様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月　 ５ 日現在

機関番号：１２５０１
研究種目：挑戦的萌芽研究
研究期間：2016～2018
課題番号：１６Ｋ１２３９１

研究課題名（和文）論理的に隙のない情報理論テキストの自動生成

研究課題名（英文）Automated Generation of No Logical Gap Readable Proof for Information Theory

研究代表者
萩原　学（Hagiwara, Manabu）

千葉大学・大学院理学研究院・准教授

研究者番号：８０４１５７２８

交付決定額（研究期間全体）：（直接経費）　　2,600,000円

研究成果の概要（和文）：Coq/MathComp上の情報理論ライブラリInfoTheoの活用として、そのライブラリの.vファイルを人間が読解可能な文章へ変換するソフトウェアの実装に挑戦した。アプローチはソフトウェアProviollaをベースとしつつ、Pythonベースの完全オリジナルソースコードを作成し、MathCompのタクティクやCoqのコマンドを自然言語へ置き換えるものである。

研究成果の学術的意義や社会的意義
情報理論の著名な諸定理を論理的な隙間なしに解説する文書が作成されることは、情報理論の基礎理論を頑健たるものにするだけでなく、初学者が誤解なく自主学習できる文献を提供できる。論理的に隙間のない証明として、計算機向けに形式化された証明が存在する。しかしこれは、人間にとってのReadabilityが欠如している。本研究の目標が達成されれば、隙間なくReadabilityのある証明が得られることになる。

研究成果の概要（英文）：As an application of the library InfoTheo on Coq / MathComp for information theory, we challenged the implementation of software that converts .v files of the library into human readable natural language. The approach is based on customization of the software Proviolla and creating completely original source code on Python, and replacing MathComp tactics and Coq commands with human readable natural language.

研究分野：符号理論

キーワード：情報理論　形式化

様　式　Ｃ－１９，Ｆ－１９－１，Ｚ－１９，ＣＫ－１９（共通）

１．研究開始当初の背景

情報理論の起源はシャノンの文献[1]と言われ，そこで情報エントロピー，情報源符号化，通信路モデル，通信路符号化といった情報理論の基本的な概念・理論が提示された．発表当時は厳密ではない概念や論理が顕在したと言われているが，後の情報理論研究者の貢献により，人間の感覚として厳密な理論へ整理された．その後，諸理論は体系的に纏められ，情報理論に関する名著が数多く出版されてきた（[2,3,4]など）．一方，厳密な理論と言えば，論理学における形式論理が挙げられる．形式論理では論理を公理論的・代数的に扱うことで，論理的正しさの厳密な定義を導入し，感覚に依らない普遍的な真理の保証や発掘を可能とする．しかし，形式論理による理論の記述は非常に複雑であり１＋１＝２の証明でさえ，形式論理で記述すると７００ページ以上の紙面が必要となる[5]．このように，理論の厳密化は人間には困難な作業である．

近年，形式論理と計算機科学の融合研究が発展し，定理証明支援系と呼ばれる計算システムの実用化が近づいている．ここで定理証明支援系とは，論理の形式論理化をサポートするソフトウェアを指す．このシステムは，形式論理に従って記述した証明中に誤りがないかどうかチェックしたり，過去に証明した補題等を計算機上の関数ライブラリとして扱えたり，複雑な論理式にノーティションを導入することで直観的な把握を可能とする，といった証明の支援を行う．

これまでに研究代表者は，定理証明支援系 Coq/SSReflect を活用し，シャノン理論の形式化に挑んできた．結果，通信路符号化順定理・逆定理，固定長情報源符号化順定理・逆定理の形式化に成功した．その後，研究分担者の葛岡氏からアドバイスを受け，可変長情報源符号化順定理・逆定理の形式化にも成功した．つまり，シャノンによる情報理論の著名な定理に対し，形式論理的に厳密な証明を与えたことになる．ところがこれらの証明は，証明言語（形式論理に対応する計算機言語）で書かれる為，人間にとって非常に読みづらい．厳密な証明を作り上げても，計算機にしか読めず，人間には理解しづらい状況に，研究代表者らは疑問を感じてきた．

