研究室公開日（オンライン）：2 月 18 日(木) 11 時，13 時，15 時（その他の日時・対面は予約対応）

応用数理ネットワーク研究室　C3-303

梅村恭司 教授（umemura@tut.jp），吉田光男 助教（yoshida@cs.tut.ac.jp）

廣中詩織 特任助手（hironaka.shiori.qp@tut.jp）

☆どのようなことを勉強するか？

文字列アルゴリズム／機械学習アルゴリズム／ノンパラメトリック統計処理

さらに，英語／プレゼンテーション／技術文章の書き方

☆テーマの考え方

ユニークな技術／最新の技術を習得して前人未踏のニーズを開拓する。

☆どのようなアプローチで研究するか？

統計処理とアルゴリズムを道具として，蓄積された大量のデータを扱う。

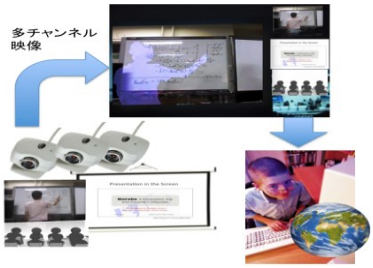
データマイニングで使われているアプローチを踏襲して研究する。

☆キーワード

応用統計学，情報検索，自然言語処理，データマイニング，計算社会科学，e ラーニング

○教育応用のためのネットワークシステム

コンピュータで映像を処理，蓄積，配信する技術を利用し教育システムへの応用をしている。遠隔授業が必要となったとういう社会変化を前向きに捉え，新しく生まれたニーズに対応するシステムを作成し，実際に使いながら改善していくというプロジェクトである。このプロジェクトによって，Web ベースのアプリケーションの構築技術と，ビデオと音声の技術が習得できる。さらに，教育という観点から，人間の振る舞いに対しての理解が深まる。



ネットワーク時代の教育のサポート

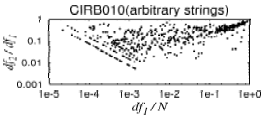
○統計的な分析のためのアルゴリズムと統計分析

確率・統計の技術はコンピュータに高度な処理をさせる基盤である。そして，これを有効に活用するには，数を数えるという操作が必要であり，これには高速なアルゴリズムがある。研究室公開においては「重要な文字列を特定する」という統計処理を紹介する。着眼点は，「重要な単語は，ある部分に繰り返して出現する」という性質である。これは，統計的に処理できることを説明し，さらに，それを実用で使うために開発したプログラムのデモンストレーションを行う。この研究成果は，競争力のあるソフトウェアの開発につながっており，共同研究先で製品化されている。

計算量：長さNのすべての部分文字列はN(N-1)/2 個
それぞれについて、文書頻度を求めたい。

実現した方法　前処理：O(N log(N))，文字列ごとに：O(log(N))

x	df(x)	df2(x)
ロ	124696	79894
ロボ	3672	2413
ロボッ	3320	2237
ロボット	3320	2237
ロボットに	577	96
ロボットにつ	30	1
ロボットについ	30	1
ロボットについて	30	1



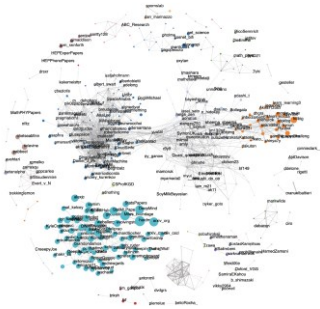
df2（x）：文字列xを2回以上含む文書数

図2　全文字列を対象とした分析

高速な文字列統計値の計数法

○ウェブ／ソーシャルメディアの大規模データ活用

ソーシャルメディア（Twitter）が日常的に利用されるようになり，そこには様々な情報が投稿されている。世界中の位置情報付きツイートやリツイートなどを大規模に活用し，フォロー関係（ソーシャルグラフ）から居住地などのユーザ属性を推定したり，ソーシャルメディアでの投稿内容を分析したりしている。さらに，ウェブのニュース記事データと組み合わせ，ソーシャルメディアで RT/Like されやすいニュースの分析や特徴の抽出，あるいは，ウェブの学術情報（論文）データと組み合わせ，学術情報を共有・取得するユーザコミュニティの分析を行っている。



論文共有コミュニティの可視化