

キー入力の癖の学習に基づく 入力間違い推定法の検討

M0113347 西村 希規

研究背景

レポートや論文をPCを使用して書くことが多くなっている。その際、特にPC操作に慣れていない学生は、誤って違う文字を入力してしまうことが考えられる。また、その文字を目視で確認するのは非常に困難である。

この問題を解決するためには、文章校正機能の必要性は高いと考えられる。

研究の目的

現在の文章校正ツールは、間違いのパターンに登録されている単語以外は文書校正されない。本研究では、PCのキーボード入力から間違えて入力した文字のパターン取得し、どのキーの入力が苦手か解析する。解析結果を用いて、間違えやすいキーの文字をマーキングする。そしてユーザーに、この文字は間違えてないか確認してもらう仕組みを構築する。

キー入力からの間違い推定

手法としては以下の通りである。

キーロガーを使用し、キーボードの入力を取得する。以下の図にあるU[BackSpace]Iが間違えて入力されたものである。この場合、Iを入力しようとしたが、間違えてUを入力してしまったものである。これをキーロガーのログから抽出する。

CHU[BackSpace]INE
キーロガーで文字を取得した時の例

先ほど、キーロガーで取得した文字の入力データに対し、機械学習を使用して、間違えて入力されたパターンがどのくらい出現したのかを計算する。テキストの分類や特徴の取得に

特化している機械学習ライブラリーである gemsim 内の Phrases という単語の頻出回数、単語同士の関連性を解析する機能を利用している。

解析すると以下の図の結果が得られる。

F [Delete]_G 6	F [Delete]_G 0.0612244897959
S [Delete]_D 4	S [Delete]_D 0.0408163265306
O [Delete]_I 3	O [Delete]_I 0.030612244898
T [Delete]_N 3	T [Delete]_N 0.030612244898
C [Delete]_B 3	C [Delete]_B 0.030612244898
C [Delete]_X 9	C [Delete]_X 0.0918367346939
C [Delete]_V 4	C [Delete]_V 0.0408163265306
; [Delete]_L 6	; [Delete]_L 0.0612244897959
B [Delete]_V 11	B [Delete]_V 0.112244897959

実際に間違えて入力した文字の回数を解析した例 実際に間違えて入力した文字の確率を計算した例

見方

F [Delete]_G 6

上記の図の場合、Gを打とうとしたが、誤ってFを入力してしまい、それをDeleteキーで消去し、正しい文字であるGを打ったケースが6回あったということが分かる。

今後の予定

キーロガーによって間違えたパターンとどのくらい間違えたのかを解析することは実現できた。今後は上記の確率の結果を使用して、特に間違える確率が高い文字は、以下の図のように文字をマーキングし、ユーザーに間違えていないか注意を促す仕組みを作成していく予定である。

文字をマーキングした例