２．研究の目的

これまで研究代表者・分担者は，定理証明支援系 Coq/SSReflect を活用し，通信路符号化順定理・逆定理，固定長情報源符号化順定理・逆定理，可変長情報源符号化順定理・逆定理になどに対して形式論理的に厳密な証明，いわゆる形式化に成功してきた．

しかし，これらの証明は，証明言語（形式論理に対応する計算機言語）で書かれる為，人間にとって非常に読みづらい．そこで，証明言語から，人間にとって読みやすい文章へ変換する証明変換システムの構築・その為の理論の基礎構築を行うことを研究目的とする．

人間には見落としがちな論理構造等の発見，証明の論理展開が細部まで書かれた情報理論の教科書の作成にも繋がり，学問的価値の更なる向上に寄与できる．

３．研究の方法

情報理論に関する証明言語上のライブラリ infotheo に対して，「依存関係解析器」「証明課程加工器」「記法 LaTeX 化器」「命令変換器」および「証明支援系カーネルとのアクセス器」を構成することで，証明変換システムを構築する．情報理論とその形式化（証明言語による記述）で実績のある研究代表者・分担者とその学生に加え，定理証明支援系のソースコードである計算機言語 OCaml の主開発者 J.Garrique 氏を研究協力者に迎えることで，証明変換システムの実現可能性が高い体制としている．また解釈論を証明変換に導入し，人間にとっての読みやすさを向上させる．

一般的なアプローチは，論文や書籍などに書かれた理論を証明言語へ変換する．本研究のアプローチは真逆であり，証明言語から論文や書籍への変換を目指す．
その戦略として，情報理論証明独特の表現に注目しテキスト化を簡便化すること，定理証明支援系のカーネルを直接操作すること，最近のマスデジタリゼーションの依存関係解析の手法を導入すること，証明言語 SSReflect の特徴である命令の少なさに着目することなどを考えている．

近年のデジタリゼーションの結果として，ドキュメント内の数式検索には ML の意味論的解釈ではなく，記号的解釈が効果的であることが報告されている．そこで，証明変換に対するこのような解釈の効果を研究する．また，証明は日常言語と異なり癖のある表現で記述されることが多く．例えば「We denote a set by X.」のような癖のある表現に着目することで変換の方法が制限され，証明言語から人間の言葉への証明変換の可能性が開くと考えている．

４．研究成果

情報理論（とくに情報源符号化定理）の形式化ライブラリから人間が読み易いテキストを自動生成するシステムの開発に取り組んだ．具体的には，情報源符号化定理の形式化を含んでいる Coq/SSReflect ライブラリである InfoTheo の .v ファイル群に対し，タクティク等を自然言語へ変換するソフトウェアの開発を行った．

ソフトウェアは Python のスクリプトとして作成した．また，Coq と Python 間のデータの橋

渡しとして，Carst Tankink により開発されたフリーソフトの proviola を参考にした．これは標準出力とエラー出力をパイプとした，Coq の入力と出力をやりとりする手法である．より具体的には，開発されたソフトウェアは次のような二段階処理により，形式化ライブラリを人間が読み易いテキストへ変換する．まず前処理として，Coq/SSReflect スクリプトに対して，proviola による film 化と同様の操作を施す．その後，前処理を済ませたファイルを LaTeX のソースファイルへと変換することで，人間が読み易いテキストを生成する．

