

令和 2 年 6 月 18 日現在

機関番号：35409

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K00163

研究課題名(和文) データベースから予測失敗の原因を探る 直腸癌・放射化学療法の目利き支援

研究課題名(英文) Exploring the causes of predictive failure from a database - Connoisseurship support for rectal cancer and radiochemotherapy -

研究代表者

金子 邦彦 (KANEKO, Kunihiko)

福山大学・工学部・教授

研究者番号：50274494

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：種々のデータのデータベースを統計的に解析するというプロセスにおいては、解析作業をコンピュータで行う手間、そして、解析作業をコンピュータで行うための設定準備作業の手間が問題になる。放射化学療法の効果に関する匿名データ、舌画像に関する匿名データ、顔写真、運動姿勢のビデオのデータベースについて、それぞれの分析に適する「解析シナリオ」を新しく定め、その解析シナリオを、専門家だけでなく、ふつうのパソコンでできるようにするという問題に取り組んだ。その結果として、オープンなソフトウェアの集合体により、容易にパソコンでの解析シナリオ実施ができることを確認。その準備作業はすべて明文化し、再現可能とした。

研究成果の学術的意義や社会的意義

ビッグデータ社会、人工知能社会の到来により、データを裏付けとした実験、解析、シミュレーションは、欠かせないものとなってきた。従来は、特定の目的を定めパッケージ化されたITシステムを用いることが多かったが、柔軟性に欠き、失敗に終わる(意味のある結論が得られない)場合もあった。近年、オープンサイエンスの時代が到来し、人工知能の技術、統計解析の技術の種々のソースコードが公開され一定の条件下で利用できるようになってきた。そうしたソースコードの集合体を集め、Windowsのパソコンで動くようにする手順を明確にした。誰でも再現できるものである。人工知能や統計解析をIT専門家でなくても簡単に実施できる。

研究成果の概要(英文)：In the process of statistically analyse of database, the time and effort to perform the analysis by a computer and the time and effort to prepare a computing environment are problems. I tackled the problem. First, I decided "analysis scenarios" of database applications. Second, I established a method to prepare a computing environment suitable for the database applications. In this study, I decided analysis scenarios for the analysis of anonymized data on radiochemotherapy, anonymized data on tongue images, face images, and videos of motor postures. Then analysis scenarios using statistical and AI methods are decided. In order to make the scenarios available on an ordinary personal computer without being an expert, I used open softwares. As a result, it is confirmed that a set of open software can easily run analysis scenarios on personal computer. All the preparatory work has been made explicit of a Web page and reproducible.

研究分野：データベース応用, 人工知能応用

キーワード：データベース応用 人工知能応用 統計解析応用 ITシステム オープンソフトウェア活用

1. 研究開始当初の背景

2015年に医学研究者との情報交換の場において、大腸癌・直腸癌のエクセル形式のデータ(医師が患部の大きさや治り具合を評価したデータ)と遺伝子のデータの解析の困難さが話題となった。データの解析には、あるITのパッケージソフトウェアが使用されており、メニュー形式で機能を選ぶものであった。メニューにない機能を簡単に増やせるものではない。メニューにある機能を種々組み合わせるが、意味のある結果が得られず失敗するという状況であった。一方で、当時、パッケージソフトウェアでなく、RやPythonといった人工知能や統計解析に向き、データベースとの連携とも容易なオープンなプログラミング言語が勃興しつつあった。RやPythonでは、当時の従来のパッケージソフトウェアよりも機能が充実している(公開パッケージという形で)上に、プログラムのソースコードを自ら製作し組み込むことも可能であり、有望であると考えた。

2. 研究の目的

医学等の専門家が、パソコンを使って、種々のデータのデータベースを統計的にあるいは人工知能を用いて解析するというプロセスにおいて、なるべく失敗なく行えるよう、機能の充実したRやPythonといったオープンなプログラミング言語を使うとする。条件は次の通りである。

- ・結果はビジュアルに表示できること。(RやPythonはビジュアルに表示する機能は容易であり研究課題としない)
- ・Windowsのパソコンで動くこと。(RやPythonの公開パッケージその他オープンソフトウェアは、Linuxでは動くが、Windowsの動作は想定されていないことが多々ある。それらのソースコードを書き換えるという移植という作業は好ましくない。移植せずに、いかに、動作させるかの課題がある)
- ・特定目的でなく一般性があること。

