



Bengali Text Summarization

Md Mahfuzur Rahaman, Sujoy Nath, Shahnur Islam
Shahjalal University of Science and Technology, Sylhet

1 Introduction

Summarization is the task of condensing a piece of of text to a shorter version that contains the main information from the original.

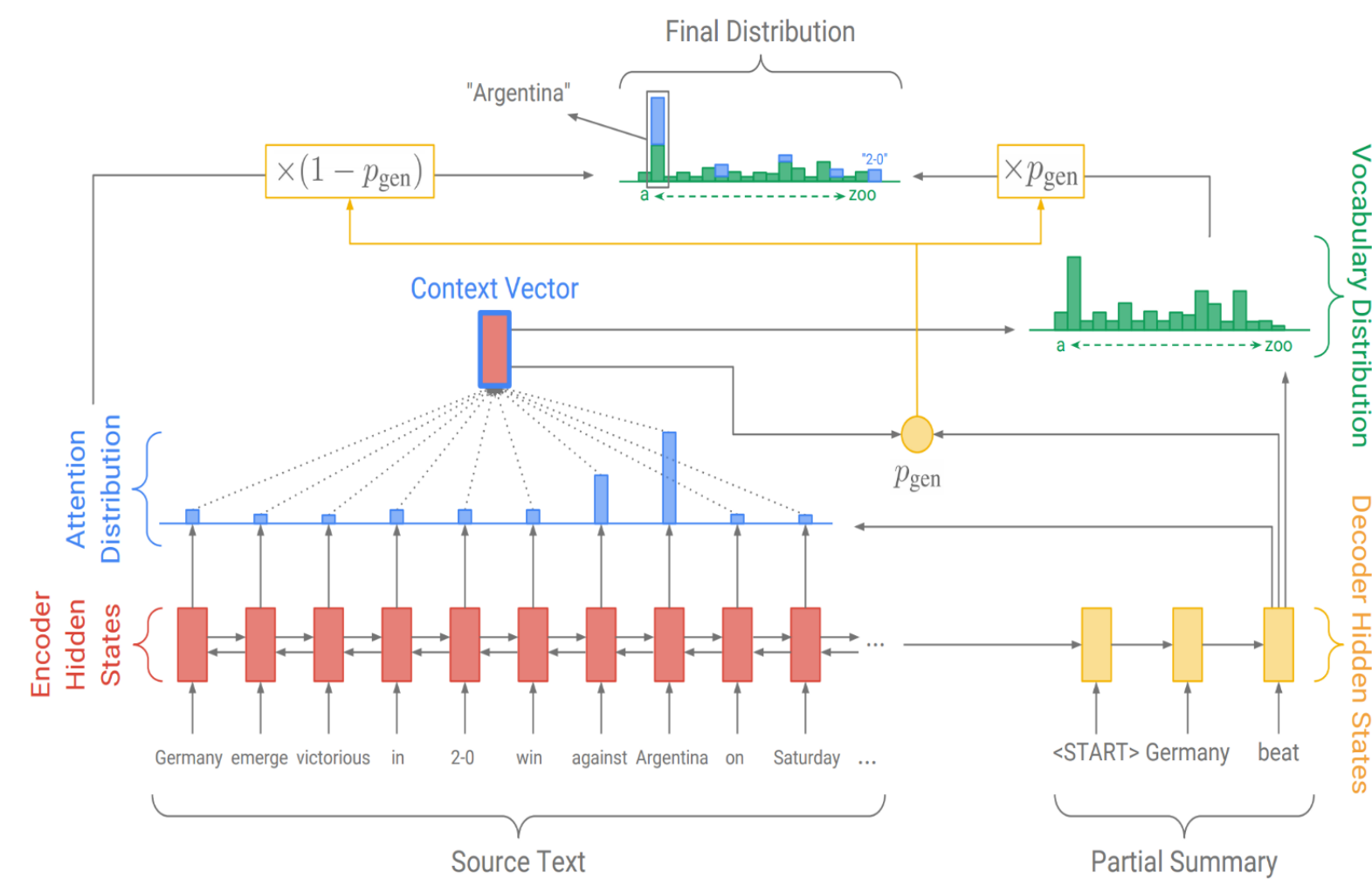
There are two basic approach to summarization :extractive and abstractive.

