




| | | |
|---------|---|---|
| 氏名・職名 | 岩沼 宏治 教授 |  |
| キーワード | データマイニング, オンライン系列マイニング WEBインテリジェンス, SATソルバー | |
| ホームページ | http://www.iwlab.org/our-lab/our-staff/iw | |
| 略歴 | 電子情報通信学会・人工知能研究専門委員会委員長、 山梨大学・IT推進本部副本部長など | |
| 受賞 | 人工知能学会全国大会優秀論文賞(1987,89,90,91)、 情報科学技術フォーラム(FIT2004)優秀論文賞 日本ソフトウェア科学会 第3回ソフトウェア論文賞(2014) | |
| 研究者から一言 | 知的情報処理の理論と技術について研究と開発を行っています。皆様のお役に立てれば幸いです。 | |

系列データマイニング

情報爆発時代と呼ばれる今日、WEBコンテンツ、メディア記事、ソフトウェア／ハードウェアから、交通情報、生命システムにいたる多種多様なデータの大規模蓄積化が急速に進んでいます。本研究室では、これら**大規模かつ複雑に構造化されたデータから有用な知識を自動抽出するデータマイニング(DM)技術**の開発に取り組んでいます。

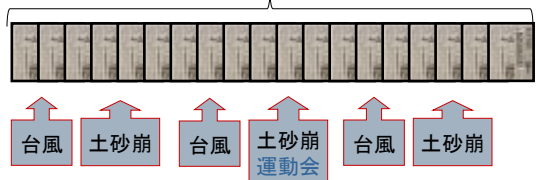
例えばコンビニでは、お客さまの**購買履歴**から、どういった商品と商品の組合せ(例:「パン」と「缶コーヒー」)が同時に購入されているかを分析することがありますが、この分析には「DM技術」が応用されています。同じ様な事例として**WEBサーバーのアクセス履歴**解析なども挙げられます。対象となるデータは履歴に限りません。情報社会におけるDM技術の用途は多岐にわたりますが、本研究室では**系列データを対象とする系列DM**を得意としています。

新聞等のメディア記事を対象とすると、明示されていない事象の時系列関係(例:「台風」の後に「土砂崩れ」が起きる)を求めることも可能となる。また、このような関連する事象を抽出することで、関連情報の自動検出や自動追跡といったこともできるようになる。

課題

通常は、頻出なパターンを有用パターンとして抽出するが、その中には、価値のないイベント(単語)を抽出してしまう傾向がある。

過去数年分の新聞記事



解決策として、本研究室では

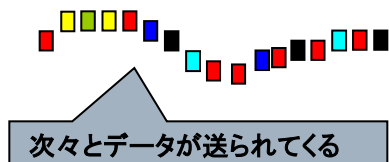
1. 頻度尺度と情報量を掛け合わせた尺度の利用し、併せて高速計算法を開発
2. イベント間の共起性を考慮した尺度の利用方法を開発

※ 共起性＝同時に出現する割合のことで、関連性の強さを表す

課題

連続して大量のデータが送られてくる場合(＝ストリームデータ)、データマイニングの処理が追いつかず、メモリが不足する可能性がある。

例) クレジットカードの利用履歴、通信トラフィック、プラント制御情報



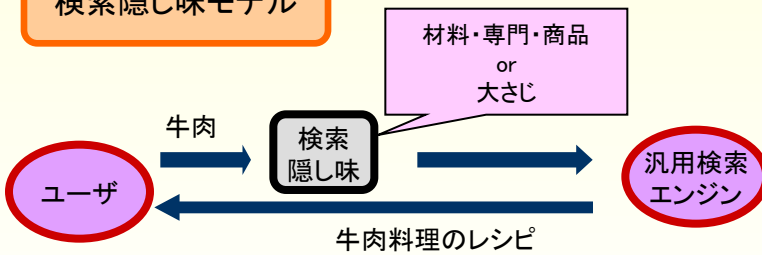
解決策として、本研究室では

データを一度だけ読んで高速に処理(オンライン型アルゴリズム)し、更にある種の誤差を許して、有効性の少ないと思われるデータの廃棄し、メモリを大幅に節約するアルゴリズムを開発。



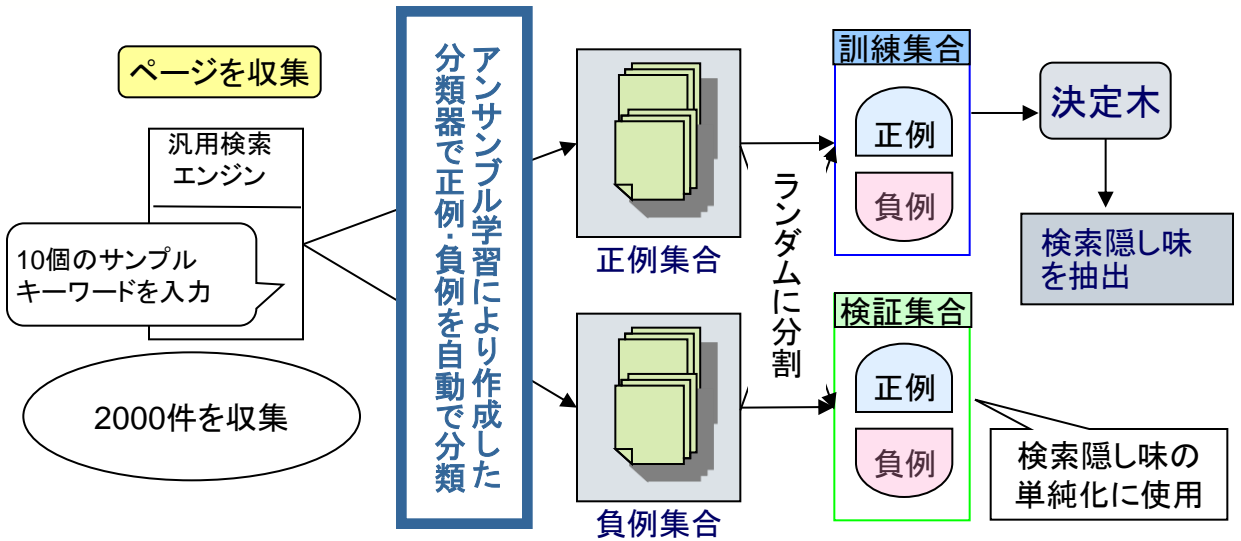
専門検索エンジンの半自動構築

検索隠し味モデル



一般的な検索エンジンを使うと意図しないサイトが上位表示されることがありますが、この問題を補う要素技術を研究開発しています。

適切な検索隠し味を追加すれば、汎用検索エンジンが、料理レシピの専門検索エンジンとして働かせることができる。そこで適切な検索隠し味を専門ごと個別に半自動生成する。



適用できる製品・分野のイメージ

- ネットワークやプラントなど監視システム
 - 商品販売監視と分析
 - 商品サービス推薦システム
 - 検索サービスの付加価値向上
- Etc...

皆様のお役に立てれば幸いです。

シーズについてのお問合せ、ご相談先
E-mail: renkei-as@yamanashi.ac.jp
Tel: 055-220-8759 Fax: 055-220-8757

