

京都大学 大学院情報学研究科 知能情報学コース

言語メディア分野

黒橋禎夫 Chu Chenhui 村脇有吾
Huang Yin-Jou Cheng Fei 清丸寛一

本研究室では自然言語処理(Natural Language Processing; NLP)の研究を行っています。そこには、言語とコミュニケーションの仕組みを計算機が扱える正確さで解明するという理論的研究と、それによって情報集約、機械翻訳、対話システム等をより高度化して人間の活動を支援するという工学的研究の2つの側面があります。

■言語理解の基礎的研究

計算機による言語理解を実現するには、計算機に常識・世界知識を与えることが不可欠です。ディープラーニング技術の進展により生み出された汎用言語モデルは、単純な仕組みであるにも関わらず、驚くほど柔軟な知識処理が可能であることがわかってきました。本研究室でも、大規模WebテキストをGPUクラスタ計算機で処理することで汎用言語モデルの訓練を行うとともに、それらのモデルを使って、計算機による文章理解の研究を進めています。また、言語が時間とともに変化することは言語の本質に関わる性質だと考え、言語の歴史的変化を統計的に解析する研究も行っています。

Q: Find the amount of Calcium hydroxide that is required to react with 2 moles of Carbon dioxide to form 2 moles of Calcium carbonate along with 2 moles of Water

A: First, use the chemical reaction predictor to determine the chemical equation for the reaction.

Reactants: ?Ca(OH)₂, 2CO₂

Products: 2CaCO₃, 2H₂O

<< Chemical reaction predictor >>

2Ca(OH)₂ + 2CO₂ → 2CaCO₃ + 2H₂O

Second, use the molar mass list to determine the molar mass of Calcium hydroxide. ...

外部モジュール
化学反応予測器の呼び出し

汎用言語モデルと外部モジュールの組合せによる質問応答
(質問Qに対する回答Aを推論過程とともにモデルが生成)

■言語処理の実社会応用に関する研究

言語処理の高度化に伴い、あらゆる分野で言語処理の利用が始まっています。実データを持つ様々な機関と協力しながら言語処理の実社会応用に関する研究を展開しています。

言語処理による知識の構造化 テキストには様々な知識が眠っていますし、対話を通じて能動的に人から知識を引き出すこともできます。テキストを解析し、「誰が何をどのように行うのか」という形で知識を構造化することで、診療記録から創薬につながる手がかりを探したり、インタビューを通じて熟練の技能者から暗黙知を引き出すことを支援するといった様々な社会的応用を進めています。

機械翻訳の実応用に向けた研究 機械翻訳はディープラーニングの導入によって大きく精度が向上し、さまざまな場面で活用され始めています。本研究室では、短文脈依存性が高いドラマのセリフを適切に翻訳するために映像を手がかりとして利用したり、文化的な違いから適切に訳すのが難しい表現を検出して翻訳しやすいように書き改めるといった研究を進めています。

対話システム 機械翻訳の実利用が視野に入った現在、自然言語処理の次のキラアプリケーションとして注目されるのが対話システムです。社会実験も推進しており、地域コミュニティの活性化や行政サービスの円滑化を目標として、尼崎市・丹波市、LINE株式会社と協力し、行政サービスに関する市民の問い合わせに自動的に応答できる対話システムを構築しました。