作成したソフトウェアは，研究体制（代表者，分担者）で協力し動作を確認した．その結果，InfoTheo の.v ファイルの一部に対して，ある程度まで自然言語化することが出来た．しかしながら，全てのファイルを完全に自然言語化するには人力による調整が必要であった．人力による調整なしの完全な自動化のためには，さらなるブレイクスルーが必要である．検討の結果，自然言語化をスムースに実現するには，.v ファイル作成時からフォーマット等のルールを定めて行うべきという知見に達した．また，動作環境が異なることで，機能の一部が正常に動作しない状況も確認された．これは，Python と Coq の OS に依存した動作によるところが大きい．どの OS でも同一のもしくは同様の動作を実現するのは非常に困難である．このような Coq の入出力操作に関する OS 依存性を明らかにしたことも，本研究によって得られた知見の一つである．

さらに，本プロジェクトの副産物的な成果として，情報源符号化における新たな理論的成果を得ることもできた．一般的な情報理論の教科書においては，情報源符号化における符号の性能評価の尺度として「情報源出力１文字当たりの平均符号語長」を考え，このとき達成可能な最適性能が「シャノン・エントロピー」と一致することを示すのが標準的である．しかしながら，符号語長の平均値だけが唯一の評価尺度であるという理論的必然性は全くない．そこで本研究では，平均値よりも一般的な，符号語長のモーメントを評価尺度とした場合について考察した．その結果，復号誤りが生じることを許容する場合には，達成可能な最適な符号語長モーメントは「スムース・レニー・エントロピー」と呼ばれる量で特徴づけられることを明らかにした．このように，本プロジェクトを通して，単に論理展開を精密化したテキストを生成するだけではなく，理論展開をより深化させ新たな成果へ発展させることもできた．

５．主な発表論文等

〔雑誌論文〕（計１６件）

①Justin Kong, Webb David J., Hagiwara Manabu, Formalization of Insertion/Deletion Codes and the Levenshtein Metric in Lean, Proc. of ISITA, pp.1-6, vol.1, 2019.査読有
DOI: 10.23919/ISITA.2018.8664354

②Shigeaki Kuzuoka, An application of universal FV codes to source coding allowing errors, IEICE Technical Report, pp.25-30, 2017 年, 査読なし

③Manabu Hagiwara, Justin Kong, Consolidation for compact constraints and Kendall tau LP decodable permutation codes, Designs, Codes and Cryptography, vol85. pp.483-521, 2017. 査読有
DOI: 10.1007/s10623-016-0313-5

④Ken’ichi Kuga, Manabu Hagiwara and Mitsuharu Yamamoto, Formalization of Bing’s Shrinking Method in Geometric Topology, Intelligent Computer Mathematics, LNCS 9791, pp.18-27, 2016. 査読有り
DOI: 10.1007/978-3-319-42547-4_2

⑤W. Matsumoto, Manabu Hagiwara, P. T. Boufounos, K. Fukushima, T. Mariyama, Z. Xiongxin, A Deep Neural Network Architecture Using Dimensionality Reduction with Sparse Matrices, Neural Information Processing, LNCS9950, pp.397-404, 2016, 査読有り
DOI: 10.1007/978-3-319-46681-1_48

⑥萩原学, ポストモダン符号理論としてのネットワーク, 置換, 形式化２：ネットワーク符号, 日本応用数理学会誌　応用数理, vo.26(2), pp.75-80, 2016, 査読なし
DOI: 10.11540/bjsiam.26.2_27

⑦萩原学, 誤り訂正符号の例と将来展望, 映像情報メディア学会誌, vol.70(7), pp.567-570, 2016, 査読なし

⑧萩原学, ポストモダン符号理論としてのネットワーク, 置換, 形式化３：置換符号, 日本応用数理学会誌　応用数理, vo.26(3), pp.125-130, 2016, 査読なし
DOI: 10.11540/bjsiam.26.3_29

⑨萩原学, ポストモダン符号理論としてのネットワーク, 置換, 形式化４：符号理論の形式化,

日本応用数理学会誌　応用数理, vo.26(4), pp.172-77, 2016, 査読なし
DOI: 10.11540/bjsiam.26.4_28

⑩Manabu Hagiwara, On Ordered Syndromes for Multi Insertion/Deletion Error-Correcting Codes, Proceeding of ISIT 2016, pp.625-629, vol.1, 2016, 査読有り