Extractive is where the machine text actual context and uses as the summary.

Abstractive is a technique used to mimic .This method makes the summarization more condense and human like

Here we have tested both of the methods and have decided to develop abstractive summarization to our project

2 Methods



5 Conclusion

Lots of work have been done so far for English Text Summarization but a few attempt have been made for Bengali Text Summarization.

We have tried to develop a Bengali text summarizer . We have tested the existing methodology and finally we have chosen the sequence to sequence RNN model .

We believe that our work is significant because we design our user study to prioritize human significance rather than simply mathematical significance

In this semester we have worked this method for English text summarization further we will modify and train this model for our Bengali dataset which will provide the actual summary of Bengali document

3 Results

For Extractive Summarization we have used TF-IDF. Here is given an output for extractive summarization

Sentence	Summary
লিওনেল মেসি, না হয় ক্রিস্টিয়ানো রোনালদো, বছরের নির্ধারিত সময়ে যেকোনো একজনের মাথাতেই ওঠে ফিফা বালন ডি'অরের মুকুট। দুশতা ঘেন একেবারেই সাজানো চিত্রনাট্যের মতো হয়ে উঠেছে। কম তো হলো না, সেই ২০০৮ সাল থেকে পাল্লা দিয়ে চলেছে দুজনের বিশ্বসেরা ফুটবলার হওয়ার দৌড়। মেসির নামের পাশে পাঁচটি, রোনালদোর নামের পাশেও তিনটি, এবার কি অন্য রকম কিছু হতে চলেছে?হয়তো যদিও এখনো রোনালদোর দিকে পাল্লা অনেক ভারী, তবে বিশ্বকাপের বছরে এই টুর্নামেন্টের সাফল্য অনেকটাই এগিয়ে দেয় এই আসরের সফলদের। ফিফা বর্ষসরার সঙ্গে আলাদা হয়ে যাওয়ায় এই পুরস্কার ভোটাভুটির অধিকার আবার শুধু সাংবাদিকদের কাছে থাকল।	লিওনেল মেসি, না হয় ক্রিস্টিয়ানো রোনালদো, বছরের নির্ধারিত সময়ে যেকোনো একজনের মাথাতেই ওঠে ফিফা বালন ডি'অরের মুকুট। দুশতা ঘেন একেবারেই সাজানো চিত্রনাট্যের মতো হয়ে উঠেছে। কম তো হলো না, সেই ২০০৮ সাল থেকে পাল্লা দিয়ে চলেছে দুজনের বিশ্বসেরা ফুটবলার হওয়ার দৌড়। মেসির নামের পাশে পাঁচটি, রোনালদোর নামের পাশেও তিনটি, এবার কি অন্য রকম কিছু হতে চলেছে?হয়তো যদিও এখনো রোনালদোর দিকে পাল্লা অনেক ভারী, তবে বিশ্বকাপের বছরে এই টুর্নামেন্টের সাফল্য অনেকটাই এগিয়ে দেয় এই আসরের সফলদের।

For Abstractive Summarization we have used sequence to sequence RNN model .Here is given an output of abstractive summarization

Sentence	Actual Document	Summary
Presedent Vladimir Putin announced plans Monday for the russain military to hold exercise in the former soviet republic of Kyrgystan the itar-tass news agency	Russain Plans military to exercise in Kyrgystan	Putin to launch military exercise in Kyrgyzstan

4 Data

For Extractive Summarization we have used Bengali document from Bengali online newspaper .

For Abstractive summarization we have used English dataset from the news collection by the associated press world wide from English Gigaword.