3. 研究の方法

次の4種類のデータベースを作成する。

- (1) 放射化学療法の効果に関する匿名データ
- (2) 舌画像に関する匿名データ
- (3) 顔写真
- (4) 運動姿勢のビデオのデータベース

うち「(1)」は、研究開始時点までに準備ができており、取り扱いに慣れていたデータであったため「(1)」で研究を着手した。「2. 研究の目的」に示したように、特定目的でなく一般性があることを目指すため、研究期間中に検討を進め、2, 3, 4のデータベースを段階的に増やした。(なお、いずれも、個人情報保護、プライバシーの確保、著作権等に関する法令順守の他を遵守しながら取り扱った)。これらデータベースについて、実際の解析シナリオを次の通り定めた。解析シナリオの作成自体も、調査、研究を必要とするものであった。

以上のデータベースそれぞれについて解析シナリオを定める必要があった。

(1) 40000近くある遺伝子データの中から放射化学療法の効果との関連の高いものを選びだすとき、遺伝子データを主成分分析したうえで、その中で、何個の主成分を選ぶことが最適かを探す問題。シルエット距離の変化により最適かを探ることができることを見出しており、分析を実施した。

(2) 舌画像の時間経過による色の変化(詳しくはLab色空間でのa, b成分の時間変化)が別の因子と相関関係にあるかを分析する問題。舌画像の色の変化は、撮影直後の数秒、その後数十秒程度、それ以後に分かれることを研究期間内に見出すことができた(撮影直後の緊張の緩和、そして疲れが出ていると想像される)。そこで、舌画像の色の変化を単純な直線近似でなく、3線分のマッチングで近似し、分析を実施した。

(3) 顔写真のマッチングによる個体識別の問題。これは、近年の人工知能技術の進展により、PythonのパッケージDlibと関連ソフトウェアにより個体識別ができることを調査、評価し、それを用いての分析を実施した。

(4) 人工知能による架空ビデオの合成における問題。これは、ある人物の運動のビデオと、別人物の写真1枚より、その別の人物のビデオを合成するものである。人工知能技術GANの進展により、この問題は精度よくできつつある。ここでは、実験室において、「GANを使う」ことを意識しない状態で写真を撮影し、それがどういう場合にうまく使えるか(あるいは使えないか)を確認するという分析を実施した。

以上、種々のデータベースと解析シナリオを定め、「2. 研究の目的」に取り組んだ。

4. 研究成果

まず、(1),(2),(3),(4)の解析シナリオについては、すべての実施を終えた。

その後、「2. 研究の目的」にあるように、Windowsのパソコンで、(1),(2),(3),(4)の解析シナリオが簡単に実施できるようなコンピューティング環境を、RやPythonの公開パッ

ページその他オープンソフトウェアで整えることができるかに取り組んだ。これらは、オープンなソフトウェアであるため入手に問題はないが、Windows での動作は苦心した。他の研究者と（医学分野）の情報交換では、エクセルとの連携、あるいは、Windows パソコンでの操作になれているという理由から、Linux は見送った。結果として、コンピューティング環境の試作を終えることができた。その準備作業はすべて明文化し、いつでも Windows パソコンで再現可能としている。将来、Windows のバージョンアップ、Visual Studio のバージョンアップにより、その明文化された作業手順はそのまま使えるものではないが、明文化されていることにより、これらバージョンアップにおける対応は困難なく、見通し良く行える見込みである。

なお、金子邦彦 Web ページ内 (<https://www.kkaneko.jp/index-j.html>) でも成果を公開している。本研究とは直接関係はないが、小型コンピュータ Raspberry Pi での取り組み（Linux をベースとしており、本研究課題での Windows 利用に起因する問題は回避できている）に、本研究課題の経験を生かすことができた。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計5件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 安保真理子, 澄田有紀, 金子邦彦
2. 発表標題 運動姿勢の比較における人工知能の活用
3. 学会等名 令和元年度(第70回)電気・情報関連学会中国支部連合大会
4. 発表年 2019年

1. 発表者名 澄田有紀, 安保真理子, 金子邦彦
2. 発表標題 運動再現技術を用いた運動技能向上の試み
3. 学会等名 令和元年度(第70回)電気・情報関連学会中国支部連合大会
4. 発表年 2019年

1. 発表者名 金子邦彦
2. 発表標題 顔画像と顔情報のデータベース設計の試み
3. 学会等名 2019年電子情報通信学会総合大会D-4-7
4. 発表年 2019年

1. 発表者名 金子邦彦
2. 発表標題 深層学習表情認識システム
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2019年

1. 発表者名 金子邦彦
2. 発表標題 深層学習を用いた複数カメラの人物再識別
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

金子研究室ホームページの中で、成果を公開している。 https://www.kkaneko.jp/index-j.html
--

6. 研究組織		
氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考