

言語のための数理的手法の開発

教授　高村 大也

研究分野：計算言語学、自然言語処理、機械学習

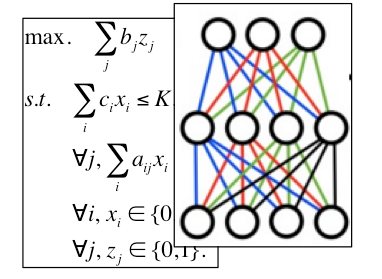
ホームページ: http://www.lr.pi.titech.ac.jp/~takamura/index\_j.html

●研究内容・目的

　人間は、単なる記号の羅列でしかないテキストから、国の経済状況を把握したり、新発売の携帯電話についての意見を読み取ったり、友人の悩みに共感したり、物語に感動したりすることができます。逆に自分の知識、意見、感情を記号の羅列に変換して、相手に伝えることもできます。そこには、コンピュータが簡単に真似できない非常に高度な知識処理が行われているはずです。自然言語処理学は、その簡単に真似できないはずの非常に高度な知識処理を、コンピュータに行わせる技術を開発するという、とても挑戦的な学問です。自然言語処理には、機械翻訳、文書要約、意見分析など様々な応用があり、世の中で強く必要とされています。これらの応用は、記号の羅列から単語を見つける単語分割、単語や文の意味を理解する意味解析、文の構造を同定する構文解析など、多くの要素技術から成り立っています。私達の研究室では、要素技術から応用まで、自然言語処理における課題を幅広く研究しています。

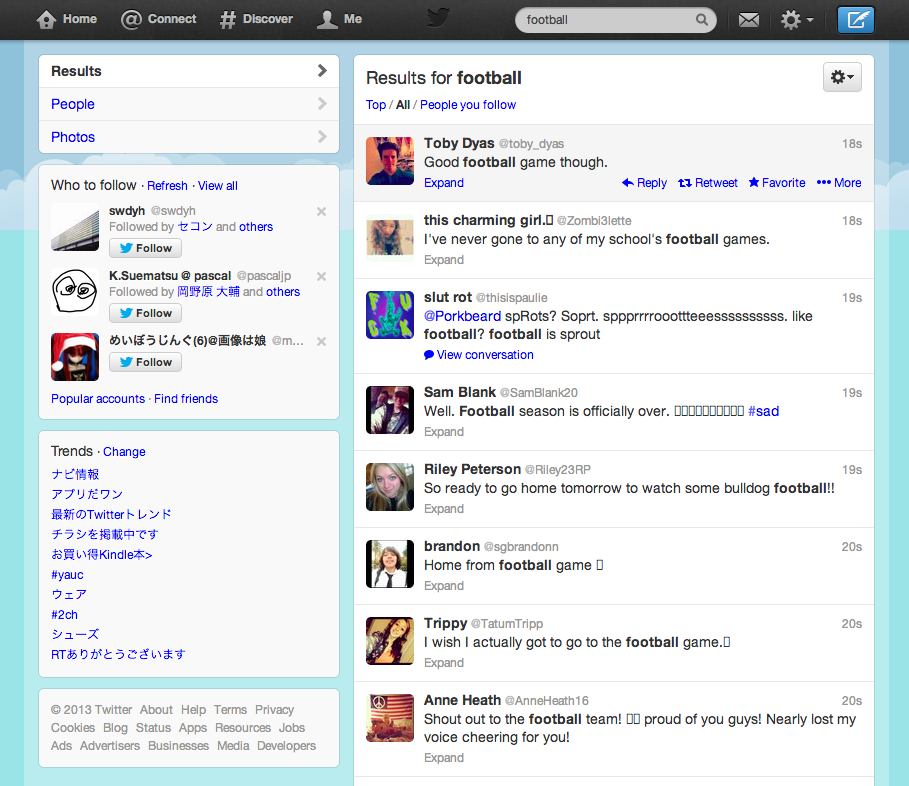
　実際に処理方法を開発し始めると、言語表現は多様でありかつ曖昧であるという事実と対面することになります。その細かな処理過程を人間が一つ一つコンピュータに教えてあげることは現実的ではありません。コンピュータ自身にデータから学んでもらうのが得策です。ここで登場するのが機械学習という技術です。私達の研究室では、線形分類器やトピックモデルから、発展途上にある深層学習まで様々な機械学習手法の中から問題に適したものを選択して利用し、必要に応じて手法を拡張して問題を解いています。

　また、多様な情報をエンコードすることができる言語という体系がどのようなものであるか、また人間の言語能力はどのような性質を持っているのか、といった疑問に対しても、計算的手法を用いることで答えを探る研究をしています。

