

二文字入力方式の実装と評価

天野 岳洋 学生番号 6930-36-6945

June 6, 2024

Abstract

本研究では、統計的な二文字入力方式の実装を行い、n-gram モデル、RNN、KyTea ベースの3つのアプローチを比較する。これらの手法の性能を評価するために text8 というデータセットを準備し、トレーニング部分でモデルを訓練し、テスト部分で評価を行いました。結果として、各モデルの精度と有効性を報告する。

1 はじめに

近年、入力方式の効率と精度は、モバイルテキストからプロフェッショナルなライティングツールまで、ユーザーエクスペリエンスを向上させるために重要な要素となっている。そこで本研究では、二文字入力方式の実装を通じて、入力効率の向上を目指しました。具体的には、n-gram モデル、リカレントニューラルネットワーク (RNN)、および KyTea を用いた3つのアプローチを検討しました。

2 方法

本研究では、以下の3つのアプローチを用いて統計的な二文字入力方式を実装しました。

2.1 n-gram モデル

n-gram モデルを使用して、先頭2文字を基に次に続く文字を予測しました。n-gram のサイズやスムージング技法などの詳細は、補完してください。

2.2 リカレントニューラルネットワーク (RNN)

RNN を使用して、文字列の連続性を考慮しながら次の文字を予測しました。RNN のアーキテクチャやトレーニングパラメータなどの詳細は、補完してください。

2.3 KyTea ベースのセグメンテーション

KyTea を使用してテキストを分割し、二文字入力方式の予測精度を向上させました。KyTea の設定や使用方法などの詳細は、補完してください。

3 言語リソース（データセット）

本研究では、データセットとして適当なテキストコーパスを使用しました。データセットの準備方法や内容については以下の通りです。

項目	詳細
データセット名	-----
トレーニングデータのサイズ	-----
テストデータのサイズ	-----
言語	日本語
ソース	-----

Table 1: データセットの仕様

4 実験評価

各モデルの性能を評価するために、トレーニングデータとテストデータを使用して実験を行いました。評価指標としては精度を用いました。各手法の精度は以下の通りです。

4.1 n-gram モデルの評価

n-gram モデルの評価結果について補完してください。

4.2 RNN の評価

RNN の評価結果について補完してください。

4.3 KyTea ベースの評価

KyTea ベースの評価結果について補完してください。

5 結論

本研究では、統計的な二文字入力方式の3つのアプローチを比較しました。それぞれの手法において異なる利点と欠点が見られましたが、全体的には_____が最も高い精度を示しました。今後の研究では、さらに大規模なデータセットを使用してモデルの改善を図るとともに、ユーザーインターフェースの最適化も検討していきます。