

公益の風 #33



東北公益文科大学大学院
公益学研究科修了生
田雨冉



Harvard-Yenching 図書館が公開している『Naxi manuscripts collection』の一部写真

中国には豊かな多民族文化があり、56の少数民族がそれぞれ独自の才能を発揮しています。私は雲南省昆明市の出身で、ナシ族末裔の一人です。ナシ族の中には「世界で唯一生きていく象形文字」として知られるトン

機械学習を活用してナシ族・トンパ文字の魅力を発信する

研究の目的は、トンパ文字の自動解析ツール開発の支援とトンパ文字を継承するために、トンパ文字経書の注釈付き画像のデータセットを構築し、そのデータをオープンデータとして、公開することです。このデータセットには現時点では103ページの写本と7000字近くの文字、1103以上の異なる絵文字の注釈が含まれています。また、提案データ

その理由の一つに、トンパ文字の注釈付きデータが不足しているということがあります。筆者の修士論文研究では、雲南省ナシ族の魅力を発信しながら、トンパ文字を自動解析するためのデータセット構築に焦点を当てています。

近年、音声認識、手書き文字認識、機械翻訳などの人工知能や音声・テキスト処理技術は、人々の日常生活に徐々に浸透してきています。これまでもナシ族について多くの人文・社会科学系の研究がされてきましたが、現代の技術や機械学習を活用するトンパ文字の研究は限られていました。

パ象形文字を持っていません。トンパ文字は、ナシ族の「トンパ」という祭司から名をもらい、主に宗教文書を記録するために使用されてきた象形文字です。



指導教員のノヴァコフスキー先生との講義時に撮影



この研究は、トンパ文字の

さらに、研究では3735の絵文字の視覚的特徴と構成要素の分析が行われ、キーワードベースのラベルフィルタリングを適用することで、アンテーションワークフローが最適化されました。これにより、トンパ文字の自動解析における効率が向上しました。

セットは、ディープラーニングモデルの学習に使用され、トンパ文字で書かれたテキストの自動分析ツールの開発に役立つことが実験的に検証されました。



院生研究報告会の様子

こうした結果から、提案されたデータセットがトンパ文字の自動解析ツールの開発に有用であることが示されます。この研究は、トンパ文字の

言語処理技術の発展に寄与するだけでなく、ナシ族の文化の保存と継承にも貢献することが期待されます。

ナシ族の文化や言語、トンパ象形文字は、中国または世界にとって、失われてはならない貴重な文化財だと考えています。本研究の成果は、指導教員のカロール・ノヴァコフスキー先生をはじめ、多くの支援者の協力によって達成されました。この研究が、トンパ文字の自動解析技術の発展において重要な一歩となり、今後の研究や言語処理技術の発展に貢献し、ナシ族の文化遺産の保存や言語研究分野における革新に寄与することを願っています。