

(続紙 1)

京都大学	博士（情報学）	氏名	張 訳心
論文題目	Multimedia Procedural Instruction Understanding through Image-Text Sequence Analysis (画像・テキストシーケンス分析によるマルチメディア手順書の理解)		
(論文内容の要旨)			
本論文は、世の中に存在する様々な作業手順指示書のデータを計算機を用いて自動的に整理・構造化し、より理解容易な形とするための技術を提案している。そのような作業手順指示書の多くは、手順をステップの列として説明し、各ステップの説明はテキストによる説明とそれを補足する画像からなるという構成を取っている。そこで、本論文ではテキストと画像のペアの列という構造を持つ手順指示書を対象とし、そのような形式の手順指示書に特に顕著な問題を選んで取り上げている。また、提案する技術の有効性を検証する実験においては、このような手順指示書の典型的な例として、料理レシピのデータを取り上げている。 本論文は10章からなる。第1章では、まず上で述べた本研究の主旨について述べ、さらに、本論文で具体的に取り上げる問題五つについて説明している。第2章では関連する既存研究についてまとめており、第3章では検証実験に用いるために自ら作成した料理レシピデータのデータセットについて説明している。 第4章では、手順指示書には様々な「行為」を表す専門用語が登場し、それらの理解が重要となることをまず指摘し、初見の行為の理解には、その行為が既知のどの行為と類似するかの情報が有用であることを述べ、そのような行為間類似度を測るための手法として、その行為が現れるテキストはどのような画像とペアになっているかに基づいて、その行為の特徴を表すベクトルを生成する手法を提案している。 第5章では、テキストと画像のペアからなる説明では、テキストと画像のどちらか一方で明示されているものは、もう一方ではしばしば省略されるということを指摘し、そのような省略されたものを補完するための手法として、省略が生じているステップと類似するステップを、テキストの類似度と画像の類似度の双方を用いて発見し、それらの類似するステップに明示されている情報を用いて補完する手法を提案している。 第6章と第7章では、手順指示書中の画像においては同一物が作業の進展に応じて外観を変えていくことがあるため、そのような物体の自動認識が課題となることを指摘し、そのような物体の自動認識の精度を向上させる方法を二つ提案している。一つは、その画像が付けられているステップが手順全体のステップ列の中の序盤か中盤か終盤かに応じて異なる認識モデルを作成する手法であり、もう一つは、カリキュラムラーニングの手法を導入して、序盤のデータから学習を開始し、中盤のデータ、終盤のデータと学習を進める手法である。第6章では、一種類の物体を認識する場合、第7章では複数種類の物体を認識する場合について論じている。 第8章ではテキストと画像のペアの列という構造を持つ手順指示書中の画像からの物体認識では、先行するステップに現れる物体の情報が有用であること、ただし、作業手順においては手順の流れに分岐や合流があるため、必ずしもステップ列中の直前の			

ステップから物体を引き継いでいるとは限らないことを指摘し、先行するステップの中で関連が深いと判断されるステップの情報を利用して物体認識の精度を向上させる手法を提案している。

第9章では、外観が酷似する二種類の物体が一つの作業手順の中にも同時に現れることは稀である（例えば、小麦粉と米粉が同じレシピに現れることは稀である）こと、ある作業手順書に現れる物体の一覧は事前知識（例えばレシピの材料一覧や家具組立の際の部品一欄）として与えられることがあること、よって、ともに現れることが稀である組合せは、そのいずれかであることさえ判定できれば、どちらであるかまでは判定できなくともよい（例えば、小麦粉または米粉であることさえわかれば、そのレシピの材料一覧からどちらであるかは判定できる）ことを指摘し、そのような点を考慮した機械学習の手法として、学習時に区別できなくてよい組合せを誤認識した際にはロス関数の値をゼロにする手法を提案している。

第10章は結論と今後の課題について述べている。

(論文審査の結果の要旨)

本論文は、世の中に存在する様々な作業の手順指示書が、それらの作業に関してわれわれが長い年月をかけて蓄積してきた知識の貴重な記録という側面を有していることに注目し、これらの貴重な知識を断絶させることなく継承していくため、あるいは、今後、人間だけでなく機械にも自動的に理解させることによってより一層活用していくために、これらの手順指示書の情報を計算機を用いて自動的に整理・構造化し、より理解容易な形とするための技術を提案している。よって、社会的価値の高い課題に取り組んでいると評価できる。また、そのような作業手順指示書の大半は、テキストと画像のペアの列という構造を持っているため、そのような形式の手順指示書に特に顕著な問題を取り上げている。よって、本研究で提案する各手法の検証実験においては料理レシピデータを用いてはいるものの、提案手法はいずれも様々なデータに適用可能な一定の汎用性を有していると考えられる。

第4章では、手順指示書には様々な「行為」を表す未知の語が登場するため、それらの理解のために、その行為が既知のどの行為と類似するかを、行為を表す語がどのような画像と同じステップに現れるかの情報を用いて推定する手法を提案している。このような「行為」を表す未知の語の理解は、手順指示書の理解においては重要な問題であるため、有用な手法の提案であると評価できる。

第5章では、ステップの説明中で省略されているものを補完するために、テキストの類似度と画像の類似度の双方を用いて類似する他のステップを発見する手法を提案している。これは、テキストと画像のペアを用いて何かを説明しているデータにおいては、どちらか一方で明示されているものは、もう一方ではしばしば省略されるということに注目した研究であり、手順指示書以外にも応用可能な汎用性の高い技術と評価できる。

第6章、第7章では、作業の進展に応じて外観を変化させる物体の画像認識の精度を向上させる手法を提案している。これも、作業手順指示書においては極めて広く見られる問題に関するものであり、また、作業手順指示書以外の様々な動画などへも応用できる可能性を持つ技術として評価できる。

第8章では、あるステップの画像中の物体の認識精度を向上させるために、先行するステップのうち、対象ステップへと物体が引き継がれていると思われるものの情報を利用する手法を提案している。これも、第6章、第7章の手法同様、動画などへも応用できる可能性を持つ技術として評価できる。

第9章では、区別できなくても構わないような組合せがある場合の画像認識の精度向上の手法を提案しており、これも作業手順指示書以外にも応用可能な汎用性の高い手法として評価できる。本論文でも、出演者リストが与えられている映画の画像からの俳優の認識にも応用できることが示されている。

以上、本論文は、人類の貴重な財産である作業手順に関する知識の継承と有効利用に資する有益な技術を提案しているものであり、また、作業手順書以外へも応用可能な汎用性の高い手法の提案ともなっているため、学術的にも実应用的にも寄与するところが大きい。よって、本論文は博士(情報学)の学位論文として価値あるものと認める。また、令和7年2月19日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。

なお、本論文は、京都大学学位規程第14条第2項に該当するものと判断し、公表に際しては、当該内容が別途公開されるまでの間、当該論文の全文に代えてその内容を要約したものとすることを認める。