**研究内容（10/7）**

**わかりやすく**

部分集合選択問題の解決におけるRBSC-SubGenアルゴリズムの挙動の分析というテーマで、Pythonプログラムコードの開発と効率化を検討した。上記のアルゴリズムはEラーニングの向上を目的としていて、英単語アプリの単語難易度を分ける仕組みの効率化を図ったものである。

**概要**

本研究では、計算統計学の部分集合生成問題に注目し、その解決にランク双列相関（RBSC）ベースのデッキ生成アルゴリズムRBSC-SubGenを利用する。RBSC-SubGenは元々、望ましい難易度の英単語の集合を自動的に多数構築するために設計されており、一般的な部分集合選択問題と多くの共通点を持つ。RBSC-SubGenにおいて、単語データベースだけでなく任意のランク付けされた項目集合への適用を検討し、先行研究で扱われていない様々なハイパーパラメータに対する耐性を研究する。すなわち、様々な条件下でRBSC-SubGenの挙動を確認し、望ましいRBSCに対する飽和と格差の観点から脆弱な側面を示す。

**役割**

希望するレベルの単語頻度関係を持つ語彙デッキを、大規模な単語集合から希望数だけ自動的に構築するRBSC-SubGenアルゴリズムの検討

**技術的な工夫**

通販サイトのおすすめ商品や（英単語などの）学習アプリのお任せ学習が、裏でどのようなアルゴリズムを動かしているのか理解が乏しく、またそのようなサービスのデータ分析についてさらに効率化できるのではないかと考え、この研究を行った。

研究で部分集合選択問題と相関係数、単純差の概念を取り入れ、データ分析の面白みを実感できた。