●研究テーマ

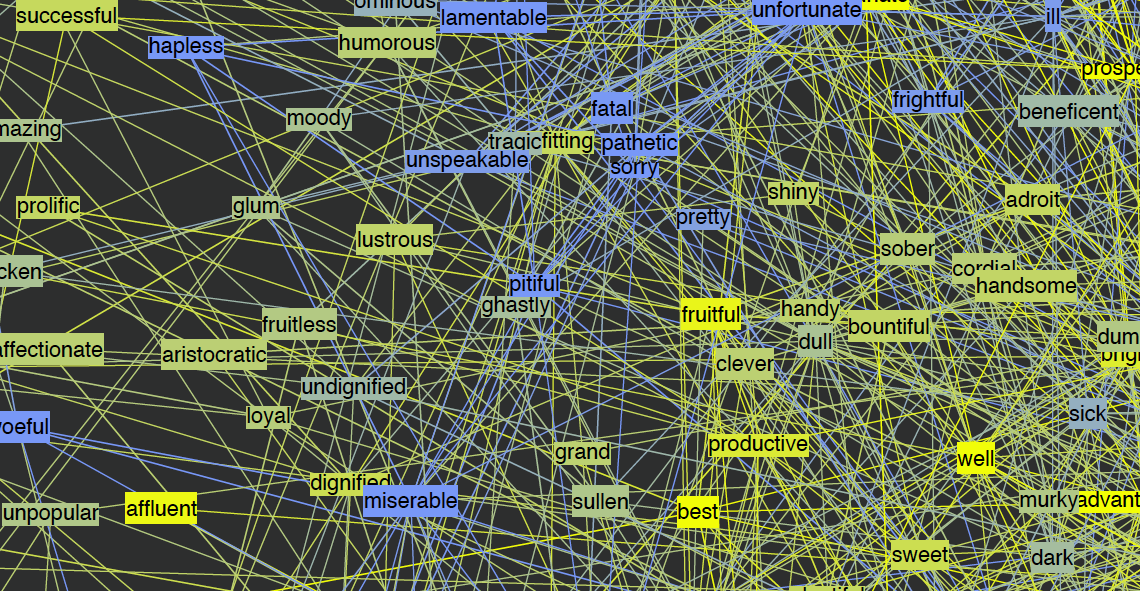
１．文書要約手法の開発

　非常に長い文書、あるいは多数の文書が与えられた時、その内容を簡潔に表す短い文書(要約)を生成する文書要約手法を開発しています。特に、最適化問題による要約のモデル化や、機械学習による定式化に力を入れています。

２．言語データを通して世界や社会を見る技術の開発

　インターネット上、特にソーシャルメディア上には、人々の意見や感情が溢れています。製品やサービスに対する意見、政治に対する意見、世間のマナーに対する意見など、多くの意見が大量に存在し、貴重な情報源となっています。何に対するどういった意見がどれくらいあるのかを自動的に把握することができれば、製品、サービス、政治、マナーなどをより良い方向に向けるための手助けになるでしょう。そのための技術の開発も、私達の研究室の重要なテーマです。

3．数理的アプローチによる言語研究

　世界には、日本語のように手の指も足の指も「指」という一つの単語で表現される言語と、英語で「finger」と「toe」と呼ぶように別の単語を持つ言語があります。このように各言語は様々な特徴を持ち、思わぬところで違いがあったりします。語彙の構造など、各言語の特徴、あるいは言語間の差異、また言語一般に存在する性質なども研究対象です。

●教員からのメッセージ

　言語に興味のある方、言語で書かれた知識に興味のある方、電化製品を購入しようと思って製品レビューを大量に読んで時間を浪費してしまい、レビューを簡潔にまとめてほしいと思ったことがある方、などなど、ぜひ一緒に研究しましょう。

●関連する業績、プロジェクトなど

1. Hiroya Takamura, Ryo Nagata and Yoshifumi Kawasaki. "Discriminative Analysis of Linguistic Features for Typological Study" (to appear), In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, 2016.

2. 菊池悠太, 平尾努, 高村大也, 奥村学, 永田昌明, "入れ子依存木の刈り込みによる単一文書要約手法", *言語処理学会論文誌 自然言語処理*, Vol.22, No.3, pp. 197-217, 2015.

3. Mitsumasa Kubo, Ryohei Sasano, Hiroya Takamura and Manabu Okumura,

"Generating Live Sports Updates from Twitter by Finding Good Reporters",

In *Proceedings of Conference on Web Intelligence*, pp. 527-534, 2013.

4. 高村大也, 奥村学, "施設配置問題による文書要約のモデル化", *人工知能学会論文誌*, Vol.25, No.1, pp. 174-182, 2010.

5. 高村大也, 奥村学, "最大被覆問題とその変種による文書要約モデル", *人工知能学会論文誌*, Vol.23, No.6, pp. 505-513, 2008.

6. 高村大也、"言語処理のための機械学習入門", コロナ社, 2010.