのサイトに対して実験を行う。

4.2 実験結果

実験によって得られた各コンテンツに対する評価を基に、サイトごとに以下の式によって選別の精度求め、その結果を図4に示した。

$$\text{選別精度} = \frac{\text{評価が○のコンテンツ数}}{\text{全コンテンツ数} - \text{評価が△のコンテンツ数}}$$

図4より明らかに負例タグを用いた場合の精度が高くなっている。しかし、この結果は必ずしも提案システムの有用性を示すものではない。その要因として、今回の実験で登録したサイトは正例のコンテンツが多かった点と、負例タグを用いないシステムでは多くのコンテンツが負例として選別された点があげられる。後者の原因として、コンテンツから名詞を抽出する際にコンテンツの内容とは関係のない名詞も多く取得されてしまい、その名詞の種類がサイトごとに偏っていたことから、興味ごとではなく、サイトごとにコンテンツの特徴ベクトルが偏ったのではないかと考えられる。そのため、今後コンテンツから名詞を抽出する際に、その内容と無関係なものの除去を適切に行えるように改良する必要がある。

5. 関連技術

5.1 フィルタリング

提案手法では、コンテンツの選別と推薦でそれぞれ異なるフィルタリング手法を採用している。選別フェーズではコンテンツベースフィルタリング¹⁰⁾が用いられる。フィードリーダーが選別対象とするコンテンツはそれぞれ作成されて間もないことから、選別に必要な情報は、アクセス履歴などではなくコンテンツ自身から抽出される必要がある。コンテンツベースフィルタリングはコンテンツの選別をコンテンツ自身の特徴とユーザの類似度とに基づいておこなうフィルタリング手法であり、そのような選別フェーズの特性によくマッチしている。推薦フェーズでは協調フィルタリング¹¹⁾が用いられる。前述したように提案システムでは、コンテンツの推薦は選別されたコンテンツがユーザたちに提示された後で実行される。したがって各ユーザがどのようなコンテンツに興味をもっているかというプロフィール情報を利用することができ、興味の類似したユーザに同一のコンテンツを推薦する手法である協調フィルタリングを効果的に利用できる。

5.2 特徴ベクトル

提案手法では各コンテンツをベクトル空間中の特徴ベクトルとして取り扱う。コンテンツ中に d 種類の名詞があらわれる場合、そのすべての名詞の集合を $\Lambda = \{\lambda^1, \dots, \lambda^d\}$ とする。また、総コンテンツ数を n 個とした場合、その集合を $C = \{c_1, \dots, c_n\}$ と表し、これらのコンテンツに対応する特徴ベクトルの集合を $V = \{v_1, \dots, v_n\}$ と表す。 v_i の各要素 $v_i^k (k = 1, \dots, d)$ は、tf-idf 尺度を用いて以下のように表現される：

$$v_i^k = tf_i(\lambda^k) \cdot idf(\lambda^k)$$

ただし $tf_i(\lambda^k) = c_i$ 中の名詞 λ^k の頻度 / c_i 中の全名詞の出現の総和、および $idf(\lambda^k) = \log(\text{総コンテンツ数} / \text{名詞 } \lambda^k \text{ が出現するコンテンツ数}) + 1$

5.3 ボロノイ分割

本システムではベクトル空間上のコンテンツの集合をクラスとし、1つのクラスで1つの興味を表している。このクラスの境界を求めるためにボロノイ分割を利用する。ボロノイ分割の例を図5に示す。ボロノイ分割とは空間上に配置された複数の点に対し、空間上の他の地点をそこから最も近い点の領域となるように空間を分割する手法である。分割された各領域はボロノイ領域と呼ばれ、空間に n 個の点が存在する場合、各点を $p_i (i = 1, \dots, n)$ と表し、その点が存在するボロノイ領域を $V(p_i)$ とする。点 a と点 b のユークリッド距離を

$d(a, b)$ で表すとすると、ボロノイ領域 $V(p_i)$ は

$$V(p_i) = \{p \mid d(p, p_i) < d(p, p_j), j \neq i\}$$

で表すことができる。このため、新たにコンテンツの特徴ベクトルをベクトル空間に配置した場合、そのコンテンツはそこから最も距離の近いコンテンツの領域に位置しており、これらは同じ興味を表していると予想することができる。

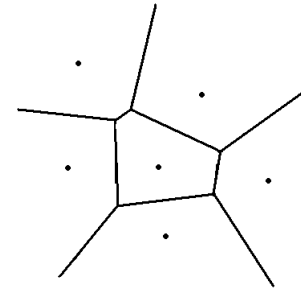


図5 ボロノイ分割の例

6. 終わりに

本稿では否定的な嗜好をユーザが明示的に示すことでその興味を詳細に表現し、その興味をフィードリーダーに用いることで、よりユーザの嗜好に適したコンテンツの選別と推薦をするシステムの提案を行い、そのプロトタイプシステムによって選別に対する評価実験を行った。今後の課題として、コンテンツからその特徴を適切に抽出する方法を考案する必要がある。

謝 辞

本研究の評価やシステムの作成に協力していただいた研究室の皆様へ感謝致します。

参 考 文 献

- 1) RSS 0.92. <http://backend.userland.com/rss092>.
- 2) RDF Site Summary (RSS) 1.0. <http://web.resource.org/rss/1.0/spec>.
- 3) RSS 2.0 Specification (version 2.0.11). <http://www.rssboard.org/rss-specification>.

- 4) AtomEnabled.org. <http://www.atomenabled.org/>
- 5) RSS リーダーとは【RSS reader】 - 意味/解説/説明/定義：IT 用語辞典. <http://e-words.jp/w/RSS%E383AAE%E383BCE%E38380%E383BC.html>.
- 6) Google リーダー. <http://reader.google.com/>.
- 7) Juan J. Samper, Pedro A. Castillo, Lourdes Araujo, J.J. Merelo, Oscar Cordon, Fernando Tricas. NectaRSS, an Intelligent RSS Feed Reader. Journal of Network and Computer Applications, Vol.31, Issue 4, pp.793-807, 2008.
- 8) 向井誠, 青野雅樹. RSS に基づく個人向け内容型情報推薦プロトタイプシステム. 情報処理学会研究報告, FI, No.94, pp.27-32, 2005.
- 9) 工藤拓. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. <http://mecab.sourceforge.net/>.
- 10) 市川裕介, 中村美穂, 畑恵介, 中川哲也. サイト横断でユーザ個々の嗜好に合ったサービス提供を実現するパーソナルサービス基盤技術, NTT 技術ジャーナル, Vol.20, No.1, pp67-70, 2008
- 11) 福原知宏. 協調フィルタリングに関する研究動向. <http://www.race.u-tokyo.ac.jp/fukuhara/Research/paper/98/cofil.pdf>.