

視覚に着目した変数命名法の分析



1. はじめに

研究背景:

自分がプログラミングを行う際、変数名の単語の形が「美しい」「美しくない」を気にして命名していることに気付いた。

例)

美しい: 「sample」「document」「variable」

美しくない: 「center」「paper」「sphere」

目的:

自分の変数命名の主観的な基準を明確にする



2. 仮説

変数名の美しさは単語に含まれるアルファベットの段(図1)の数に影響されると考えた

- ・上段例... 「t」, 「i」, 「f」
- ・中段例... 「c」, 「w」, 「e」
- ・下段例... 「p」, 「q」, 「g」

上に突き出る

下に突き出る



図1 各アルファベットの段の例

3. 英単語の調査

サンプルの収集

英単語帳の英単語236個を対象にデータを収集した

主観的データ:

自分の中で変数名が美しいか (合格、不合格)

- ・合格... 169個 (約71.6%)
- ・不合格... 67個 (約28.4%)

ほとんどの単語が合格である

客観的データ:

単語に含まれるアルファベットの段の個数

- ・合格「sample」... 上段: 1個 中段: 4個 下段: 1個
- ・不合格「paper」... 上段: 0個 中段: 3個 下段: 2個

各段の影響力を調査

単語に含まれるアルファベットの段数の個数が、合格の影響力に関係あるかをロジスティック回帰にて分析する

(分析A) 全ての段を説明変数に利用した場合

⇒ 上段: 0.37 中段: -0.12 下段: -0.63

(分析B) 上段と下段を説明変数に利用した場合

⇒ 上段: 0.41 下段: -0.60

結果:

分析Bは上段と下段の影響力(絶対値)が似ている上段と下段の個数の差により合格の予測が可能

4. 特徴の抽出

合格率より合否の予測:

- ・上段と下段の個数の差での合格の割合を調べた(図2)ただし、合格と不合格の単語数には差がある
- ・そこで、合格単語を不合格単語に正規化して合格率を算出した

$$\text{合格率} = \frac{\text{正規化した合格単語数}}{\text{不合格単語数} + \text{正規化した合格単語数}}$$

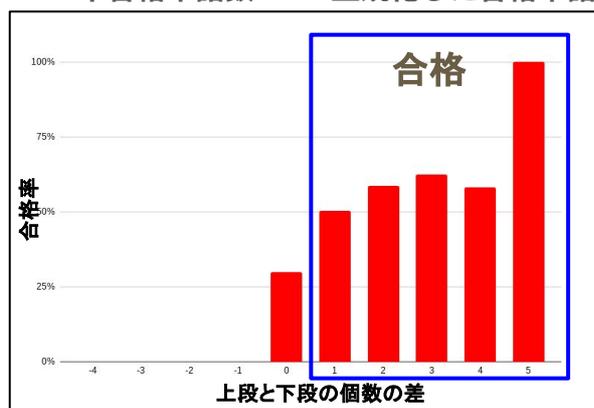


図2 上段と下段の個数の差から調べた合格率

結果:

合格率が50%を超えたものを合格と予測する

5. 検証

収集したサンプル500個程度の英単語データを対象に検証を行った

この検証では、「合否の予測」と「主観的データ」どの程度一致してるかを調べる

合否の予測:

上段と下段の個数の差が1以上であれば合格

主観的データ:

自分の中で変数名が美しいか (合格、不合格)

結果:

予測精度

約75.58%

考察:

変数名を美しいと感じる平均的な基準から約25%離れている。主観的な基準としての精度であるが、他人の意見を集められれば更に精度が高められると感じた。

6. まとめ・今後の展望

まとめ:

単語の上段の個数と下段の個数のバランスである程度美しいかどうかが決まることが分かった。

今後の展望:

- ・これらの分析結果を利用して、プログラマーに変数の美しさを表示してくれる支援機能を提供したい
- ・他の人はどのような考え方で変数を命名をしているのかを明らかにしていき、精度を上げていきたい

引用文献:

「英単語検定」 <https://www.eitangokentei.com/chu1-eitango/>