⑪Justin Kong and Manabu Hagiwara, Nonexistence of Perfect Permutation Codes in the Ulam Metric, Proceeding of ISITA 2016, pp.727-731, vol.1, 2016, 査読有り

⑫Manabu Hagiwara, Kyosuke Nakano and Justin Kong, Formalization of Coding Theory using Lean, Proceeding of ISITA 2016, pp.527-531, vol.1, 2016, 査読有り

⑬Manabu Hagiwara and Kyosuke Nakano, Formalization of Binary Symmetric Erasure Channel Based on Infotheo, Proceeding of ISITA 2016, pp.512-516, vol.1, 2016, 査読有り

⑭Shigeaki Kuzuoka, Variable-length coding for mixed sources with side information allowing decoding errors, Proceeding of ISITA 2016, pp.161-165, vol.1, 2016, 査読有り

⑮Shigeaki Kuzuoka and Shun Watanabe, On distributed computing for functions with certain structures, Proceeding of ITW, pp.6-10, vol.1, 2016, 査読有り
DOI: 10.1109/ITW.2016.7606785

⑯Shigeaki Kuzuoka, On the smooth Renyi entropy and variable-length source coding allowing errors, Proceeding of ISIT, pp.745-749, vol.1, 2016, 査読有り
DOI: 10.1109/ISIT.2016.7541398

〔学会発表〕（計６件）
①Shigeaki Kuzuoka, On universal FV coding allowing non-vanishing error probability, the 10th Asia-Europe Workshop on Information Theory (AEW10)（国際学会）, 2017 年

②萩原学, C 型ルート系に付随する挿入削除誤り訂正符号, 情報理論とその応用シンポジウム 2016, 2016 年 12 月 13 日～2016 年 12 月 16 日, 高山グリーンホテル（岐阜県）

③中野恭輔, 萩原学, Coq/SSReflect による通信路の同型性の形式化, 情報理論とその応用シンポジウム 2016, 2016 年 12 月 13 日～2016 年 12 月 16 日, 高山グリーンホテル（岐阜県）

④葛岡成晃, An application of Iriyama's Lemma for multiterminal source coding systems, 情報理論とその応用シンポジウム 2016, 2016 年 12 月 13 日～2016 年 12 月 16 日, 高山グリーンホテル（岐阜県）

⑤Manabu Hagiwara, Shifted Young diagrams and binary I/D error-correcting codes, SIAM Conference on Discrete Mathematics（国際学会）, 2016 年 06 月 06 日～2016 年 06 月 10 日, 米国　ジョージア州

⑥葛岡成晃, 関数計算のためのデータ圧縮－関数の二分法によるアプローチ－, 電子情報通信学会情報理論研究会（招待講演）, 2016 年 05 月 19 日～2016 年 05 月 20 日, 小樽経済センター（北海道）

〔図書〕（計２件）
①萩原学, アフェルト・レナルド, Coq/SSReflect/MathComp による定理証明, 森北出版, 総ページ数２２４, 2018 年

②萩原学（編者）, 進化する符号理論, 日本評論社, 総ページ数 206, 2016 年

６．研究組織

(1)研究分担者
研究分担者氏名：葛岡　成晃
ローマ字氏名：Shigeaki Kuzuoka
所属研究機関名：和歌山大学

部局名：システム工学部
職名：准教授
研究者番号（8 桁）：60452538

(2)研究協力者
研究協力者氏名： ジャック　ガリグ
ローマ字氏名：Jacques Garrigue

研究協力者氏名： 中野 恭輔
ローマ字氏名：Kyosuke Nakano

研究協力者氏名： ジャスティン　コング
ローマ字氏名：Justin Kong